

Semi-synthetic Document Image Generation Using Texture Mapping on Scanned 3D Document Shapes

V.C Kieu^{*†}, Nicholas Journet^{*}, Muriel Visani[†], Rémy Mullot[†] and Jean Phillippe Domenger^{*}

^{*}Laboratoire Bordelais de Recherche en Informatique - LaBRI
University of Bordeaux I, Bordeaux, France

[†]Laboratoire Informatique, Image et Interaction - L3i
University of La Rochelle, La Rochelle, France

Email: {vkieu, journet, domenger}@labri.fr, {muriel.visani, remy.mullot}@univ-lr.fr

Abstract—This article presents a method for generating semi-synthetic images of old documents where the pages might be torn (not flat). By using only 2D deformation models, most existing methods give non-realistic synthetic document images. Thus, we propose to use 3D approach for reproducing geometric distortions in real documents. First, our new texture coordinate generation technique extracts texture coordinates of each vertex in the document shape (mesh) resulting from 3D scanning of a real degraded document. Then, any 2D document image can be overlaid on the mesh by using an existing texture image mapping method. As a result, many complex real geometric distortions can be integrated in generated synthetic images. These images then can be used for enriching training sets or for performance evaluation. Our degradation method here is jointly used with the character degradation model we proposed in [1] to generate the 6000 semi-synthetic degraded images of the music score removal staff line competition of ICDAR 2013¹.

I. INTRODUCTION

The performance of document analysis and recognition methods depends on the size and the quality of the training data, no matter whether the data is obtained from real life or synthesized [2]. Using synthetic data (in addition to real life data) has some advantages: rapid generation process, low cost and high accuracy in the ground-truth. Therefore, degradation models in general and geometric distortion models in particular are widely used to generate synthetic images. These images are then used to evaluate or compare the robustness of various document analysis methods (i.e. for symbol recognition & spotting system evaluation in [3], handwriting recognition algorithm in [4], segmentation, restoration. . .) and to enrich the training database in order to improve the system performance [5], [6].

There are three main approaches to degrade an image document: adding noise on an entire or a part of document, degrading characters, and distorting the shape of documents. The noise models are early and widely used to test the robustness of different document image analysis methods towards noise. The noise has both global and local effects. In [7], the document degradation model which simulates four types of classic noises (Gaussian noise, high frequency noise, hard pencil noise and motion blur noise) is proposed by Zhai *et al.* This model is used for assessing the robustness of their line detection algorithm. The noise resulting from the physics of the image acquisition process is modelled and used to

evaluate the performance, as the perturbation model of Loce *et al* [8]. In addition, the effects of ten parameter model for character degradation are studied by Baird [9]. He also described the most frequently observed degradations and their causes in [2]. Based on his work, a degradation model and a method for estimating its parameters have been proposed in [10] for improving OCR. Kanungo *et al* [11] proposed a popular nonlinear local document deformation model mimicking additive noise due to the scanning process. The model was further validated in [12]. Dedicated to binary images, this model adds “salt and pepper” noise, i.e. flipping pixels from background to foreground and vice versa, in the neighbourhood of the characters. However, most of these models only work with binary images. Therefore, several authors simulate most common grayscale defects due to the age of the document itself and the printing/writing process for performance evaluation works. For example, in [13], the bleed-through deformation model is presented. This model is used to compare the robustness of two OCR algorithms when the bleed-through intensity increases. We also proposed a grayscale character degradation model in [1] which simulates most common defects in ancient documents such as ink splotches, white specks or streaks.

