MRC – THE PROPOSED DOCUMENT IMAGE COMPRESSION SCHEME

Costin-Anton Boiangiu [1*]
Luiza Grigoraş [2]

ABSTRACT

In this part we propose a new MRC compression scheme, based on k-means clustering for image decomposition into layers and on image interpolation and resampling for filling in sparse layers. This scheme also uses JPEG2000 for the actual compression of the foreground and background layers and JBIG2 for mask compression. Additionally, more than 30 resampling filter functions representing three main filter families (polynomial, exponential and windowed-sinc) have been implemented and their effects on this compression scheme analyzed. From the best ones selected, several conclusions and recommendations have been derived, based on an image quality analysis using PSNR and OCR mean text confidence.

KEYWORDS: *MRC, Document Compression, Image Compression, Data Compression, Image Processing, OCR, Resampling Filters.*

THE PROPOSED COMPRESSION SCHEME

Among the aims of developing this MRC-based compression algorithm were the following: to obtain the best compression rate possible, based on optimal decomposition of an image into overlapping layers; to reduce or even eliminate artifacts introduced by digital devices used to obtain the image or image irregularities coming from the page that was scanned (the case of old, more degraded books); to increase clarity of text against the background, enhancing readability and the number of OCR successful character recognitions. The paper at hand continues the work presented by Boiangiu and Grigoraş (2017).

The steps of the proposed compression scheme are depicted in Fig 1. The preprocessing stages ensure the quality and efficiency of the actual compression. In the segmentation stage, the image is decomposed into the foreground and background layers using a k-means clustering algorithm, specifically 2-means clustering, as there are two groups of pixels that need to be identified: those pertaining to the foreground and those pertaining to the background. In this case, clusters represent groups of pixel values: higher ones for the background, which is lighter, lower ones for the foreground, which is darker. The algorithm is applied in YCrCb color space, to better suit JPEG/JPEG2000 algorithms, in which luminance/luma values are the most important. At first, clusters are composed only of their

---

[1*] corresponding author, Professor PhD Eng., "Politehnica" University of Bucharest, 060042 Bucharest, Romania, costin.boiangiu@cs.pub.ro

[2] Engineer, "Politehnica" University of Bucharest, 060042 Bucharest, Romania, luiza.grigoras@cti.pub.ro

centroids, which are established at Black and White ([0, 0, 0] and respectively [1, 1, 1] - in normalized RGB space). The clusters are constructed further on by assigning each pixel color to one of them, based on the Euclidean distance from the cluster's centroid, in the established colorspace, a distance which has to be smaller than a pre-established threshold. Afterward, new centroids are calculated as the mean value of all pixels in each cluster and the process is repeated. The algorithm ends after no notable displacement of the centroids is observed or after a maximum number of iterations. Although the algorithm is a simple one, it produces satisfactory results for image decomposition into layers.
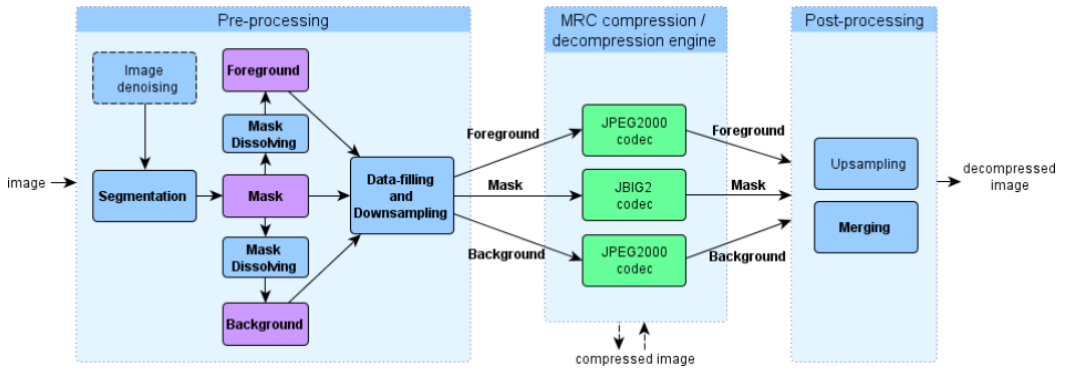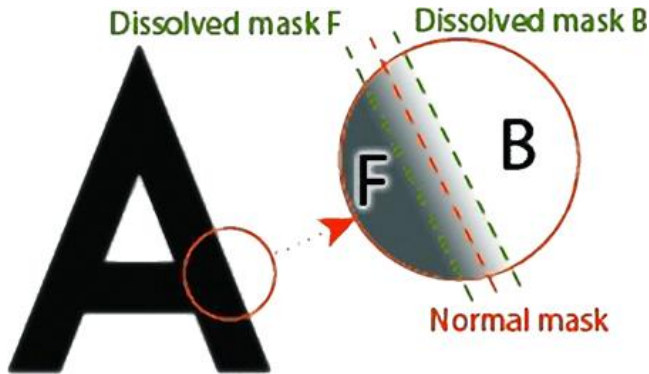

Figure 1. Proposed MRC compression scheme

Figure 2. Illustration of the smooth transitions problem. The edge is not sharp, thus some pixels will be classified as belonging to the foreground (F), and some to the background (B), as shown by the "Normal Mask" delimiter. By dissolving the mask when classifying pixels for foreground or for background, the delimiter is moved so that the resulting layer will contain only appropriate pixels.

