

# Multi-Exposure Document Fusion Based on Edge-Intensities

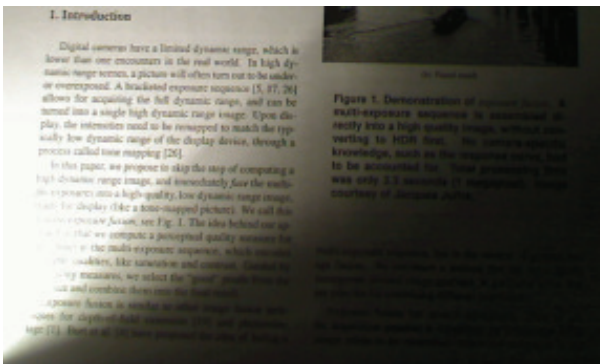
Marco Block, Maxim Schaubert, Fabian Wiesel, and Raúl Rojas  
Freie Universität Berlin  
Institut für Informatik und Mathematik  
Takustr. 9, 14195 Berlin, Germany  
{block,schaubert,wiesel,rojas}@inf.fu-berlin.de

## Abstract

*This paper presents a new algorithm for fusing images of text-documents taken with different exposures. It is compared to several standard block oriented exposure- and focus-blending-algorithms. The recognition rate of a publicly available OCR-engine is used as a benchmark to quantify the results. Experiments show in average an improvement in the recognition rate from 0.46 to 0.64 by employing exposure blending as preprocessing step to an OCR. The presented algorithm of blending high-pass filtered images instead of original images further increases the recognition rate to 0.95.*

## 1. Introduction

With current off-the-shelf cameras, there are often situations, where no single exposure time enables one to capture the whole intensity range as exemplified in figure 1.



**Figure 1. Camera cannot capture the full range of intensities resulting in regions with under- and overexposure**

Parts of the image can be underexposed, while others

are overexposed at the same time. One common approach is take several images of the same scene with various exposures by varying shutter speed or aperture, and combine them to a single High-Dynamic-Range-Image (HDR-Image), later tone-mapping can applied to create a standard image.

However, this requires constant illumination and knowledge of the inverse camera intensity function, which has to be calibrated.

A more simple approach is the direct blending of multi-exposure images according to a blend function. Previous publications evaluated their results with either natural scenes or buildings and focused on achieving a visually appealing result. Contrary to natural scenes, documents are mostly black-and-white, and have large blank areas. This could be a challenge for previous solutions. Also, further document processing does not require the representation of the whole range of intensities in the image, but rather the whole local contrast for better readability.

In the following chapter, we present several published methods of block-wise exposure blending. We then propose a new approach based on the blending of high-passed filters according to the intensities of the edges. The algorithms are evaluated in this new problem domain. Instead of relying on the reader to evaluate the results aesthetically, we quantify the results by determining the recognition rate of the publicly available OCR-system Tesseract. The proposed pre-processing step results in a significantly higher recognition rate.

## 2. Related Work

*Exposure blending* or *exposure fusion* are often used synonyms for the creation of DRI-images [11]. Usually a weighted mean of either intensities [8] or CIELAB colours [4] are used. The weight can be determined by either radial basis function [4] or a power function of contrast, saturation and well-exposedness.

Do we consider the creation of a high-contrast image as a problem of image-composition, then we not only can improve on exposure, but also on sharpness and the depth of field [7, 5].

A common approach is dividing the images  $\mathbf{I}_j, j = 1 \dots N$  in  $d \times d$  blocks and evaluating each block according to some quality measure. A survey of several quality measures used for multi-focus image fusion is given in [6].

From all blocks covering the same area, one selects the block, which has the best result according to the quality measure. The resulting image would introduce discontinuities between the blocks, therefore a continuous blending function is required.

Goshtasby [4] employs rational Gaussian surfaces:

$$w_i(x, y) = \frac{G_i(x, y)}{\sum_{j=1}^M G_j(x, y)},$$

$$G_i(x, y) = e^{-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}}$$

where  $(x_i, y_i)$  is the centre of the  $i$ th-block of a total of  $M$  blocks, and  $\sigma$  the width of the blending function.

The final composition  $\mathbf{H}$  can be written as the weighted sum over the selected blocks:

$$\mathbf{H}(x, y) = \sum_{i=1}^M w_i(x, y) \mathbf{I}_{k_i}(x, y),$$

where  $k_i$  is the index of the image with the highest quality in the  $i$ th-block.

The parameters  $d$  and  $\sigma$  are determined by maximising the result of the quality measure for the whole composed image, as it has been proposed by Goshtasby [4].

