Local multiscale blur estimation based on toggle mapping for sharp region extraction

Luc Gillibert[1], Théodore Chabardès[1], Beatriz Marcotegui[1]

[*]MINES ParisTech, PSL Research University, CMM - Centre for Mathematical Morphology, 35 rue Saint Honoré - Fontainebleau, France
[*]{luc.gillibert,theodore.chabardes,beatriz.marcotegui}@mines-paristech.fr

Abstract: In this paper a multiscale local blur estimation is proposed based on the existing local focus measure that combines gradient and toggle mapping. This method evaluates the quality of images regardless of their content (not in an autofocus context) and can predict OCR accuracy based on local blur. The resulting approach outperforms state of the art blur detection methods. Quantitative results are given on DIQA database. Moreover we demonstrate its usefulness for extracting a region of interest from partially blurry images. Results are shown on images acquired by a project devoted to smartphone based text extraction for visually impaired people. In this context sharp region extraction is essential since it allows warning the users when their picture is unusable. Moreover it saves computing time.

Keywords: Blur estimation, Toggle mapping, DIQA database, LINX database

1. Introduction

Mobile phones offer a number of increasing capabilities, especially high resolution photography and high computational power that can be used for a new range of image processing applications. Text is omnipresent and brings essential information in our daily life. Visual impairment deprives a significant percentage of the population of this important information. The goal of the LINX project is to develop an automatic reader on smartphone to help visually impaired people in their daily life, making textual information available to them.

In order to reach this goal, several challenging tasks need to be solved: text detection, text extraction and Optical Character Recognition (OCR). Each of these tasks is time consuming and must not be executed on useless pixels. Visually impaired people cannot guarantee the quality of the image they acquire. Images can be degraded by bad lighting conditions, shadows, out-of-focus blur, motion blur, perspective problems or noise. Moreover the use of handheld devices aggravates the blur problem. Therefore extracting regions containing sharp enough structures is an extremely useful functionality for this kind of application.

For several decades, Image Quality Assessment (IQA) techniques have attracted great attention from the scientific community. These techniques can be classified into three categories:

- Full Reference (FR): require the original image as a reference. The most popular FR objective metrics are the mean square error (MSE), Peak Signal to Noise Ratio (PSNR) and cross-correlation between both images [23]. The structural similarity index (SSI) [31], attempts to quantify the visibility of errors based on the degradation of structural information. The use of

FR IQA is limited because the reference image is often unavailable.

- Reduced reference (RR): are designed to predict the quality of degraded images with only partial information of them. The comparison is based on extracted features instead of on the whole image. Examples of these techniques can be found in [16] where the comparison is based on the statistical modeling of the discrete cosine transform (DCT) and [32] based on wavelets. The original image is still required in order to extract the features to be compared.
- No-reference or blind techniques: the measure is computed on the degraded image itself. This is the most challenging task, in which the quality measure relies exclusively on a single acquired image and corresponds to our application context (LINX project).

A wide variety of approaches for NR blur estimation exist. They operate either in the spatial or in the frequency domain. Among the spatial techniques the simplest metrics rely on gradient or laplacian computation. Indeed, the blur reduces the contrast of contours and tends to weaken the gradient. The higher the gradient, the sharper the image. Thus, the sum of thresholded gradient or the squared gradient [11, 26], Brenner function [36], multiscale gradient [6] and laplacian [19] are examples of these techniques. Other authors have considered the variance or other histogram measures such as entropy [12, 27, 20], assuming that a focused image conveys more information than an unfocused one. Finally, some techniques are based on edge width estimation, after an edge detection step [18, 4]. Concerning frequential approaches, they rely on the assumption that a defocused image results from the convolution of the image which produces a decrease in the high frequencies. Thus, the sum of the spectrum components above a given cut frequency [10], the ratio between the high frequency coefficients over low frequency ones or the distribution of null coefficients (blurred images are likely to have all of their high frequency coefficients set to zero) have been used as focus measures. Discrete Cosine Transform (DCT) [17, 25] and wavelet transform [34, 33] have been used for this purpose. The correlation of wavelet coefficients over scales has also been used [7, 30]. In [2, 1] a sharpness index is derived from the phase component of the Fourier Transform, based on the fact that edges correspond to points of maximum phase congruence.

Most previous approaches have been established for autofocus purposes, where the quality of an image can be judged compared to the quality of other images with similar content. Consequently, they do not fit our requirements. In order to lift this restriction some local approaches have been proposed, applying the method to a given neighborhood [22], that needs to be defined or on the result of an edge detection approach [4, 21].

Recently, the specific problem of document image quality is becoming a topic of intense research. Several papers attempt to predict the OCR accuracy when applied to an image, based on IQA measures of the given image [3, 21, 35]. The interesting DIQA dataset, described in section 3.2, has been introduced in this context [15]. We will benchmark our approach on it.

Lately [13] proposed a local blur estimation based on Fuzzy-C-Means clustering approach after a binarization step. They combine this local estimation with the character size in a learning based strategy, in order to predict the OCR accuracy. The learning approach suits a document oriented application but it is not convenient for our LINX project.

Image quality assessment remains an open problem, in particular for generic scene-text images such as those of our LINX project. This paper is an extended version of [5]. In the original paper, a local blur estimator that does not require a prior contour detection, nor a neighborhood definition was introduced. This paper introduces the following novelties:

- A parameters sensitivity study.