The mask dissolving step is necessary before using the mask for separating the foreground and background layers from the original image, in order to counteract the undesired effects generated by soft transitions of edges from foreground to background, also described by Zaghetto and De Queiroz (2008) and illustrated in Fig 2. In this case, the segmentation algorithm cannot clearly classify all edge pixels as pertaining to the foreground or to the background layer. Therefore, the foreground layer will contain pixels from the background, which hinder good compression and produce artifacts on the edges in the resulting image; similar for the background layer. The mask dissolving

preprocessing step implies enlarging or shrinking the holes in the foreground or background layer so that no border pixels will be contained.
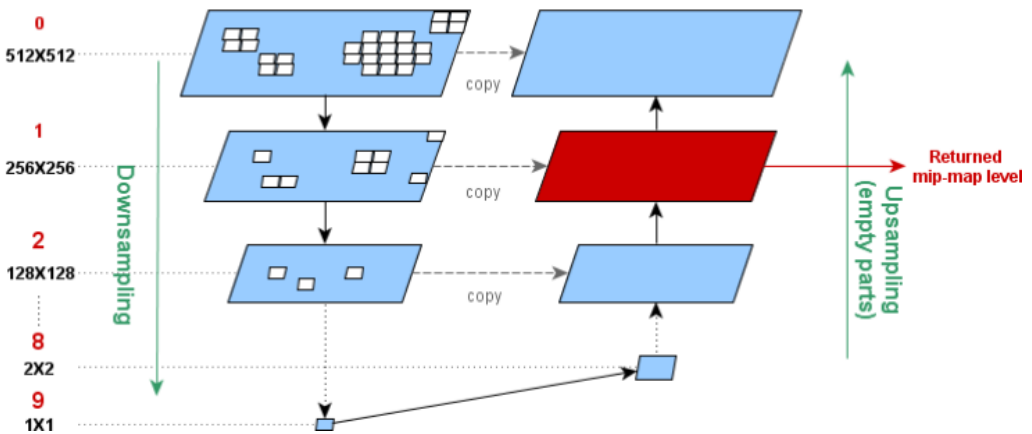

Figure 3. Pyramid of downsampled and upsampled versions of the original foreground/background layer. The white pixels indicate holes in the layer.

The proposed method for data-filling implies using a simplified version of the single image super-resolution technique proposed by Glasner et al. (2009). This super-resolution scheme is based on resampling the image: downsampling and upsampling back. Fig 3 schematically describes the process. A pyramid of images is created by successively downsampling the original image until the most basic level is reached, that of pixels resolution. This is a non-empty pixel, thus a complete information derived from the existing one in the image. This value is then used to fill in the gaps in the closest mipmap level. Further on, information is taken from the newly filled in layer and propagated upward in the same manner, until original image resolution is reached back. At this level, a new image is obtained, preserving the original existing information, but with no gaps. As exemplified by Mukherjee et al. (2002), empty pixels are not taken into account when interpolating. During the upsampling process, only the empty portions of the image have to be filled with information from higher mipmap levels; the rest of the pixels are copied from the original levels, in order to preserve the existing information. The final image layer to be passed to the specific compressor can be a downsampled one; when going upwards, the algorithm may stop at a higher mipmap level. For RGB images, each color channel is resampled individually. The downsampling ratio can be defined as: $1 : MipMapPower^{2MipMapLevel}$, where MipMapPower represents the actual downsampling factor, i.e., the ratio between the width or height of adjacent level images.

The compression engine uses JPEG2000 compression for the foreground and background layers and JBIG2 compression for the mask layer. These algorithms have been chosen because they are quite recent and have been developed with the specific purpose of improving (and even replacing) older compression schemes for continuous-tone images and bi-level images respectively.

