Objective Quality Measurement for Geometric Document Image Restoration

Christoph H. Lampert and Thomas M. Breuel

German Research Center for Artificial Intelligence (DFKI) GmbH
D-67663 Kaiserslautern, Germany, {chl, tmb}@iupr.net

1 Introduction

Many algorithms to remove distortion from document images have be proposed in recent years, but so far there is no reliable method for comparing their performance. In this paper we propose a collection of methods to measure the quality of such restoration algorithms for document image which show a non-linear distortion due to perspective or page curl.

For the result from these measurement to be meaningful, a common data set of ground truth is required. We therefore started with the buildup of a document image database that is meant to serve as a common data basis for all kinds of restoration from images of $3D$-shaped document. The long term goal would be to establish this database and following extensions in the area of document image dewarping as an as fruitful and indispensable tool as e.g. the NIST database is for OCR, or the Caltech database is for object and face recognition.

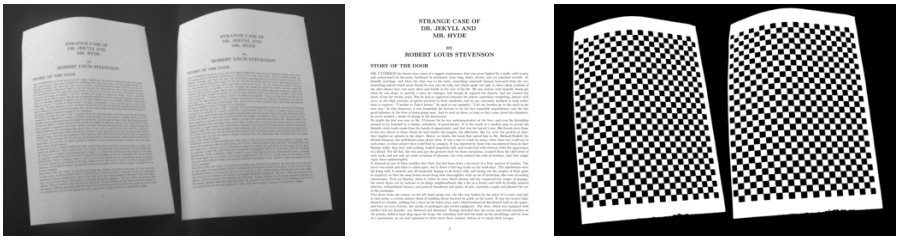
2 A Database of Document Images and Ground Truth

So far, the different publishes approaches use different input images to measure their performance, and the input material usually is adapted to the capabilities and target of the system to be tested. Our goal is to provide a common basis for many different approaches instead. We therefore decided to build up a new collection of document images with enough additional information to support different methods. Namely, for each document the database contains

a) a stereo image pair showing the document as it lies in 3D
b) a flat binary image of the document as ground truth,
c) a stereo image pair showing geometric ground truth in same shape as a)
d) an ASCII file of textual ground truth.

Examples of the image components are shown in Figure 1(a)–1(c). In the rest of this section, we describe the components in the order of their generation.

Text Ground Truth (ASCII GT): We used a selection of texts from the Project Gutenberg collection for our experiments. These texts are under a free license, thus we avoid copyright problem when publishing the database. We only include English texts, because this allows a simple ASCII encoding scheme, without the need for special characters.



(a) Warped document image (b) Flat GT (c) Geometric GT

Fig. 1. Typical entry of the document database.

Binary Ground Truth Image (flat GT): The ASCII GT files are typeset using LaTeX, using different layouts, fonts, and font sizes. The output is converted into binary images of size 2400×3000 pixels at a resolution of 300 dpi. The values were chosen to make the resulting image fit onto A4 as well as onto Letter paper.

Document Images and Geometric Ground Truth (geometric GT): The flat GT images are printed in dark blue color on a color laser printer. Additionally, a regular yellow checkers pattern is printed onto each page. The idea of this is to later be able to separate the color channels again, without one pattern disturbing the other. The resulting printouts are bent into in different shaped, typically like a book that lies open on a table. Of those, stereo image pairs are taken using a calibrated rig, allowing to also test algorithms that rely on stereo vision for shape reconstruction (e.g. [1, 2]). The captures are performed with different camera position, viewing angle and lighting conditions. We also vary the image resolution between approximately 50 and 300 dpi.

From the captured images, the blue color channel is separated which almost only shows the checkers pattern. It is binarized and morphologically cleaned, to serve as basis for the geometric GT measurements (see Section 3).

The remaining red and green channels show only the printed text. They are not binarized, but combined and converted to an 8 bit gray scale image. That way, also algorithms are supported that rely on shading information, e.g. [3].

3 Measure for Objective Quality Evaluation

Which measure of 'quality' is the right one to judge the performance of a restoration algorithm depends mainly on the application. In the following, we formalize some of those measures to be used in conjunction with the proposed document database. The aim is to in future allow a simple comparison between different methods and different publications.

Visual Inspection: The simplest way for a human to compare restoration systems is by a visual comparison of their output with their input. Obviously, judging such image pairs is a very informal and subjective way of evaluation.