- Multiscale extension of the original method in order to deal with documents with variable font size.
- The selection of a sharp region of interest (possibly empty if the whole image is blur) in order to avoid processing doomed to failure.

The paper is organized as follows:

Section 2 summarizes the toggle mapping based blur estimator approach. The sensitivity of this blur estimator to its parameters is studied in section 3 and some experimental results are given. The performance of the approach is evaluated on DIQA dataset.

In section 4 a region of interest is extracted from the local blur information. Then, in section 5, the extraction of the region of interest is generalized for multiscale blur detection.

Section 6 discusses the choice of the scales used and shows relevant sharp regions extracted from LINX database with variable font sizes. Finally section 7 concludes this work.

2. Toggle mapping based blur estimation

Toggle mapping (TM) operator was introduced in [14]. From an image I, two transformations are computed: the dilation, $\delta_B(I)$ and the erosion $\epsilon_B(I)$, the maximum and the minimum value within the structuring element B. Equation 1 describes the TM operator:

$$TM_B(I) = \begin{cases} \delta_B(I) & if\ \delta_B(I) - I\ <\ I - \epsilon_B(I) \\ \epsilon_B(I) & otherwise \end{cases} \tag{1}$$

Thus, each pixel is replaced by one of these two transformations, selecting the one that is closest to the original pixel value. This principle was generalized in [28], with the use of other transformations involved in the toggle process.

TM was designed as a contrast enhancement operator. It was also used as scene text segmentation tool in [8]: pixels replaced by the dilation are set to 1 and those replaced by the erosion are set to 0.

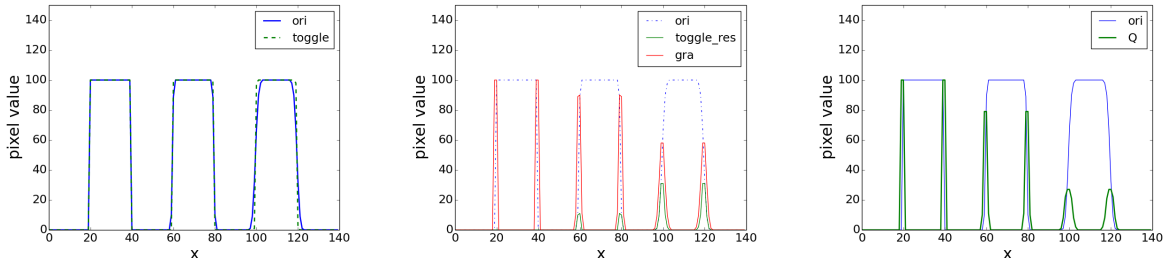
A local blur estimation based on TM is introduced in [5].

We define the residual image as the absolute difference between the original image and the TM:

$$TMR_B(I) = \begin{cases} \delta_B(I) - I & if\ \delta_B(I) - I\ <\ I - \epsilon_B(I) \\ I - \epsilon_B(I) & otherwise \end{cases} \tag{2}$$

A blur boundary leads to a high residue, and this residue increases with the size of the structuring element involved in the TM. This is true up to a TM size equivalent to the blur edge width. A blur estimation could then be built on the evolution of the TM residues, with a series of structuring elements of increasing size. However a sharp boundary and a homogeneous region would have the same series of low TM residue value. In order to distinguish between these two situations we compare these residues with the gradient value. The morphological gradient is defined as the difference between the maximum and the minimum value in the neighborhood of a pixel ($\delta_B(f) - \epsilon_B(f)$), while the TM replaces each pixel with the closest extremum in the neighborhood, that is either the maximum or the minimum. Thus, by definition, TMR is lower than the gradient. However if the TM uses a larger structuring element its residue can reach the gradient value, in which case, the pixel is classified as blur.

Figure 1 shows a simple 1D example with three objects, from sharp on the left to blurry on the right. Figure 1(a) illustrates the toggle mapping operator in green dashed line. The corresponding



(a) original and TM (b) TM residue and gradient. (c) I_Q

Fig. 1. *Intermediate steps of local quality estimation on three 1D objects. (a) original objects, from sharp on the left to blurry on the right. (b) TMR increases with blur while the gradient decreases. (c) The final I_Q, the difference between TMR and the gradient.*

residue is shown in green and the gradient in red in figure 1(b). We can observe that the residue of the toggle mapping increases with the blurriness while the gradient decreases with it. Figure 1(c) shows the I_Q, the quality measure defined as the difference between the gradient and the toggle mapping residue: $I_Q = max\,(0, gradient_{B_N}(I) - TMR_{B_M}(I))$, with B_N the structuring element B of size N and $M > N$. As expected, contours are ranked according to their contrast and sharpness, while homogeneous regions have a small I_Q value.

Figure 2 shows the flow chart of the algorithm. First, a bilateral filter [29] is applied in order to get rid of noise. Then I_Q is computed as the positive difference between the gradient and the TMR. It is a local measure. It can be thresholded for identifying sharp regions. Small connected components of the thresholded I_Q can be removed with an area opening. Other morphological filters will be introduced in section 4. If a global quality score, Q_{score}, is required for the whole image it can be deduced by averaging the thresholded I_Q image, taking into account only the values over the threshold. Thus the quality measure does not depend on the quantity of text in the images. As shown in section 3, the removal of small regions has a low impact on the Q_{score}. However it is important for sharp region selection. Figure 3 shows the impact of the bilateral filter on the process. Black pixels correspond to pixels of I_Q under the threshold. We can observe a much cleaner result after the filtering. This will be an essential aspect for the sharp region selection. Moreover, the quantitative contribution of this filter and its parametrization will be studied in section 3.