## 2.1. Variance

A simple quality measure is the variance of intensities in the block  $B$ :

$$\mathbf{V}(I, B) = \sum_{(x,y) \in B} (I(x, y) - \mu)^2,$$

$$\mu(I, B) = \frac{1}{|B|} \sum_{(x,y) \in B} I(x, y)$$

A higher variance suggests a wider range of intensities, indicating a better exposed image, but may also imply more noise.

## 2.2. Entropy

Goshtasby proposed to use the entropy of a region of an image as a measure to identify the part, which contains the most information [4]. The entropy of a block  $B$  is determined by

$$E(I, B) = \sum_{j=0}^{255} -p_j(I, B) \log(p_j(I, B)),$$

where  $p_j(I, B)$  is the relative frequency of the intensity value  $j$  in the block  $B$  in image  $I$ .

## 2.3. Spatial Frequency

Eskicioglu et al published a quality measure called spatial-frequency, also known as energy of gradient [3]:

$$\mathbf{SF}(I, B) = \sum_{(x,y) \in B} (I(x, y) - I(x - 1, y))^2 + (I(x, y) - I(x, y - 1))^2$$

## 2.4. Energy of Laplacian

Energy of Laplacian is simply the squared magnitude of the Laplacian. This is another measure for the power in the high frequency spectrum. We use the form as given in [6] and summarise the values over the block  $B$ :

$$\mathbf{EOL}(I, B) = \sum_{(x,y) \in B} \mathcal{L}(I)(x, y)$$

with

$$\mathcal{L}(I) = \begin{bmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{bmatrix} * I$$

where  $(*)$  stands for the convolution operator.

## 2.5. Sum of Modified Laplacian

Nayar et al [9] noted, that the  $x$ - and  $y$ -components of the Laplacian can have opposite signs, resulting in a low magnitude, despite a large gradient. To alleviate this problem, Nayar et al are proposing a modified Laplacian

$$\mathbf{ML}(I, x, y) = |2I(x, y) - I(x - \delta, y) - I(x + \delta, y)| + |2I(x, y) - I(x, y - \delta) - I(x, y + \delta)|,$$

where  $\delta$  can be chosen according to the properties of the image. In our case, we followed [6] in selecting  $\delta = 1$ .

The sum over the modified Laplacian of the whole image-block yields the quality measure:

$$\mathbf{SML}(I, B) = \sum_{(x,y) \in B} \mathbf{ML}(I, x, y),$$

for  $\mathbf{ML}(I, x, y) \geq T$

### 3. Pixel-wise Exposure Blending

The rather arbitrary division in  $d \times d$  axis-aligned blocks seems inflexible, as a casted shadow, or the cone of a spot light will seldom follow such a pattern. Therefore a pixel-based approach seems worthwhile.

We suggest a simple extension of the block-oriented approaches above: Instead of dividing the image in non-overlapping blocks, one defines a window of  $d \times d$  around a pixel and determines its quality.

#### 3.1. Pixel-based Entropy

Similar to the original entropy-based method, one could simply select the pixel with the highest entropy

$$\mathbf{H}^{[Max]}(x, y) = \arg \max_j (E(\mathbf{I}_j, B_d(x, y))),$$

where  $B_d(x, y)$  is a  $d \times d$  block around  $(x, y)$ .

As one expects from a discrete selection, the decision boundaries will become visible and produce visual artefacts.

To avoid such discontinuities, a more gradual selection seems prudent. Since entropy is by definition non-negative, and, due to the overlapping nature of the  $d \times d$  windows, rather continuous, it seems prudent to define the resulting pixel as the weighted mean of the original pixels at the same position, weighted according to their respective entropy:

$$\mathbf{H}^{[Pix]}(x, y) = \sum_{j=1}^N w_j(x, y) \mathbf{I}_j(x, y),$$

$$w_j(x, y) = \frac{E(\mathbf{I}_j, B_d(x, y))}{\sum_{k=1}^N E(\mathbf{I}_k, B_d(x, y))}$$

One has to consider, that the entropy in all images may be zero. This happens, when all pixel in a window have the same intensity, in which case the blended pixel assumes that value.

#### 3.2. Continuous Edge Intensity

Another quality measure is the local edge intensity  $\mathbf{L}$  in the image  $\mathbf{I}$ . By folding a Gaussian kernel  $\mathbf{G}$  with the image  $\mathbf{I}$  and then subtracting this smoothed image from the original image, we obtain the edge intensity image:

$$\mathbf{L} = \max(0, \mathbf{I} - \mathbf{G} * \mathbf{I})$$

We limit the values to non-negative values. The spectrum of edge intensities depends on the width of the Gaussian kernel  $\sigma$ .