RESULTS AND DISCUSSION

The performances of the proposed MRC codec have been evaluated with regard to machine-printed document images, some containing graphics and also line art (Fig 4).
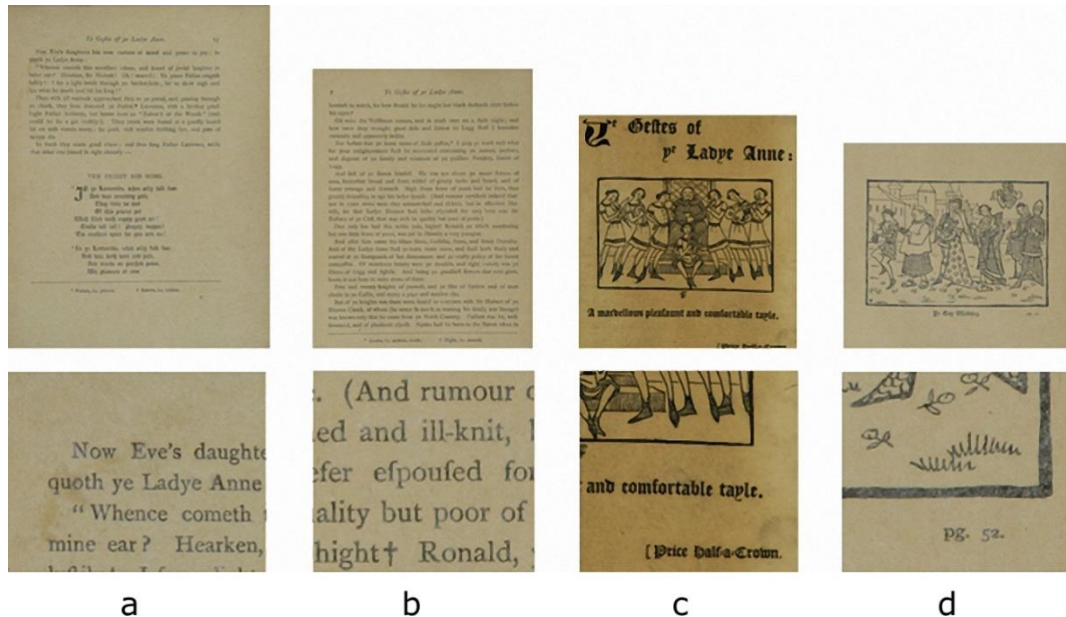

Figure 4. Images used for testing: (a) "Book Page 1", (b) "Book Page 2" , (c) "Minstrels", (d) "Wedding". The age of the scanned book pages and the scanning process may produce unwanted artifacts (exemplified below the test images) in the final images, making automated text recognition harder: the texture of paper, stains, fading of letter edges or other contours, background noise.

The case studies presented in the following subsections were conducted in order to determine the best parameters for the data-filling stage, which influence the entire MRC compression process. These parameters are the resampling filters, first of all, followed by the mipmap power and downsampling ratios for the foreground and background layers. The first case study performs a strictly quantitative, objective analysis, determining the best filters based on the size of the obtained image. The second case study performs a finer analysis of these best filters, based on qualitative objective measures (PSNR and OCR confidence), from which several final conclusions have been derived.

Quantitative Measurements

The purpose of the first case study was to determine the best-suited resampling filters for the proposed MRC compression scheme. For the results to be relevant, JPEG2000 lossless compression was used and the resulting foreground and background layers were kept at original image sizes (mipmap level to return is zero). The best filters were considered to be those which determined the best compression ratio (the smallest final image size). A part of the results of this test is exemplified in Fig 5.

For all images and mipmap powers, similar filter performances have been registered. It can be observed that the polynomial filters perform best, as suggested by their grouping in the left extremity of the graph. The exceptions from this category are the Cubic Spline filter with parameter α=-1 and the Box filter, which sometimes showed average performance, comparable to that of Flat-Top windowed-sinc, or worst performance, depending on the content of the image. Performances of the Gaussian filter are similar to those of the polynomials. From the BC family, the Notch filter produces very good results, with Mitchell, Robidoux and Catmull-Rom following at some distance. At the other extremity are the windowed-sinc filters. From this category, the sinc windowed by the Flat-Top function performs best, followed (not closely) by Nuttal, Blackman-Harris and Blackman-Nuttal.

As emphasized in Section 1, an efficient resampling filter would be one that produces a smooth layer, property requested by JPEG2000 compression. In the light of this statement, the results can be easily explained. The most blurring polynomial filters perform best. The Gaussian filter, the Cubic Splines, and B-Spline, with their excessive blurring, produce the best results. Even though simple, the Box and Triangle filters perform quite well. This can be justified by the fact that the Box filter produces large patches of constant color, which can be well compressed (even though transitions between these patches may be abrupt) and the Triangle filter produces smooth gradients (Thyssen, 2017). The behavior of the filters from the BC-family can be explained through the fact that they are designed as the best compromise between the level of detail in the image and the number of artifacts, diminishing blurring, which, in this case, is advantageous. The windowed-sinc filters are designed to preserve details in the image, having sharpening effects, and thus produce poorer results.
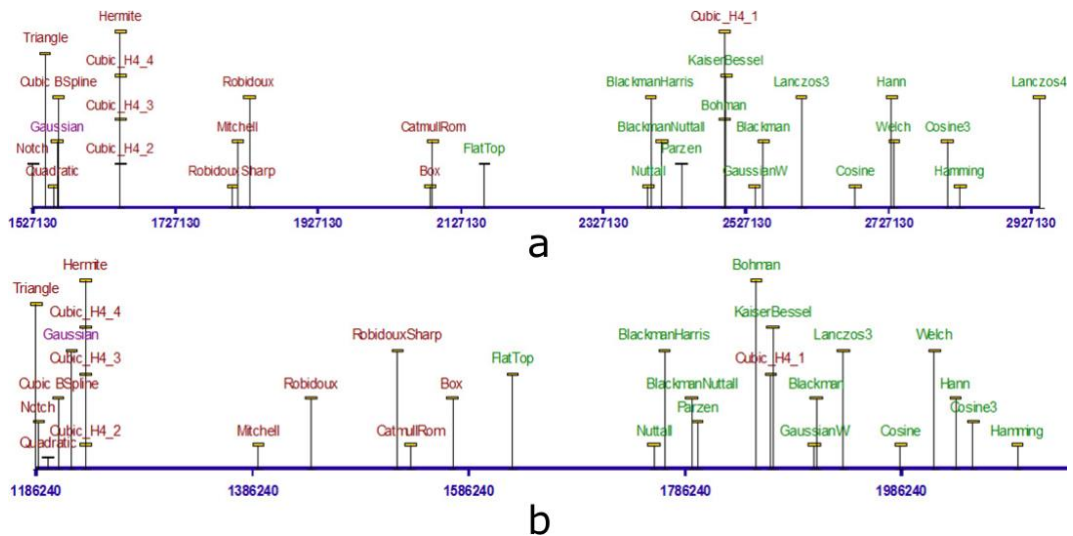