3. Results and parameters sensitivity

3.1. Qualitative results

A good correspondence between observed quality and our score on LINX database has been noted. The quality of the extracted regions of section 6 will confirm this point.

3.2. Quantitative results

Quantitative results are given on DIQA database, an interesting dataset recently published [15]. It is composed of 175 document images acquired by a mobile phone. Images are directly acquired by a camera, so the blur is not simulated. 25 documents have been acquired, 6 to 8 times each, in different conditions: from perfect focus to completely out-of-focus images. The quality of each image is given by the OCR accuracy on the processed image. Three different OCRs have

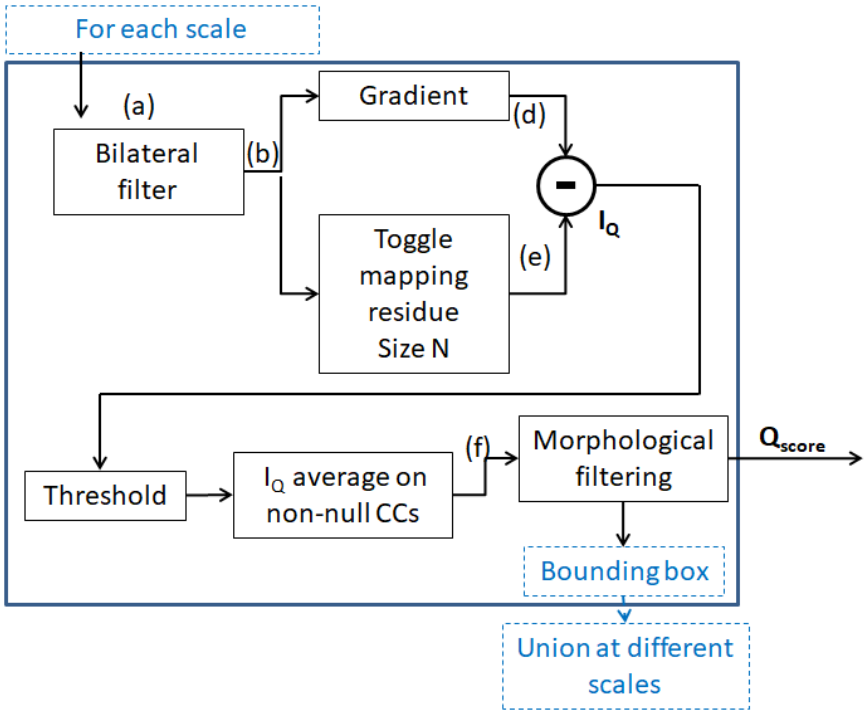
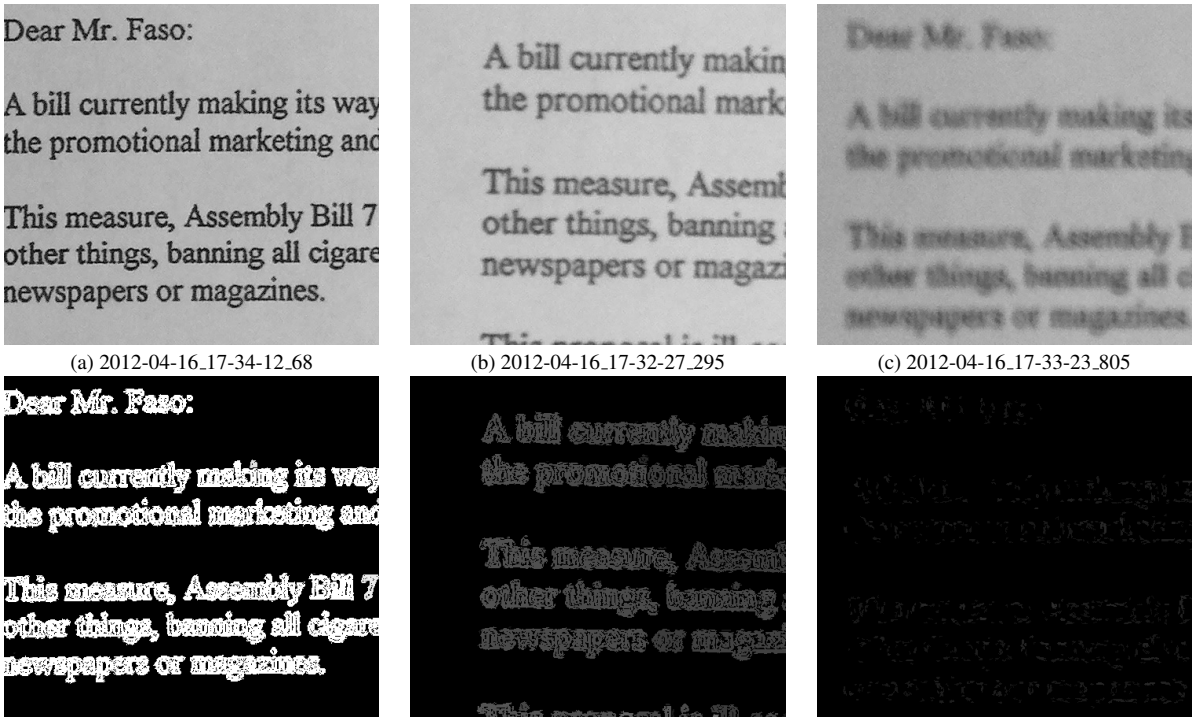