The geometric distortions also impact on the system’s performance. The causes of these defects are numerous, for example, the thick and bound document, the incorrect position of camera or document, lens distortion and of course the state of degradation of the old document itself. . . Several document geometric distortion models have previously been proposed. In [11], Kanungo also presented a perspective distortion model. He considered the physical deformed form of thick and bound documents near the spine as a circular arc segment. This model takes into account the defects of the illumination and optical distortion process as well. Jian Liang *et al* [14] generalized the model by using the reverse mapping process in which the thick and bound documents are modelled as the developable surfaces. These surfaces are approximately divided into a group of planar quadrilaterals to simplify the reverse mapping process. The authors then used synthetic images to evaluate the robustness of their geometric rectification algorithm in [14]. Almost all of the geometric distortion models use simple geometric transformation functions and approximate calculations to deal with complex shapes that lead to non-realistic results. Indeed, it is difficult to generate synthetically many geometric distortions observed in real documents (i.e. dents, small folds, tears. . . see examples in Fig. 1) by using geometric transformation functions because they follow no mathematical

¹<http://www.cvc.uab.es/cvcmuscima/competition2013/index.htm>

rules. However, since the ultimate goal of all algorithms is to

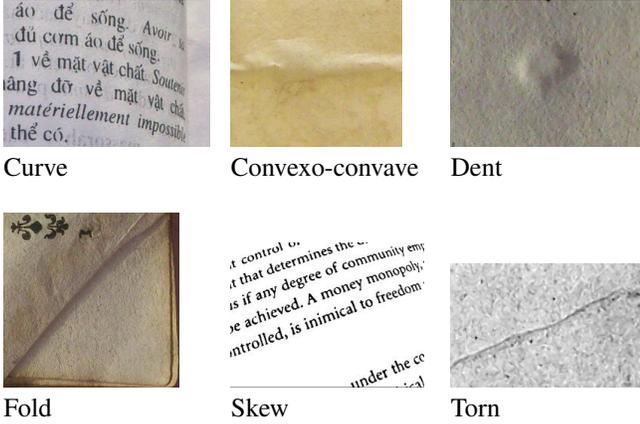


Fig. 1. Geometric distortions in real document images

handle with the real data, we really need to generate these distortions for evaluation works. Thus, we propose a method based on the texture mapping. By using our method, we can take advantage of effectively simulating real-life of 3D texture mapping technique to reproduce real defects. The method has three steps as mentioned in Fig. 2: (A) Real degraded document 3D scan, (B) Texture coordinate generation and (C) Texture image mapping. The main contribution of this article lies on a new proposition for generating realistic semi-synthetic gray-level document images.

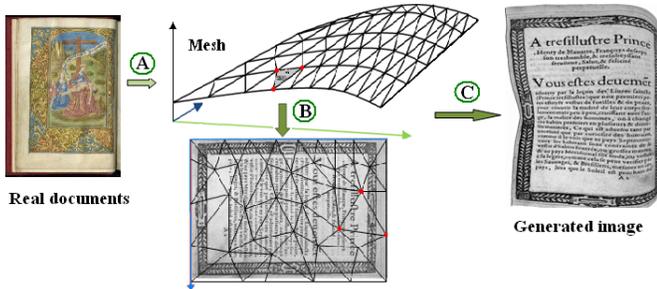


Fig. 2. Texture mapping on real 3D document shape process

This paper is organized as follows: in Section II, our method is described. Experimental results for visual validation are given in Section III. Conclusions and future works are provided in Section IV.

II. OUR METHOD

A. Real degraded document 3D scan process

Hardware. The main purpose of this step is to create a polygon mesh that represents the real document shape as in Fig. 3. The Kréon Aquilon laser 3D scanner at laboratory LaBRI is employed for this work (scan rate up to 1.2 million points per second and the accuracy for distances up to $60\mu m$). The result mesh M is a collection of m vertices ($v_1, v_2 \dots, v_m$), n edges ($e_1, e_2 \dots, e_n$) and p triangles ($t_1, t_2 \dots, t_p$). Every vertex of the mesh will be mapped to a point in the plane in the next step.