Figure 5. Image sizes obtained for all tested filters. (a) "Book Page 1" and (b) "Minstrels". The dimensions obtained and the filters are marked on the axis. Filter names are colored as follows: red - polynomial filters, green - windowed-sinc filters, magenta - exponential filters. A similar grouping of filters can be observed in both images.

For the second case study, a number of 14 filters from all categories have been selected, as follows: Triangle, Hermite, Quadratic, Cubic B-Spline, Mitchell, Catmull-Rom, Notch, Robidoux, Gaussian, Flat-Top, Parzen, Nuttall, Blackman-Nuttall, Blackman-Harris. The Box and Robidoux Sharp filters have been omitted, because of their inconsistency between tests. The cubic splines Cubic_H4_2, Cubic_H4_3, Cubic_H4_4 have also been omitted because their performances in terms of the size of image and PSNR are almost equivalent to those of Hermite, and the latter has been preferred over them, having better time results as well.

Qualitative Measurements

The purpose of the second study was to further distinguish between performances of the filters for which similar compression ratios have been previously obtained. The emphasis was on the quality of the image: both PSNR and OCR mean text confidence metrics have been used to evaluate filter performances. The text confidence was obtained using the Tesseract OCR engine (Smith, 2007).

Several mipmap powers have been tested, their influence on the quality of images being illustrated in Fig 6. PSNR values did not decrease drastically; good qualities for all images have been obtained even for high values, such as 7 or 8. It has also been observed that PSNR values varied inversely proportional to the foreground mipmap power and directly proportional to the background mipmap power. For time performance reasons, higher values for mipmap powers have been preferred (5-8) and smaller values (1-2) for mipmap levels to return.
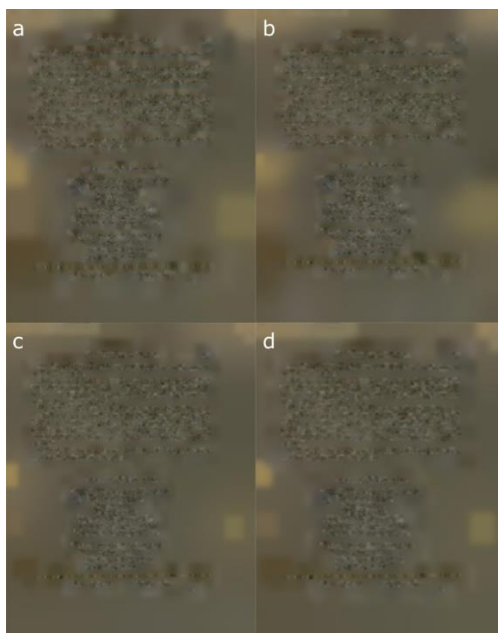