Fig. 2. *Algorithm flowchart.*

(a) original image (b) $Q_{score} = 17$ (c) $Q_{score} = 32$

(d) original image (e) $Q_{score} = 9$ (f) $Q_{score} = 17$

Fig. 3. *Filter influence in I_Q computation. Left column: original image. Middle column: Thresholded I_Q computed from original image. Right column:Thresholded I_Q from bilateral filtered image.*



(a) 2012-04-16_17-34-12_68 (b) 2012-04-16_17-32-27_295 (c) 2012-04-16_17-33-23_805

(d) OCR accuracy = 0.97; $Q_{score} = 55$ (e) OCR accuracy = 0.88; $Q_{score} = 14$ (f) OCR accuracy = 0.24; $Q_{score} = 5$

Fig. 4. *Three DIQA examples. A sharp image (left column), intermediate blur (middle column) and a blur image (right column). First row: original images. Second row: I_Q image and the respective OCR accuracy and Q_{score}.*

been used: Tesseract, FineReader and Omni. The overall accuracy of these OCRs on the whole database are:

	Tesseract	FineReader	Omni
OCR accuracies:	0.53	0.76	0.72

Figure 4 shows 3 crops from this dataset, with an OCR accuracy ranging from 0.97 to 0.24 for FineReader. Our quality score for these images ranges from 55 to 5. Spearman correlation is commonly used to assess the coherence of two variables without taking into account their precise values. It compares their ranking indexes instead of the variables themselves. In our example, the ranking based on the OCR accuracy or on our quality score is the same (1.- (a); 2.-(b); 3.-(c)). Thus, for this simple example with only three images, our quality score is perfectly correlated with the OCR accuracy: the Spearman correlation between them is then equal to 1.

Some works [15, 24] have tried to estimate the OCR accuracy based on some features extracted from the image. The score used for comparing different methods is the median of the Spearman correlation applied to each document set. The scores reported are over 90%. We think that the proposed score overestimates the quality of the OCR accuracy prediction. Indeed, if the Spearman correlation is computed for each document set, the content of images to be compared is similar. The context is then close to an autofocus situation which is much easier than comparing the quality of images with different content. This problem is pointed out in [24]: a failure is reported for an image with a low focus measure whereas the OCR accuracy is over 90%, provoked by the large white space in the page. Another reported issue concerns an image with a high focus measure but a low OCR accuracy. The huge headline leads to a high focus measure while most of the text is

Approach	FineReader	Tesseract	Omni
Combination of classical blur descriptors [24]	0.6467	-	-
Phase coherence [1]	0.5569	0.6318	0.5174
Our approach	0.8426	0.9016	0.7277

Table 1 Spearman correlation between OCR accuracy and quality score obtained in the same conditions by different state of the art approaches.

out-of-focus. We address these situations in this paper.

Moreover, as stated in [24], by reporting the median value, outlier classes in which the methods might not perform well are disregarded. They propose to compute the Spearman correlation directly for the 175 images of the database. We adopt this proposal that we find more reliable. The authors of this paper have tested all combinations of 24 classical blur descriptors. The best results were obtained with the combination of 4 of them: gaussian derivative, gradient energy, squared gradient and histogram range. They report a Spearman correlation of 0.6467 for the whole database as a single set for FineReader OCR. In the same conditions we obtain a Spearman correlation of 0.8426. For benchmarking purposes table 1 summarizes the Spearman correlation obtained in the same conditions with different state of the art approaches.

A higher Spearman correlation indicates the method's ability to rank images according to the expected OCR accuracy. Therefore, we can state that our method outperforms state of the art approaches for OCR accuracy estimation.

3.3. Parameters sensitivity

The parameters have been manually tuned for Tesseract accuracy estimation, as Tesseract has been chosen in the framework of our LINX project. We will now test the sensitivity to each parameter by changing its value without changing the value of the other parameters. If not specified otherwise, parameters used are:

- Structuring element: square.
- Bilateral filter: $\sigma_{spatial} = 2$ and $\sigma_{gray} = 20$. The sensitivity to these parameters is studied on table 2 and table 3.
- Gradient size $B_1 = 1$ (3×3 pixels).
- TM size $B_2 = 2$ (5×5 pixels).
- Binary threshold on I_Q: 3. The sensitivity to this parameter is studied on table 4.
- Area opening of size 5. The sensitivity to this parameter is studied on table 5.

The two parameters of the bilateral filter can vary in a wide range without any significant impact on the Spearman correlation. As shown on table 2, the use of a bilateral filter significantly improves the scores. Furthermore $\sigma_{spatial}$ can vary from 2 to 5 with less than a 1% variation in the Spearman correlation for the three OCR. Table 3, studies the sensitivity to σ_{gray} parameter. It can vary from 15 to 35 with less than a 1% variation in the Spearman correlation for Tesseract.

As shown on table 4, the sensitivity to the threshold on the I_Q is low if the threshold is greater than 2. Moreover the sensitivity to the size of the area threshold on the I_Q is very low. As shown

on table 5, the size of the area opening on the I_Q can vary from 2 to 11 with less than one per thousand variation in the Spearman correlation for Tesseract and Omni.