One can identify fore- and background of the image by the signum of  $\mathbf{L}$ . This method was also successful used to binarise document images [2].

As we are not interested in the absolute intensities as much as the local contrast, we blend the filtered images  $\mathbf{L}_j, j = 1 \dots N$  by calculating for each pixel the weighted-average of the local edge intensities

$$\mathbf{H}^{[Cont]}(x, y) = \sum_{j=1}^N w_j(x, y) \mathbf{L}_j(x, y),$$

$$w_j(x, y) = \frac{L_j(x, y)}{\sum_{k=1}^N L_k(x, y)},$$

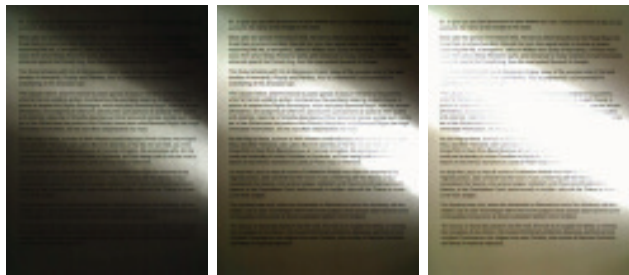
which can be summarised to:

$$\mathbf{H}^{[Cont]}(x, y) = \frac{\sum_{j=1}^N L_j^2(x, y)}{\sum_{k=1}^N L_k(x, y)}$$

### 4. Experiments and Results

The hardware was a QuickCam Pro 9000 from Logitech, which was fixed to the table, thereby minimising eventual displacement. While such a setup does not reflect real world situations, the problem of aligning of multi-exposure images can be considered as an independent one and has to be solved and evaluated separately.

The camera provides a native resolution of 2-mega-pixels. For the experiments, 40 different documents with known content with various typefaces and font-sizes were photographed with 3 different shutter times (example see figure 2). Surrounding lights were positioned to provide for different illumination situations for each document. Day-light provided another light-source, increasing the light spectrum even further.



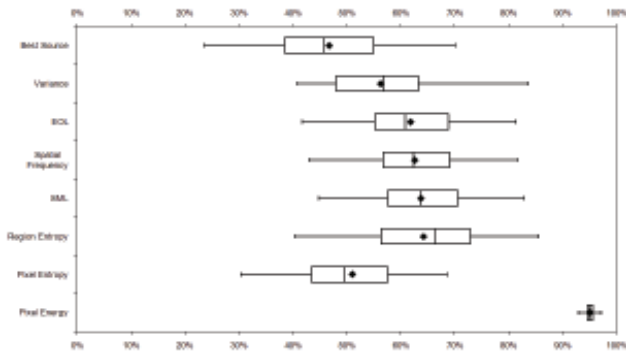
**Figure 2. Three 2-mega-pixel images with different exposures. Left image: Top-right and lower-left is underexposed. Right image: Central-right part is overexposed**

For each of the 40 sets, the presented methods were used to generate a resulting fused image, which was run through the OCR-system *Tesseract* Version 2.03 [1]. The accuracy of the recognition was determined according to the Needleman-Wunsch-algorithm [10].

As can be seen in figure 3, even when the source image with the best recognition rate is selected, we only achieve an accuracy of 0.47 in average, confirming the need for image fusion techniques. All the presented block-based approaches provide an improvement and, except for the variance-based 0.56, even a significant one over a single-exposure image. For further reference, the numerical values are shown in figure 1.

Entropy (0.64), spatial-frequency (0.63), energy-of-laplacian (0.62) and sum-modified-laplacian (0.64) do not differ significantly. Still, the entropy seem to provide the best results among those four quality measures. Simply applying this method on a per-pixel basis does not yield better results (0.51). On the contrary, it is even bested by the worst block-based solution, as the method introduces new artefacts in the resulting image. By high-pass filtering the images, we decrease the differences in intensity between the various exposures. This reduces the artefacts introduced by pixel-wise blending.

Simply blending those high-pass filtered images instead of the original ones according to edge-intensities, we achieve a recognition rate of 0.95. The result is not only significantly better, but even at its worst with 0.92, it still excels the best result of the runner-up (entropy with 0.86).