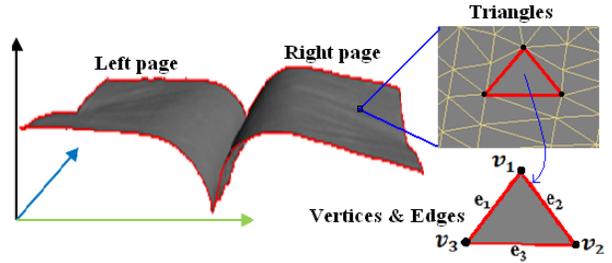


Fig. 3. An mesh of a document shape and triangles inside

B. Texture coordinate generation

The texture coordinate generation, also known as the unwrapping technique, maps every point in the curved surface to the plane. In the field of computer graphics, it generally involves many tasks to deal with high quality images (i.e. segmentation, optimization), which makes the process very time consuming. In mathematics, a curved surface can be firstly divided into planar strips (i.e quadrilaterals, triangles). And then, each strip is mapped to the plane by solving the reverse mapping equation as described in [14]. The process, nevertheless, takes a lot of time when the sampling result needs to be smooth. The other method presented in [15] projects directly each single triangle in the surface onto texture coordinates without considering the distance relation of vertices, which may lead to errors (e.g. stretching or a triangle with the acute angle less than 5 degrees becomes straight). Thus, we propose a technique that takes into account the distance between vertices by unfolding directly a 3D line into the texture coordinates.

In our approach, the mesh of previous step is transformed to the position in which the fore-edge or the spine of document shape is parallel to the plane Oyz before 3D lines are defined. Let P_y be a plane containing at least one vertex v_i provided that $P_y \perp Oy$ (see Fig. 4). Let y be the y -coordinate of the vertex v_i . The plane P_y intersects the mesh M in a 3D line $L_y = \{S_1, S_2, \dots, S_k\}$ where S_k is an intersection point of P_y with an edge e_l . Because $P_y \perp Oy$, all the points S_k have the same y -coordinate. For example, in Fig. 5, the intersection of the plane P_{300} and the mesh M is the 3D line $L_{300} = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9\}$ in which S_6, S_8 are two vertices and its y -coordinate is 300. In the next step, the intersection points in the 3D line will be mapped with points in the texture coordinates (u, v).

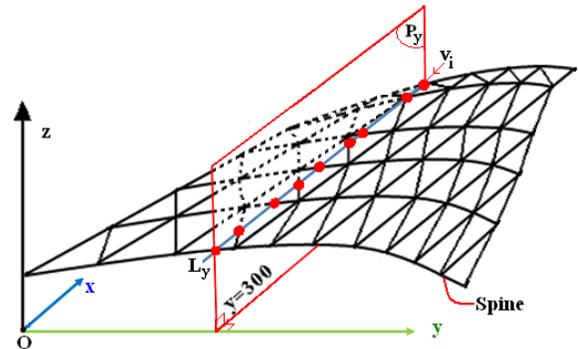


Fig. 4. The plane P_y , which passes through the vertex v_i , intersects with the mesh M in 3D view

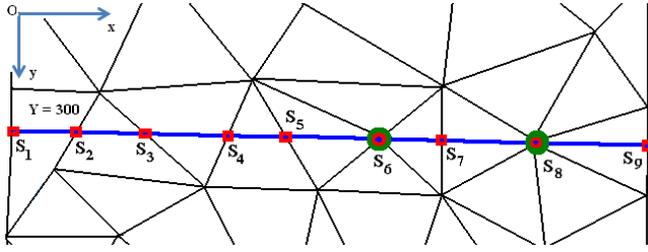


Fig. 5. Zoom of the line intersection in Fig. 4 of the plane P_{300} with the mesh M in vertical view; S_6 and S_8 are two vertices

The 3D line L_y can be then unfolded by projecting it onto the texture coordinates (u,v) on condition that the distance of two adjacent points is conserved. So, let L'_y be the projected line in the texture coordinates and $\{S'_1, S'_2, \dots, S'_k\}$ be the corresponding projected points. As a result, the coordinates $S'_k(u,v)$ are the texture coordinates of the vertex S_k . In our example, the line $L'_{300} = \{S'_1, S'_2, S'_3, S'_4, S'_5, S'_6, S'_7, S'_8, S'_9\}$ is the projection of the line L_{300} provided that $S_1S_2 = S'_1S'_2, S_2S_3 = S'_2S'_3, \dots, S_8S_9 = S'_8S'_9$ (see in Fig. 6).