Figure 6. Foreground outputs for the Notch filter. "Book Page 1" image, different foreground mipmap powers: (a) 4, (b) 5, (c) 6, (d) 7. In this case, a power of 5 produces the finest transitions, while powers of 6 and 7 make the foreground more uniform, but isolate patches of color. These values do not necessarily produce the same effects for all filters; each filter has specific best-functioning parameters.
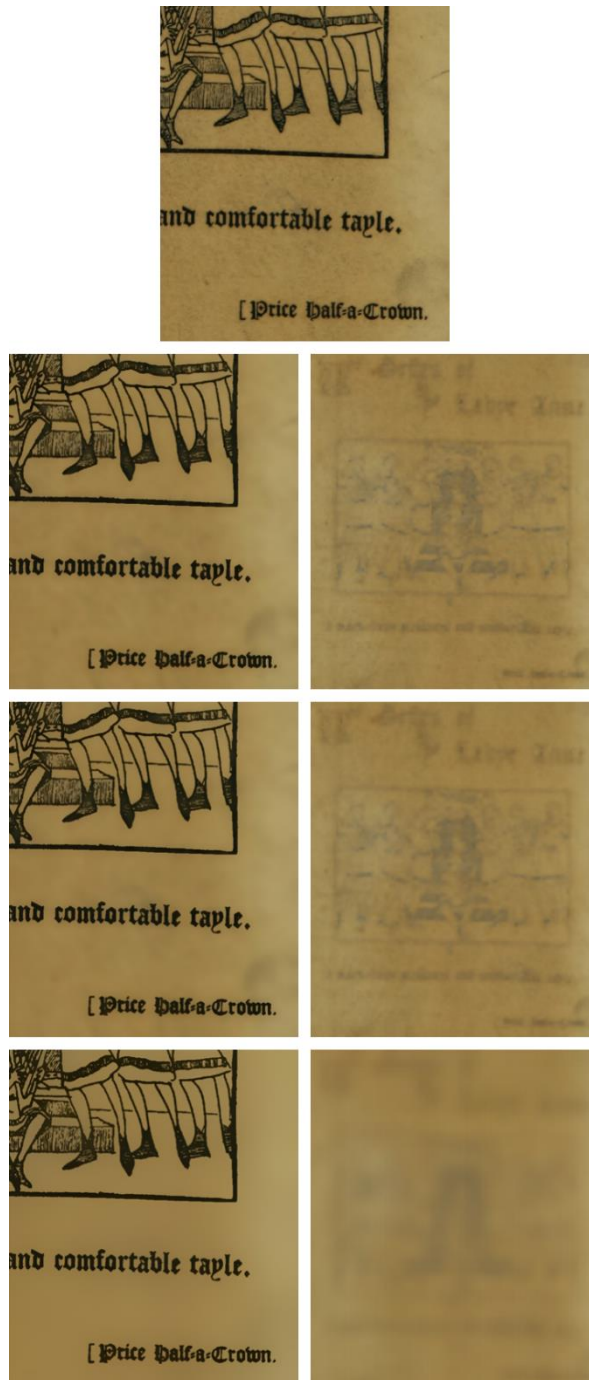
Figure 7. Progressive background smoothing of image "Minstrels" (detail), Cubic B-Spline filter, various combinations of parameters (mipmap power foreground, mipmap power background, mipmap level to return foreground, mipmap level to return background), with corresponding background layer of entire image (below each one): 1st line - original, 2nd line - 5, 6, 1, 1, 3rd line - 7, 8, 1, 1, 4th line - 5, 6, 2, 2.
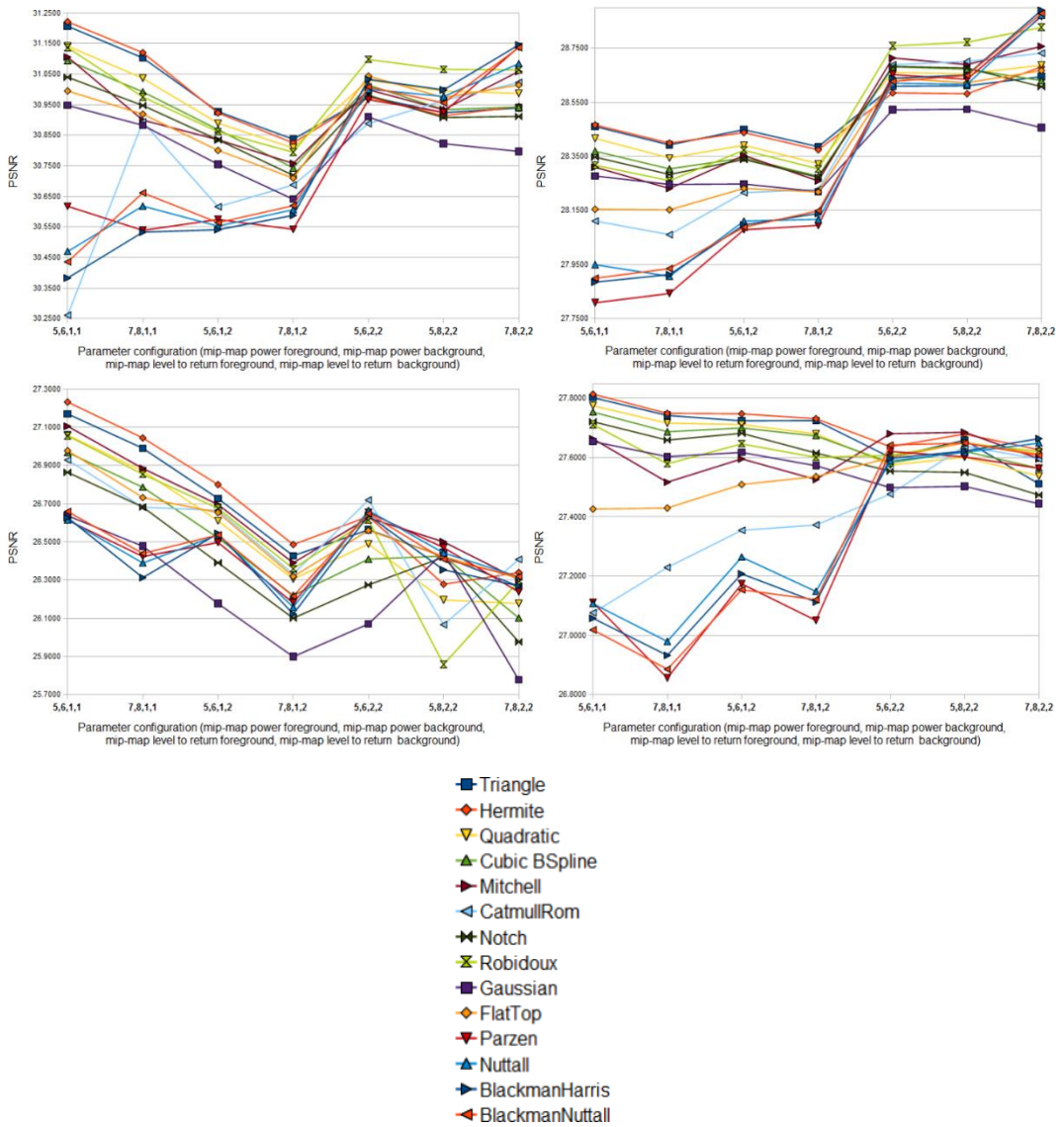
Figure 8. PSNR graphs for images: Top to bottom: "Book Page 1", "Book Page 2", "Minstrels", "Wedding". Evolution of filter performances is shown comparatively, over several configurations of parameters (in ascending order of compression rate).