$\sigma_{spatial}$	FineReader	Tesseract	Omni
0	0.7410	0.7110	0.6219
1	0.8213	0.82	0.6979
2	0.8426	0.9016	0.7277
3	0.8399	0.902	0.7266
4	0.8396	0.902	0.7263
5	0.8395	0.9024	0.7262

Table 2 Sensitivity to the $\sigma_{spatial}$ used for the bilateral filter based on Spearman correlation between OCR accuracy and quality score.

σ_{gray}	FineReader	Tesseract	Omni
5	0.7476	0.7171	0.631
10	0.8428	0.8586	0.7155
15	0.8532	0.9006	0.7305
20	0.8426	0.9016	0.7277
25	0.8374	0.9003	0.7247
30	0.836	0.8986	0.7215
35	0.8344	0.8961	0.7165

Table 3 Sensitivity to the σ_{gray} used for the bilateral filter. The Spearman correlation between OCR accuracy and quality score is given for FineReader, Tesseract and Omni.

Binary threshold	FineReader	Tesseract	Omni
1	0.8254	0.8114	0.6942
2	0.8499	0.8969	0.7271
3	0.8426	0.9016	0.7277
4	0.8385	0.9015	0.7276

Table 4 Sensitivity to the binary threshold on the I_Q based on Spearman correlation between OCR accuracy and quality score.

The sensitivity to these parameters is low. For all the tested parameters variation, the results of the the Spearman correlation obtained on the DIQA image database are higher than the 0.6467 reported in the literature. Therefore our method outperforms the state of the art for OCR accuracy estimation.

If the goal of the approach is to estimate a global Q_{score}, then the area opening can be ignored, because the sensitivity to the size of the area opening on the threshold I_Q is very low. But this area opening is very useful for the extraction of a region of interest in section 4.

4. Region of interest on heterogenous blur

If the image is heteregeneously blurry, the local blur measure I_Q, described in section 2, allows to extract a region sharp enough for the OCR. If there is no such sharp region, the whole image is

Area threshold	FineReader	Tesseract	Omni
2	0.844	0.9011	0.7277
3	0.8429	0.9012	0.7278
4	0.8428	0.9012	0.7277
5	0.8426	0.9016	0.7277
6	0.8416	0.9014	0.7272
7	0.8405	0.9013	0.727
8	0.84	0.9015	0.7275
9	0.8395	0.9014	0.7275
10	0.8392	0.9013	0.727
11	0.8387	0.9011	0.7266

Table 5 Sensitivity to the area opening size on the thresholded I_Q based on Spearman correlation between OCR accuracy and quality score.

rejected.

A naive approach would be to select a rectangular region of interest containing all the non-zero pixels of this thresholded I_Q. This naive approach will fail in some cases. For example, spurious detection in noisy regions. Another problem with the thresholded I_Q is the presence of thin non-textual objects. Typically: borders of paper sheets, isolated straight lines, straight shadows. They are irrelevant for text extraction. They can be ignored, reducing significantly the area of the extracted region of interest.

Thin objects and small isolated sharp regions on the thresholded I_Q can be removed before the extraction of the final rectangular region of interest with some simple morphological filters. This filtering process is the following:

1. $F_1(I) = \epsilon_{B_1}(\delta_{B_1}(I))$ (closing)
2. $F_2(I) = \delta_{B_2}(\epsilon_{B_2}(F_1(I)))$ (opening of the closed image)
3. $F_3(I) = \gamma^S_{area}(F_2(I))$ (area opening of $F_2(I)$)

Where γ^S_{area} is the area opening of surface S. All connected components of surface smaller than S are removed.

The first morphological closing, F_1, is used for merging small sharp regions and for filling small objects. Characters and lines are sharp on their edges, but have a low I_Q value on their uniform regions.

Thin objects and small isolated objects are removed with the morphological opening, F_2, and the area opening F_3. The edges of characters are unaffected by this filter because the morphological closing has filled characters, making them large. This step of the process is illustrated on figure 6(c).

After the morphological filters, the bounding box including all remaining pixels is our region of interest.

5. Multiscale approach

Large characters may be blur but still readable (figure 5). The method introduced in section 3 will reject them. This problem is not present on the DIQA database, because the character size is very

Fig. 5. *Example with large characters in a blurry context, but still readable. Left column: original images. Right column: miss-detection of readable characters.*

consistent in the database, but is present on the LINX database.

For solving this problem, we introduce a multiscale approach. The input image (scale $s = 1$) is processed at different scales s. At scale s the image is decimated by a factor 2^s in each coordinate, by using a bilinear filter. The whole process described in section 4 is applied to each scale and the union of resulting bounding boxes is computed (see figure 2. Blue dashed line boxes have been added.

A full example is shown on figure 6 for scale 1 and $\frac{1}{32}$. At scale $\frac{1}{32}$, the morphological filter removes the border of the table, on the right of the image, but the cover of the disk is not suppressed.

At each scale a new bounding box is produced, containing readable text for this scale. The final bounding box is the box containing the union of the bounding boxes of all the scales.