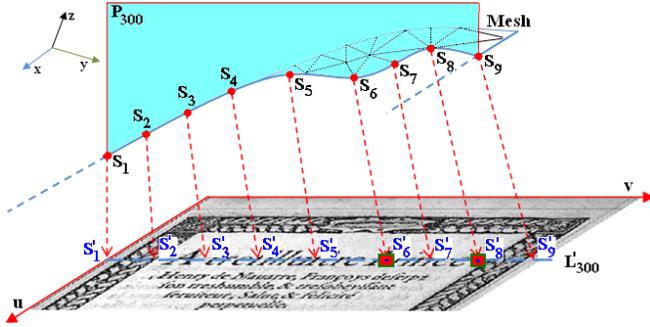


Fig. 6. Projection of the 3D line L_{300} onto the texture coordinates (u,v)

We keep only the texture coordinates of vertices. In our example, S'_6 and S'_8 are saved for the next step. The plane P_y slides from the first vertex to the last vertex of the mesh M until all the texture coordinates of the vertices are calculated. These coordinates are normalized for fitting with the image input in the texture image mapping step.

C. Texture image mapping

Once the mesh has been unfolded, any 2D document image can be overlaid on the mesh by using the texture mapping technique already supported in the library OpenGL. As a consequence of the previous step, the triangles in the mesh have one by one correspondences in the texture coordinates. Generally, in order to pack a document image on the mesh, we put the image on the texture coordinates as a mask (see Fig. 7). Then, we consider each triangle in the texture coordinates. The image pixels inside of that triangle are used to fit in the mesh. For the unfilled pixels, we need an interpolation procedure. The bi-linear interpolation is used in this case. Thanks to the texture mapping technique and Phong reflection model [16], we can calculate exactly where a pixel of the input image is mapped to the output, and we can estimate the grey-level value of this pixel.

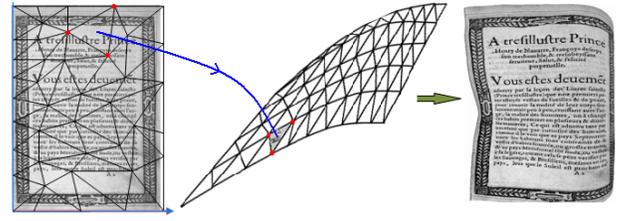


Fig. 7. Packing an image on the mesh with extracted texture coordinates

III. EXPERIMENTAL RESULTS

The real document 3D scan process takes a long time because of scan operations and document collection steps (i.e. finding books, selecting representative pages, installing pages on the scanner...). At the first time, we have considered 20 books among which we selected and scanned four pages that contain almost all of geometric defects. Therefore, we have four meshes (four document shapes) to assess the effectiveness of our approach. First, preliminary visual validation results are given to illustrate the presence of distortions in semi-synthetic generated images. And then, a set of 6000 images is provided as a database for benchmarking in the context of ICDAR 2013 contests.

A. Preliminary visual validation result

Synthetic grayscale images from our database DIGIDOC are used to pack on our meshes. The observed defects in the real documents are reproduced. In Fig. 8, the generated documents illustrate the presence of convexo-concaves, folds, tears. These defects appear frequently in the real document image because of the ageing process and the physical impact of the utilization of the documents.