Figure 9. OCR graphs for images: Top to bottom and left to right: "Book Page 1", "Book Page 2", "Minstrels". Evolution of filter performances is shown comparatively, over several configurations of parameters (in ascending order of compression rate). These w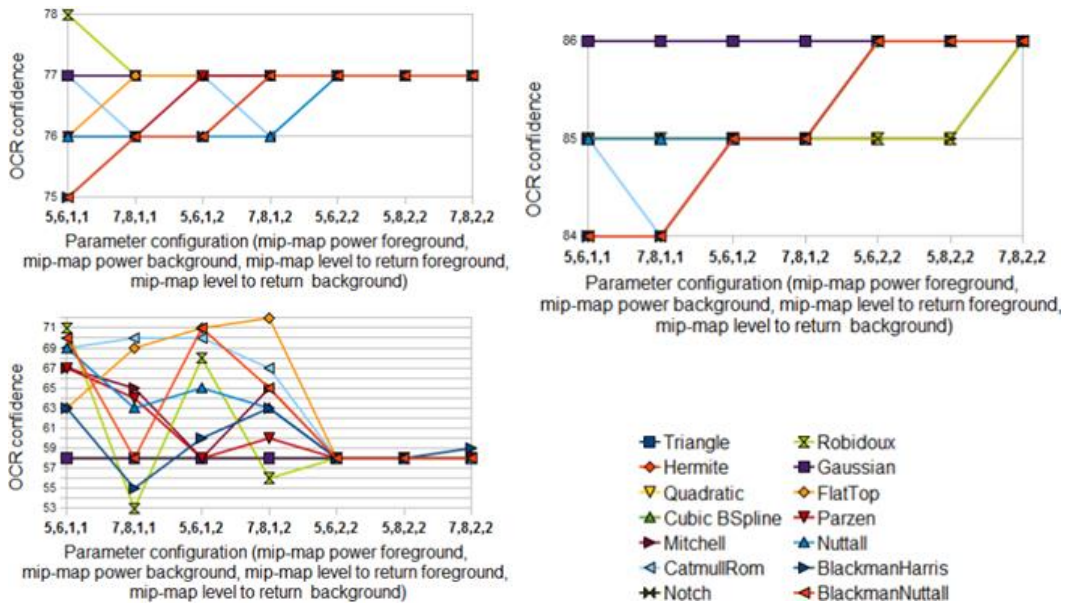ere chosen as a compromise between all filters, producing good smoothing of foreground and background and ensuring high compression rates and smaller computation time.

PSNR tests results are quite similar for all images and all combinations of mipmap powers. The graphs of these results are shown in Fig 8. The top performers were (constantly) Quadratic, Notch, Cubic B-Spline, Hermite, and Triangle. Mitchell, Robidoux, Catmull-Rom, Gaussian and Flat-Top performances vary especially with the types of images. In some cases ("Book Page 1", "Wedding"), Mitchell comes to the front as the best and Robidoux places itself quite high as well. The Flat Top window has constant performances and distinguishes itself as the best (in most cases) from its family of windowed-sinc filters also in terms of PSNR, placing itself well among the polynomials. For severe compressions of images "Minstrels" and "Wedding", the windowed-sinc filters (especially Blackman-Nuttall) have comparable performances with the others and even outperform the Notch filter and some of the polynomials. This phenomenon is less obvious in the case of "Book Page 1" and "Book Page 2" images (Blackman-Harris performs best).

OCR confidence tests results are depicted in Fig 9. For "Wedding" and "Minstrels" images, the selected windowed-sinc filters (the Flat Top filter can be outlined again) and BC-family filters (excepting Notch) have usually registered best results. For images with many more lines of text ("Book Page 1" and "Book Page 2"), OCR relative performance was rather constant for each filter. OCR confidence values improved with the smoothing of the background (Fig 7), i.e., the increase of mipmap powers and mipmap levels to return. For "Book Page 1", Gaussian, Notch and Cubic B-Spline stood out, followed by Quadratic and Robidoux. For "Book Page 2", Triangle, Hermite, Quadratic, Cubic B-Spline, Gaussian and Notch filters were constantly the best. Also, for all images, the OCR

confidences become uniform with high parameter values (high compression rate), all filters registering the same values approximately.