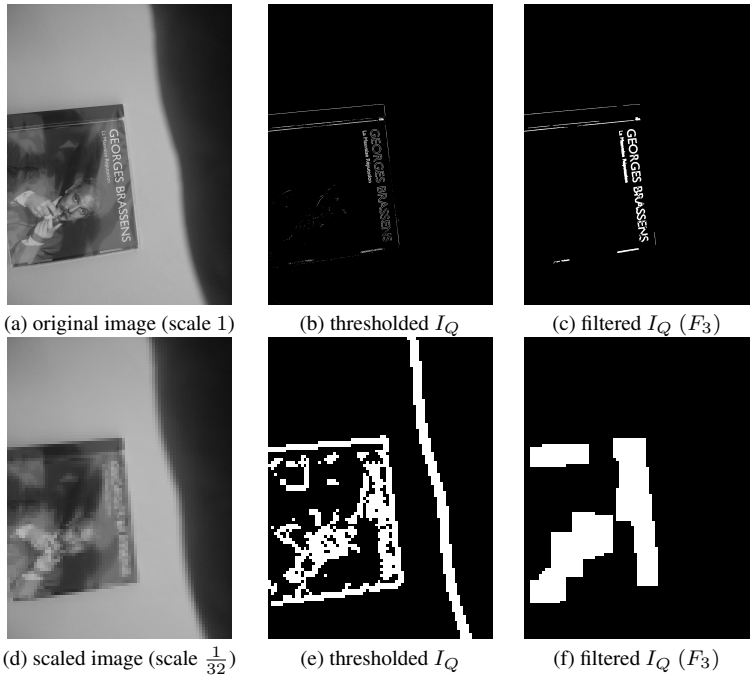
The multiscale approach solves the problem of large blurry characters. The additional computational cost is low because the process is applied to reduced-size images. More details on the computational time are given in table 7 of section 6.

6. Experimental results for the region of interest

6.1. Choice of the parameters

The muti-scale approach can use any arbitrary set of scales. By construction, the largest windows are obtained with the full set of scales ($\{1, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$). This result is considered as a reference and the Jaccard index (the area of the intersection divided by the area of the union) in order to study the influence of the scale set selection. Table 6 summarizes the result on the whole LINX database. We can observe that:

- a single scale misses 43% of the surface detected by the full set of scales.
- combining only two scales (1 and $\frac{1}{16}$) detects 89% of the surface obtained by the full set of scales.



(a) original image (scale 1) (b) thresholded I_Q (c) filtered I_Q (F_3)

(d) scaled image (scale $\frac{1}{32}$) (e) thresholded I_Q (f) filtered I_Q (F_3)

Fig. 6. *Example of the extraction of the region of interest (scale* 1 *et* $\frac{1}{32}$*). (a) is the input image (scale 1). (b) is the thresholded* I_Q*. This image is closed (*F_1*), then open with a morphological opening (*F_2*) and an area opening (*F_3*) in (d). The same process is repeated at scale* $\frac{1}{32}$ *(d,e,f). The pixels left in (c) and (f) are used for the extraction of the region of interest.*

Figure 7 illustrates qualitative results of these differences. First column illustrates the single scale results: significant text is outside the detected bounding box. Second column shows the result of combining two scales (1 and $\frac{1}{16}$): bounding boxes include all the significant text. Third column shows the full set of scales result, leading to very similar results compared with the previous one.

Experimental results on the LINX dataset show that two conveniently chosen scales, in our case $\{1, \frac{1}{16}\}$, lead to a good trade-off between quality and computing time.

The parameters for the bilateral filter, the TM operator and the estimation of the I_Q do not depend on the scale. But the size of the morphological operations used for filtering unwanted objects in the final crop depends on it. For scale 1 the operations are the following: Closing with a square structuring element of size 5, opening with a square structuring element of size 1 and area

Set of scales	$\frac{S(Intersection)}{S(Union)}$
$\{1\}$	0.57
$\{1, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$	1.00
$\{1, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$	0.99
$\{1, \frac{1}{8}, \frac{1}{16}\}$	0.95
$\{1, \frac{1}{16}, \frac{1}{32}\}$	0.94
$\{1, \frac{1}{16}\}$	0.89
$\{1, \frac{1}{32}\}$	0.85

Table 6 Comparison between reference results (scales $\{1, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$) and various combinations of scales.