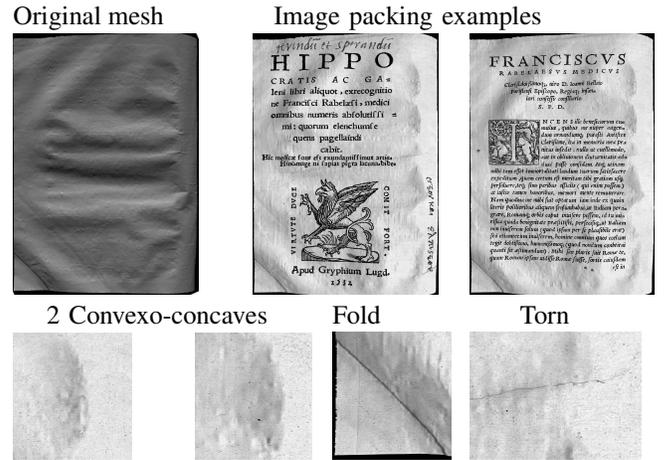
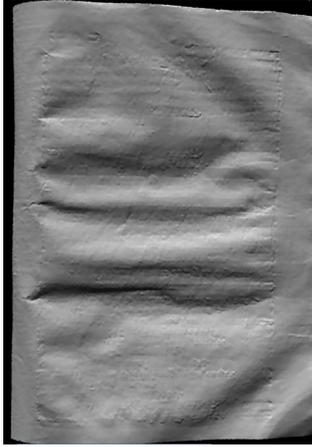


Fig. 8. Generated synthetic images and its defects such as convexo-concaves, fold and torn

In Fig. 9, both the original mesh and the semi-synthetic degraded image of a right page are given. We can see clearly illumination defects due to the convexo-concave defects, one side being darker and the other one brighter (see the bottom-left images in Fig. 9). The effect of light makes the synthetic images more realistic. Besides, the dents may appear in the ancient documents (e.g. when the wood-character stamps have strongly been pressed on the sheet of paper).

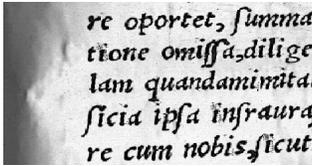
Original mesh



Generated image



Illumination defect



Dent

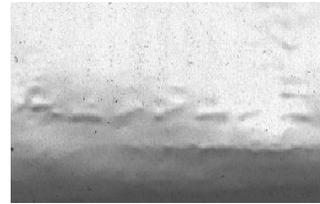
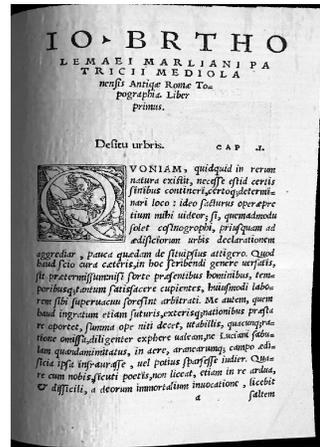
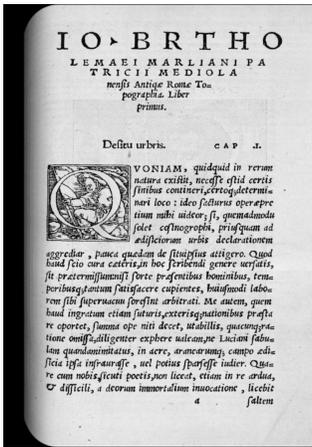


Fig. 9. Generated synthetic images and its defects such as illumination defect, dent

Fig. 10 provides a visual comparison of our model with the global distortion model of Kanungo [11] (see the left image of Fig. 10). Only the region near the document spine is distorted in Kanungo’s model. In addition, the non-linear optical point spread is modelled by applying a Gaussian blur function; thus, the result of the model is blurred.

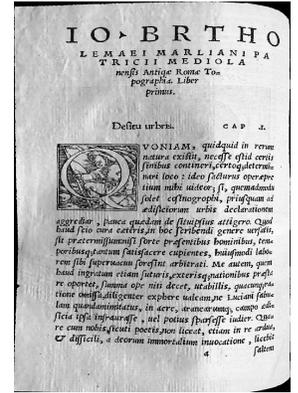


Result of Kanungo’s model Result of our method

Fig. 10. Example of Kanungo’s model in [11] (left) and result of our method (right)

Fig. 11 provides a visual comparison of our model with J. Liang’s model [14], which uses the reverse mapping process (left image in Fig. 11). In our result (the right image in Fig. 11), the real distortions such as dents, folds or irregular curves can

be simulated, which makes the curved surface more realistic.