**Figure 3. Box-and-whiskers plot of the benchmark results of the various implemented algorithms**

## 5. Conclusion and Future Work

We shown that standard *exposure blending* can be used for document-fusion and introduced OCR-recognition rate as a benchmark for this certain application. In context of this benchmark, we improved on existing algorithms. In order to eliminate the influence of the recognition algorithm and preprocessing steps implemented in the chosen OCR-system, we intend to expand the benchmark to other OCRs. Other blending- and fusion-algorithms stand to be evaluated

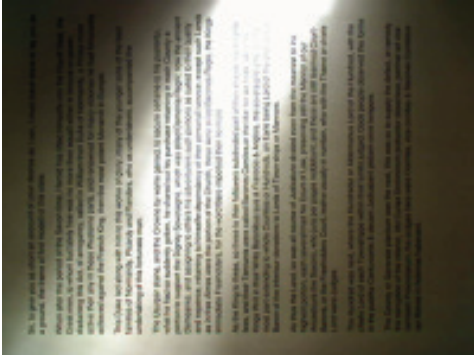
in this new context of document fusion.

**Table 1. Numerical results of the various implemented algorithms in the new benchmark**

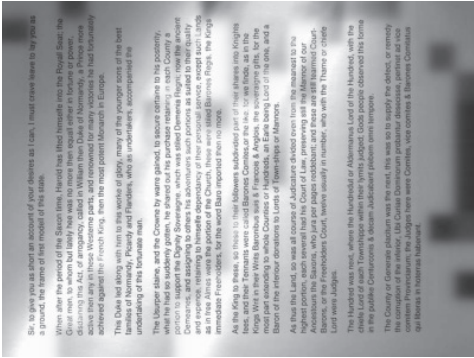
Method	Min	1st Q	Avg	3rd Q	Max
Best Source	0.24	0.39	0.47	0.55	0.70
Variance	0.41	0.48	0.56	0.63	0.84
EOL	0.42	0.55	0.62	0.69	0.81
Spatial Freq.	0.43	0.57	0.63	0.69	0.82
SML	0.45	0.58	0.64	0.71	0.83
Region Entr.	0.40	0.56	0.64	0.73	0.86
Pixel Entr.	0.30	0.44	0.51	0.58	0.67
Edge-Inten.	0.92	0.95	0.95	0.96	0.97

## References

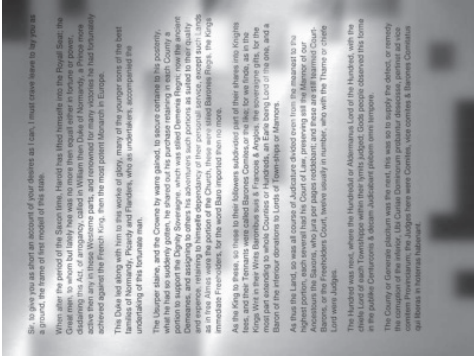
- [1] Tesseract, <http://code.google.com/p/tesseract-ocr/>.
- [2] M. Block and R. Rojas. Local contrast segmentation to binarize images. In *Third International Conference on Digital Society*. IEEE, Feb. 2009.
- [3] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Trans. Commun.*, 43(12):2959–2965, 1995.
- [4] A. A. Goshtasby. Fusion of multi-exposure images. *Image and Vision Computing*, 23(6):611–618, June 2005.
- [5] M. Grundland, R. Vohra, G. P. Williams, and N. A. Dodgson. Cross dissolve without cross fade: Preserving contrast, color and saliency in image compositing. *Comput. Graph. Forum*, 25(3):577–586, 2006.
- [6] W. Huang and Z. L. Jing. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28(4):493–500, Mar. 2007.
- [7] S. T. Li and B. Yang. Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing*, 26(7):971–979, July 2008.
- [8] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion. In *Pacific Conference on Computer Graphics and Applications*, pages 382–390, 2007.
- [9] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):824–831, 1994.
- [10] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarity in the amino acid sequences of two proteins. *J. Mol. Biol.*, 48:443–453, 1970.
- [11] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. Morgan Kaufmann Publishers, December 2005.



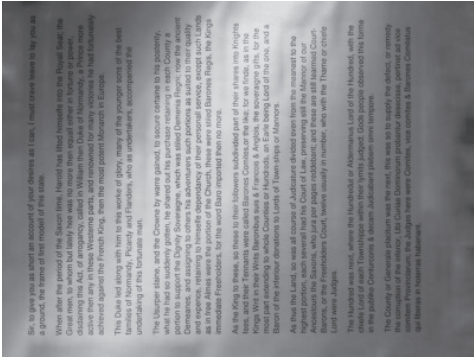
(a) Best Source



(b) Variance



(c) Energy of Laplacian



(d) Spatial frequency



(e) Sum of Modified Laplacian



(f) Region-based entropy



(g) Pixel-based Entropy



(h) Edge-Intensity