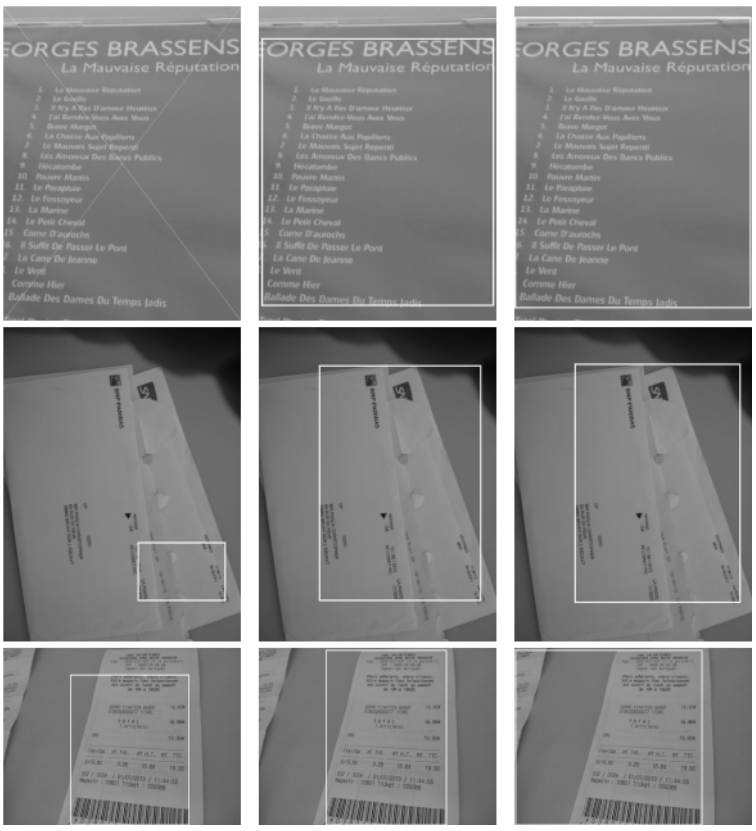


Fig. 7. *Comparison between different scale sets on a subset of the LINX database. Left column: monoscale approach. Middle column: multiscale approach with scales* $\{1, \frac{1}{16}\}$. *Right column: multiscale approach with the full set of scales* $\{1, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$.

opening with a surface of 500.

At smaller scales, the closing tends to merge different structures that the subsequent opening is not able to remove. For this reason, we have heuristically reduced the closing size while increasing the opening size. Due to the increased size of the morphological opening, the area opening is no longer critical. Hence, the operations are the following: closing with a square structuring element of size 2, opening with a square structuring element of size 3 and area opening with a surface of 50.

The processing time with the multiscale approach is very close to the processing time with the monoscale approach, as shown on table 7. The additional cost for the multiscale approach with respect to the simple scale is 1.6%.

The processing time is given in seconds for the whole LINX database: 160 images of size 3264×2448 (8 mega-pixels). The processing is done on an Intel Core i7-3800 (Sandy Bridge-E C2). The code is written in C++, based on the SMIL library [9], and uses only one core. The total processing time for the whole LINX database is 95.07 seconds. This time can be decomposed in three: 27 seconds for the bilateral filter, 28 seconds for the scaling, the toggle mapping and the morphological operations, and 40 seconds for the area opening. The average processing time for one image is 0.59 seconds, but this time can be reduced. The current algorithms are not parallelized and not fully vectorized. There is still room for optimizations. In particular, area opening seems to be a good candidate for optimization.

Set of scales	Processing time
$\{1\}$	93.53 seconds
$\{1, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$	115.43 seconds
$\{1, \frac{1}{16}\}$	95.07 seconds

Table 7 Processing time in seconds for the whole LINX database (160 images, 8 mega-pixels each). The processing is done on an Intel Core i7-3800 (Sandy Bridge-E C2).

6.2. Qualitative results

Qualitative results on LINX database are given in figure 8. Convincing results are obtained. The uniform background is correctly rejected (images 1, 2, 7-9 images in raster order), readable text is selected in the upper part of images 4 and 6 while blurry one (in the lower part) is rejected, the wall in image 5 is selected because of the details on it (cables and wall socket) and image 3 is rejected, as there is no sharp region on it.

Fig. 8. *Example of LINX images with extracted region of interest.*

7. Conclusions and perspectives

A method for selecting a region of interest for characters extraction and recognition has been introduced in this paper. It reduces the data to be processed and therefore the time and the energy needed for achieving these tasks. This will lead to a more reactive and efficient application on

smartphones.

Qualitative results on LINX examples show that blurry characters that are large enough for being read by the OCR, despite the blur, are correctly selected by the algorithm. The regions of interest are relevant and ignore large uniform or blurry regions of the images.

From a quantitative perspective, the quality indexes obtained are correlated with the OCR accuracy. On the DIQA images, the Spearman correlation reveals a good correspondence between the Q_{score} and the OCR accuracy.

For FineReader, a Spearman correlation of 0.8426 is reached, much better than the best score reported in the literature (0.6467) [24], combining classical blur descriptors. For Tesseract the results are even better, with a Spearman correlation of 0.9016, to be compared to the 0.6318 obtained with the global phase coherence method described in [1]. The method is also robust, the sensitivity to variation of the parameters is low.

In the future we will address problems specific to motion blur. In some cases, a strong motion blur can be considered as sharp by our approach. With a directional approach, such errors can be avoided. Another interesting development would be to process images in real-time on a smartphone for providing a shooting assistance. The shooting process is hard for visually impaired people, and providing them with real-time guidance would be helpful.

*Acknowledgments.*This research was supported by Cap Digital Business Cluster LINX Project.

8. References

[1] Gwendoline Blanchet and Lionel Moisan. An explicit sharpness index related to global phase coherence. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, pages 1065–1068. IEEE, 2012.

[2] Gwendoline Blanchet, Lionel Moisan, and Bernard Rougé. Measuring the global phase coherence of an image. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 1176–1179. IEEE, 2008.

[3] Luis R Blando, Junichi Kanai, and Thomas A Nartker. Prediction of ocr accuracy using simple image features. In Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on, volume 1, pages 319–322. IEEE, 1995.

[4] Gang Cao, Yao Zhao, and Rongrong Ni. Edge-based blur metric for tamper detection. Journal of Information Hiding and Multimedia Signal Processing, 1(1):20–27, 2010.

[5] Théodore Chabardès and Beatriz Marcotegui. Local blur estimation based on toggle mapping. In Mathematical Morphology and Its Applications to Signal and Image Processing - 12th International Symposium, ISMM 2015, Reykjavik, Iceland, May 27-29, 2015. Proceedings, pages 146–156, 2015.