CONCLUSIONS

In this part we presented an MRC-based codec which uses a compression scheme based on a simple super-resolution idea for the data-filling of sparse foreground and background layers and JPEG2000 compression for the same layers. Emphasis was placed on the data-filling stage, of great importance on preparing layers for actual JPEG2000 compression. The performances of the codec have been evaluated in terms of size and quality of the final image and time of compression, by varying the parameters which influence the data-filling stage: the resampling filters, the mipmap powers and the mipmap levels to return. This concludes the work described in Paşca (2013).

Mip-map powers and mipmap levels to return have great influence on the size and quality of the final image. An equilibrium between the values of these parameters is preferable. We recommend medium compression rates - tested combinations of 5, 6, 1, 2 or 7, 8, 1, 2 (mipmap power foreground, mipmap power background, mipmap level to return foreground, mipmap level to return background). Higher values for foreground powers (8-9) and lower values for background powers (4-5) can also be used; the same filter characteristics will remain valid. We also recommend that the background layer should be at least as heavily compressed as the foreground (by using higher values for mipmap powers and the levels to return), as this way a smoother, clearer background is obtained for the text which becomes easier to read and to be processed by an OCR. Noise and other artifacts from the original paper or caused by the scanning process are eliminated. This is proven by the increased values of PSNR for high background mipmap powers and higher OCR confidence for increased levels to return.

Filter performances vary greatly with the type of content the image has and with the aforementioned parameters. Their performances reflect theoretical predictions and several of them can be recommended for usage. Some filters registered constant performances in terms of both image size and quality in most cases, while others sometimes performed the best and sometimes average or even the worst. Generally, for MRC compression using JPEG2000, polynomial filters are recommended, because they produce the most smoothing of the layers. The Gaussian filter produces similar blurring, but with high cost of computation time and sometimes poor PSNR and OCR results.

Constant and good performances for all images have been registered by Hermite, Quadratic, Cubic B-Spline and even Triangle, along with Notch and Flat-Top. These filters represent a good compromise between size, computation time and quality of the image. The rest of the BC-family and the windowed-sinc filters, along with the Gaussian have shown fluctuations in performances, sometimes severe.

For small to medium compression rates, Triangle and Hermite filters registered good results on all types of images. With the increase of compression rate, the BC-family and the windowed-sinc filters advanced more on the PSNR scale. For medium to high compression rates, BC-family filters distinguished themselves, namely Mitchell (graphics and line art images), Catmull-Rom (graphics and line art images) and Robidoux (text

images); for high compression rates, windowed-sinc filters obtained the best performances, especially Blackman-Harris and Blackman-Nuttall.

The compression scheme can be further improved by implementing more suggestions from the MRC recommendation (ITU-T Recommendation T.44, 2005), such as cropping the layer so that only the actually useful part is subdued to compression and using MRC header options to specify offsets and regions of constant color. The segmentation algorithm could be modified so that the image is split into more than two layers: k-means clustering with k equal to 3 or 4, determining k based on certain heuristics. Also, a mathematical study and filter design matching the characteristics of the resampling filter to those of the wavelet filter used in JPEG2000 compression might prove useful.

Future work will also be conducted on three main directions in order to improve the multi-layer separation technique, without the usage of the k-means clustering, by encompassing modern approaches like model-based fitting background information (Minaee and Wang, 2015), sparse-smooth decomposition (Minaee et al., 2015), and morphological-based methods (Mtimet and Amiri, 2013). This will ensure that, alongside with the choosing of the most appropriate filter function and automatic tuning of parameters involved in the plane-filling method, the MRC codec will also benefit from selecting the most suitable plane separation technique, in order to achieve even better performance.

REFERENCES

[1]   Bottou, L., Haffner, P., Howard, P.G., Simard, P., Bengio, Y., LeCun, Y. (1998). High-Quality Document Image Compression with DjVu. J. Electron. Imaging, 7, pp. 410-425.

[2]   De Queiroz, R.L., Buckeley, R., Xu, M. (1999). Mixed Raster Content (MRC) Model for Compound Image Compression. In Proceedings of SPIE Visual Communications and Image Processing, volume (3653), pp. 1106-1117.

[3]   De Queiroz, R.L. (2000). On Data Filling Algorithms for MRC Layers. In Proceedings of the IEEE International Conference on Image Processing, volume (2), pp. 586-589, Vancouver, Canada.

[4]   De Queiroz, R.L. (2005). Compressing Compound Documents. In Barni, M. (ed.), The Document and Image Compression Handbook. Marcel-Dekker.

[5]   De Queiroz, R.L. (2006). Pre-Processing for MRC Layers of Scanned Images. In Proceedings of the 13th IEEE International Conference on Image Processing, pp. 3093-3096, Atlanta, GA, U.S.A.

[6]   Glasner, D., Bagon, S., Irani, M. (2009). Super-Resolution from a Single Image. In Proceedings of the 12th IEEE International Conference on Computer Vision, pp. 349-356, Kyoto, Japan.