For usage in scientific publications, it has a number of drawbacks, in particular that there is no objective numerical scale to which the results could be reduced. Therefore, different methods can only be compared by presenting images for all of them. In a printed publication, a selection has to be taken of which images to present, which introduces an additional bias.

Visual inspection remains a useful tool for the researcher to judge his or her own work, but in general, it cannot be relied on to measure the results of different approaches between different publications.

OCR Improvement: Many image restoration systems are designed to serve as a preprocessing step for the extraction of information from the document images, usually the text. The restoration step then is targeted at improving the OCR performance, not at trying to generate output that resembles some unknown original or to look pleasing to a human.

To benchmark a restoration step in this setup, we measure the increase of recognition rate that an OCR system achieves between working on the original image and on the restored version. For this we measure the *word accuracy* that is calculated as

$$accuracy = \frac{n - e}{n}, \tag{1}$$

where n is the number of words in the text GT and e the word-based edit distance between the text GT and the OCR system's output.

A problem of using OCR improvements as a quality measure is that the performance of the restoration system is always measured in conjunction with the performance of the OCR system itself. Therefore, measurements have to rely on the same OCR engine, or the results are not comparable with each other in a meaningful way.

Reduction of Geometric Distortion: When a system targets at removing geometric distortion that was caused by perspective or page curl, i.e. to "flatten" an image, then an obvious quality measure is to determine how much of the distortion remains after the restoration.

For arbitrary input images this is not possible, but for our database images it is, because we included the geometric GT. Our method is similar to [4] and works in the following way: from the distorted document image, the dewarping parameters are calculated. The subsequent dewarping itself is then performed not with the document as input, but with the geometric GT image.

The result is a more or less regular checkers pattern of 24×30 squares. We extract its interior corner points and determine the best linear transformation of those to a grid of unit square cells, allowing translation, rotation and scaling in x and y-axis. The result gives us a correspondence between transformed point centers $\boldsymbol{p}_i$ and grid points $\boldsymbol{g}_i$, and for each such pair we can calculate the displacement $d_i = ||\boldsymbol{p}_i - \boldsymbol{g}_i||$. Useful quantities to judge the remaining distortion are then the *mean and variance of the displacement vector*:

$$mean = \frac{1}{23 \cdot 29} \sum_i d_i \qquad var = \frac{1}{23 \cdot 29} \sum_i (d_i - mean)^2. \tag{2}$$

The transformation step makes the results independent under translation, rotation and scaling of the output picture. This is a desirable property, because if we just aim at measuring how "flat" the restored image has become, those planar transformations should not play a role.

Pixel Error (PSNR): The strictest method to test how well an algorithm restores an original document from a distorted image is by a pixel wise comparison of its output with the flat GT image. As distance measure we can then use the L^2-distance between the two images or equivalently the *Peak Signal to Noise Ratio (PSNR)*:

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{\sum_{x,y} \mid g(x,y) - r(x,y) \mid^2} \tag{3}$$

where (x, y) ranges over all image positions and $g(x, y)$ and $r(x, y)$ are the image intensities of the flat GT image respectively the restored image. The PSNR is measured in dB with larger values meaning a better reconstruction.

Again we allow a registration step before calculating the distance. In the current setup, we are interested in a as good restoration of the actual image as possible. We therefore allow only translation and isotropic scaling, because in the presence of text lines a restoration system should be able to detect and correct all other parameters, e.g. rotation, by itself.

4 Conclusion

We have presented different algorithms to judge the performance of document image restoration algorithms. Depending on the application, visual inspection, OCR accuracy, remaining geometric distortion or pixel-by-pixel similarity can be the most suitable measure of quality.

To provide a common basis for independent tests, we have also presented a database for geometricly distorted document images that is currently being built up at the German Research Center for Artificial Intelligence. We will finish the database until December 2005, and then make it publicly available.

References

1. Yamashita, A., Kawarago, A., Kaneko, T., Miura, K.T.: Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system. In: International Conference on Pattern Recognition. (2004) 482–485
2. Ulges, A., Lampert, C.H., Breuel, T.M.: Document capture using stereo vision. In: ACM Symposium on Document Engineering. (2004) 198–200
3. Zhang, Z., Tan, C.L., Fan, L.: Restoration of curved document images through 3D shape modeling. In: Int. Conf. on Computer Vision and Pattern Recognition. (2004) 10–15
4. Brown, M.S., Seales, W.B.: Document restoration using 3d shape: A general deskewing algorithm for arbitrarily warped documents. In: International Conference on Computer Vision. (2001) 367–374