[6] Ming-Jun Chen and Alan C Bovik. No-reference image blur assessment using multiscale gradient. EURASIP Journal on Image and Video Processing, 2011(1):1–11, 2011.

[7] Alexandre Ciancio, ALN Targino da Costa, Eduardo AB da Silva, Amir Said, Ramin Samadani, and Pere Obrador. No-reference blur assessment of digital pictures based on multifeature classifiers. Image Processing, IEEE Transactions on, 20(1):64–75, 2011.

[8] Jonathan Fabrizio, Beatriz Marcotegui, and Matthieu Cord. Text segmentation in natural scenes using toggle-mapping. In Image Processing (ICIP), 2009 16th IEEE International Conference on, pages 2373–2376. IEEE, 2009.

[9] Matthieu Faessel and Michel Bilodeau. Smil simple morphological image library. In Séminaire Performance et Généricité, LRDE, 2014.

[10] Lawrence Firestone, Kitty Cook, Kevin Culp, Neil Talsania, and Kendall Preston. Comparison of autofocus methods for automated microscopy. Cytometry, 12(3):195–206, 1991.

[11] Frans CA Groen, Ian T Young, and Guido Ligthart. A comparison of different focus functions for use in autofocus algorithms. Cytometry, 6(2):81–91, 1985.

[12] Ray A Jarvis. Focus optimization criteria for computer image-processing. Microscope, 24(2):163–180, 1976.

[13] Van-Cuong Kieu, Florence Cloppet, and Nicole Vincent. Ocr accuracy prediction method based on blur estimation. In Document Analysis Systems (DAS), 2016 12th IAPR Workshop on, pages 317–322. IEEE, 2016.

[14] Henry P Kramer and Judith B Bruckner. Iterations of a non-linear transformation for enhancement of digital images. Pattern recognition, 7(1):53–58, 1975.

[15] Jayant Kumar, Peng Ye, and David Doermann. A dataset for quality assessment of camera captured document images. In Camera-Based Document Analysis and Recognition, pages 113–125. Springer, 2014.

[16] Lin Ma, Songnan Li, Fan Zhang, and King Ngi Ngan. Reduced-reference image quality assessment using reorganized dct-based image representation. Multimedia, IEEE Transactions on, 13(4):824–829, 2011.

[17] Xavier Marichal, Wei-Ying Ma, and HongJiang Zhang. Blur determination in the compressed domain using dct information. In Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, volume 2, pages 386–390. IEEE, 1999.

[18] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. Perceptual blur and ringing metrics: application to jpeg2000. Signal Processing: Image Communication, 19(2):163–172, 2004.

[19] Shree K Nayar and Yasuo Nakagawa. Shape from focus. Pattern analysis and machine intelligence, IEEE Transactions on, 16(8):824–831, 1994.

[20] N Ng Kuang Chern, Poo Aun Neow, and VMH Ang. Practical issues in pixel-based autofocusing for machine vision. In Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on, volume 3, pages 2791–2796. IEEE, 2001.

[21] Xujun Peng, Huaigu Cao, Krishna Subramanian, Rohit Prasad, and Prem Natarajan. Automated image quality assessment for camera-captured ocr. In Image Processing (ICIP), 2011 18th IEEE International Conference on, pages 2621–2624. IEEE, 2011.

[22] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. Pattern Recognition, 46(5):1415–1432, 2013.

[23] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics, 10(4):30–45, 2009.

[24] Marçal Rusiñol, Joseph Chazalon, and Jean-Marc Ogier. Combining focus measure operators to predict ocr accuracy in mobile-captured document images. In Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on, pages 181–185. IEEE, 2014.

[25] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. IEEE Transactions on Image Processing, 21(8):3339–3352, 2012.

[26] Andrés Santos, C Ortiz de Solorzano, Juan José Vaquero, JM Pena, Norberto Malpica, and F Del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. Journal of microscopy, 188(3):264–272, 1997.

[27] John F Schlag, Arthur C Sanderson, Charles P Neuman, and Francis C Wimberly. Implementation of automatic focusing algorithms for a computer vision system with camera control. Carnegie-Mellon Univ., the Robotics Inst., 1983.

[28] Jean Serra. Toggle mappings. From pixels to features, pages 61–72, 1988.

[29] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In Computer Vision, 1998. Sixth International Conference on, pages 839–846. IEEE, 1998.

[30] Zhou Wang and Alan C Bovik. Reduced-and no-reference image quality assessment. Signal Processing Magazine, IEEE, 28(6):29–40, 2011.

[31] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on, 13(4):600–612, 2004.

[32] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In Electronic Imaging 2005, pages 149–159. International Society for Optics and Photonics, 2005.

[33] Hui Xie, Weibin Rong, and Lining Sun. Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images. In Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on, pages 229–234. IEEE, 2006.

[34] Ge Yang and Bradley J Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on, volume 3, pages 2143–2148. IEEE, 2003.

[35] Peng Ye and David Doermann. Learning features for predicting ocr accuracy. In Pattern Recognition (ICPR), 2012 21st International Conference on, pages 3204–3207. IEEE, 2012.

[36] TTE Yeo, SH Ong, R Sinniah, et al. Autofocusing for tissue microscopy. Image and vision computing, 11(10):629–639, 1993.