Result of J. Liang’s model

Result of our method

Fig. 11. Example of J. Liang’s model in [14] (left) and result of our method (right)

B. Generating semi-synthetic images for performance evaluation

The music score CVC-MUSCIMA database, which contains handwritten music score images at both gray and binary levels, is largely used to test the robustness of different optical music recognition and removal staff line algorithms [17]. For the removal staff line competition of ICDAR 2013, a set of 6000 semi-synthetic versions of the 1000 images original images has been generated (among which 4000 for training and 2000 for test). These images are generated by using jointly the proposed method and the local character degradation model described in [1] (see in Fig. 13). They are challenging for staff removal algorithms, especially because the 3D degradations can distort the staff lines, making their detection and removal more difficult. Indeed, the addition of a white speck in the middle of a staff line can lead to disconnections within the lines, which could be tricky for most staff line extraction algorithms. Similarly, adding a dark speck connected to a staff line makes segmentation more difficult, as specks might be confused with musical symbols.



Fig. 12. An original image of the CVC-MUSCIMA database is used to generate synthetic images of music score removal staff line ICDAR 2013

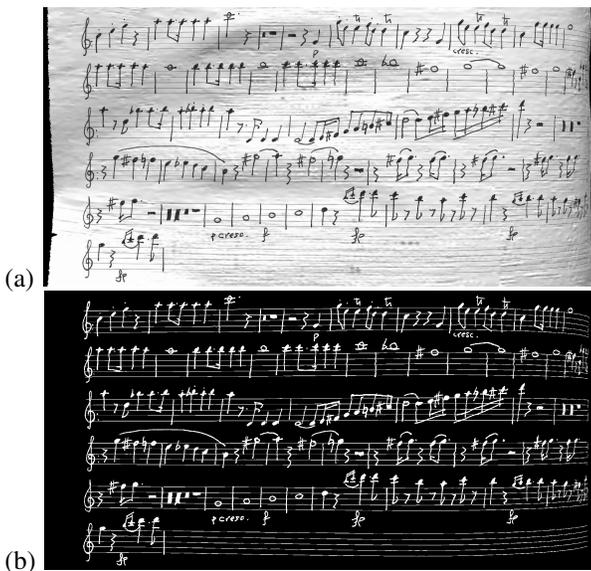


Fig. 13. Generated synthetic images of the original image in Fig. 12: (a) gray distorted image, (b) binary distorted image with staff lines

These two examples above show that our method can be applied in various contexts for evaluating the performance of systems. These experimental results (see more results on²) are encouraging, but they need to be formally validated as in [18]. We are currently collaborating with other researchers to carry out a series of tests on the different synthetic databases generated by our models and prove the effectiveness of our approach for enriching the training stage of different systems.

IV. CONCLUSION

In this paper, we present a method based on 3D meshes and texture coordinate generation for simulating geometric distortions in semi-synthetic images. This method is adapted to document images to deal with the unwrapping errors (i.e. stretching); therefore, the method can simulate real degradation better than the existing methods. There are three steps in this method. First, real documents are scanned to obtain document shapes (meshes). Then, we run (once per mesh) our new texture coordinate generation method, specially fitted for document images, and therefore more effectively in our context than most existing techniques. Finally, any 2D document image can be packed in the document shape by using an existing texture mapping technique. The Phong reflection model is used to improve the image quality. Our generation process enables to create synthetic images quickly; the only limitation in the variety of documents we can produce only lies in the number of 3D pages available (meshes obtained by 3D scanning of ancient document pages). The obtained images are very realistic for observed geometric distortions such as dents, folds, convexo-concaves, tears. The method is integrated with other degradation models in a system dedicated to a semi-synthetic old document image generation. This system can generate benchmark databases that will be further used to evaluate or compare the performances of different systems (e.g. the musical score competition at ICDAR 2013) and to enrich the training data in various contexts.