[7]   Pașca L. (2013), Hybrid Compression Using Mixed Raster Content, Smart Resampling Filters and Super Resolution, Diploma Thesis, unpublished work (original author name Pașca L., actual name Grigoraș L.).

[8]   Haneda, E., Bouman, C.A. (2011). Text Segmentation for MRC Document Compression. IEEE Trans. Image Process, pp. 1611-1626.

[9]   Harris, F.J. (1978). On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. Proc. IEEE, 66, pp. 55-83.

[10]   Hauser, H., Groller, E., Theussl, T. (2000). Mastering Windows: Improving Reconstruction. In Proceedings of the IEEE Symposium on Volume Visualization, pp. 101-108. Salt Lake City, UT, U.S.A.

[11]   ITU-T Recommendation T.44 Mixed Raster Content (MRC), (2005).

[12]   Lakhani, G., Subedi, R. (2006). Optimal Filling of FG/BG Layers of Compound Document Images. In Proceedings of the 13th IEEE International Conference on Image Processing, pp. 2273-2276, Atlanta, GA, U.S.A.

[13]   Lehmann, T.M., Gönner, C., Spitzer, K. (1999). Survey: Interpolation Methods in Medical Image Processing. IEEE Trans. Med. Imaging, 18, pp. 1049-1075.

[14]   Minaee, S., Abdolrashidi, A., Wang, Y. (2015). Screen content image segmentation using sparse-smooth decomposition. 49th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, pp. 1202-1206. DOI: 10.1109/ ACSSC.2015.7421331.

[15]   Minaee, S., Wang, Y. (2015). Screen content image segmentation using least absolute deviation fitting. 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, pp. 3295-3299. DOI: 10.1109/ ICIP.2015.7351413.

[16]   Mitchell, D.P., Netravali, A.N. (1988). Reconstruction Filters in Computer Graphics. ACM SIGGRAPH Comput. Graph, 22, pp. 221-228.

[17]   Mtimet, J., Amiri, H. (2013). A layer-based segmentation method for compound images. 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), Hammamet, pp. 1-5. DOI: 10.1109/ SSD.2013.6564005.

[18]   Mukherjee, D., Chrysafis, C., Said, A. (2002). JPEG2000-Matched MRC Compression of Compound Documents. Proceedings of the IEEE International Conference on Image Processing, volume (3), pp. 73-76.

[19]   Parker, J.A., Kenyon, R.V., Troxel, D.E. (1983). Comparison of Interpolating Methods for Image Resampling. IEEE Trans. Med. Imaging, 2, pp. 31-39.

[20]   Pavlidis, G. (2017). Mixed Raster Content, Segmentation, Compression, Transmission. Signals and Communication Technology Series, Springer Singapore, DOI: 10.1007/ 978-981-10-2830-4.

[21]   Rabbani, M., Joshi, R. (2002). An overview of the JPEG 2000 still image compression standard. Signal Process. Image Commun, 17, pp. 3-48.

[22]   Smith, S.W. (1997). The Scientist and Engineer's Guide to Digital Signal Processing, 1st ed., pp. 285-296. California Technical Publishing, San Diego, CA, U.S.A.

[23] Smith, R. (2007). An Overview of the Tesseract OCR Engine. In Proceedings of the 9th International Conference on Doc Analysis and Recognition, volume (2), pp. 629-633, Curitiba, Parana, Brasil.

[24] Thévenaz, P., Blu, T., Unser, M. (2000). Image Interpolation and Resampling. In Handbook of Medical Imaging. Processing and Analysis; Academic Press series in biomedical engineering, pp. 393-420. Academic Press, San Diego, CA, U.S.A.

[25] Thyssen, A. (2017). ImageMagick v6 Examples - Resampling Filters. http:// www.imagemagick.org/ Usage/ filter (Accessed: January 25, 2017).

[26] Turkowski, K. (1990). Filters for Common Resampling Tasks. In Glassner, A.S. (ed.), Graphics Gems, pp. 147-165. Academic Press, San Diego, CA, U.S.A.

[27] Unser, M. (1999). Splines: A Perfect Fit for Signal and Image Processing. IEEE Signal Process. Mag., 16, pp. 22-38.

[28] WOLFRAM (2017). Filter-Design Window Functions. https:// reference.wolfram.com/ language/ guide/ WindowFunctions.html. (Accessed: January 25, 2017).

[29] Zaghetto, A., De Queiroz, R., L. (2007). MRC Compression of Compound Documents Using H.264/AVC-I. Simpósio Brasileiro de Telecomunicações, Recife, Brasil.

[30] Zaghetto, A., De Queiroz, R.L. (2008). Iterative Pre- and Post-Processing for MRC Layers of Scanned Documents. In Proceedings of the 15th IEEE International Conference on Image Processing, pp. 1009-1012, San Diego, CA, U.S.A.

[31] Boiangiu, C.A., Grigoraș L. (2017). "MRC – The Theory of Layer-Based Document Image Compression", The Proceedings of Journal ISOM, Vol. 11 No. 1 (Journal of Information Systems, Operations Management).