ACKNOWLEDGEMENT

Authors would like to thank Jérôme Charton and the SmithDR team for discussions on the 3D scan process and the texture mapping application. This research is done within the DIGIDOC project financed by the ANR (Agence Nationale de la Recherche).

REFERENCES

- [1] V. Kieu, M. Visani, N. Journet, J. P. Domenger, and R. Mullot, "A character degradation model for grayscale ancient document images," in *Proc. of the ICPR*, Tsukuba Science City, Japan, Nov. 2012, pp. 685–688.
- [2] H. S. Baird, "The state of the art of document image degradation modeling," in *In Proc. of 4 th IAPR International Workshop on Document Analysis Systems, Rio de Janeiro*, Rio de Janeiro, Brazil, 2000, pp. 1–16.
- [3] M. Delalandre, E. Valveny, T. Pridmore, and D. Karatzas, "Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems," *Int. J. Doc. Anal. Recognit.*, vol. 13, no. 3, pp. 187–207, Sep. 2010.
- [4] M. Mori, A. Suzuki, A. Shio, and S. Ohtsuka, "Generating new samples from handwritten numerals based on point correspondence," in *Proc. 7th Int. Workshop on Frontiers in Handwriting Recognition*, Amsterdam, Netherlands, 2000, pp. 281–290.
- [5] T. Varga and H. Bunke, "Effects of training set expansion in handwriting recognition using synthetic data," in *Proc. 11th Conf. of the Int. Graphonomics Society*. Scottsdale, AZ, USA: Citeseer, Nov. 2003, pp. 200–203.
- [6] —, "Generation of synthetic training data for an hmm-based handwriting recognition system," in *Proc. 7th ICDAR*, Edinburgh, Scotland, August 2003, pp. 618–622.
- [7] D. D. Jian Zhai, Liu Wenyin and Q. Li, "A line drawings degradation model for performance characterization," in *Proc. 7th ICDAR*, Edinburgh, Scotland, August 2003, pp. 1020–1024.
- [8] R. Loce and W. Lama, "Half-tone banding due to vibrations in a xerographic image bar printer," *Journal of Imaging Technology*, vol. 16, no. 1, pp. 6–11, 1990.
- [9] H. S. Baird, *Structured Document Image Analyse*. New York, USA: Springer-Verlag, 1992, ch. Document image defect models, pp. 546–556.
- [10] E. B. Smith, "Modeling image degradations for improving ocr," in *16th European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, August 2008, pp. 1–5.
- [11] T. Kanungo, R. M. Haralick, and I. Phillips, "Global and local document degradation models," in *Proc. of the ICDAR*, Tsukuba Science City, Japan, Oct. 1993, pp. 730–734.
- [12] T. Kanungo, R. Haralick, H. Baird, W. Stuezle, and D. Madigan, "A statistical, nonparametric methodology for document degradation model validation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1209 – 1223, 2000.
- [13] V. Rabeux, N. Journet, and P. Domenger, "Document recto-verso registration using a dynamic time warping algorithm," in *Proc. of the ICDAR*, Beijing, China, November, 2011, pp. 1230–1234.
- [14] J. Liang, D. DeMenthon, and D. S. Doermann, "Geometric rectification of camera-captured document images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 591–605, 2008.
- [15] A. Ulges, C. H. Lampert, and T. Breuel, "Document capture using stereo vision," in *In Proceedings of the ACM Symposium on Document Engineering*. ACM, 2004, pp. 198–200.
- [16] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, Jun. 1975.
- [17] A. Fornes, A. Dutta, A. Gordo, and J. Lladós, "The icdar 2011 music scores competition: Staff removal and writer identification," in *Proc. of ICDAR*, sept. 2011, pp. 1511–1515.
- [18] Y. Li, D. Lopresti, G. Nagy, and A. Tomkins, "Validation of image defect models for optical character recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 99–108, Feb. 1996.

²http://www.labri.fr/perso/vkieu/content/Databases/music_removalStaff.html