Automatic Frame Extraction for Improving Content-Based Image Retrieval of Historical Photographs

Ana Paula B. Lopes[†§], David L. Flam[†], Natália. C. Batista[†], Sandra E. F. de Avila[†], Jussara M. Almeida[†], Arnaldo de A. Araújo[†]

[†]Computer Science Department – Minas Gerais Federal University
Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, Minas Gerais, Brazil

[§]Exact and Technological Sciences Department – Santa Cruz State University
Campus Soane Nazaré de Andrade, km 16 Rodovia Ilhéus-Itabuna, Ilhéus, Bahia, Brazil

{paula, david, natalia, sandra, jussara, arnaldo}@dcc.ufmg.br

ABSTRACT

In this paper, we propose and evaluate a method for automatic frame extraction from a collection of historical photographs. These frames are very noisy and were demonstrated to significantly affect the results of content-based image indexing and retrieval in the photograph images. The method is based on parallelogram detection that uses a Hough transform variation called Tiled Hough Transform in which the image is split into tiles to reduce computational complexity. This detector is then extended to combine detected parallelograms into a resulting frame. Two key contributions of this work are: (1) a new effective technique to solve the photographs frame problem, and (2) the use of a set of statistical and experimental design techniques either to fine-tune the proposed method and to demonstrate its effectiveness.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Experimental Design; I.4 [**Image Processing and Computer Vision**]: Applications

General Terms

Design, Experimentation

Keywords

Content-Based Image Retrieval, Historical Photographs, Cultural Heritage

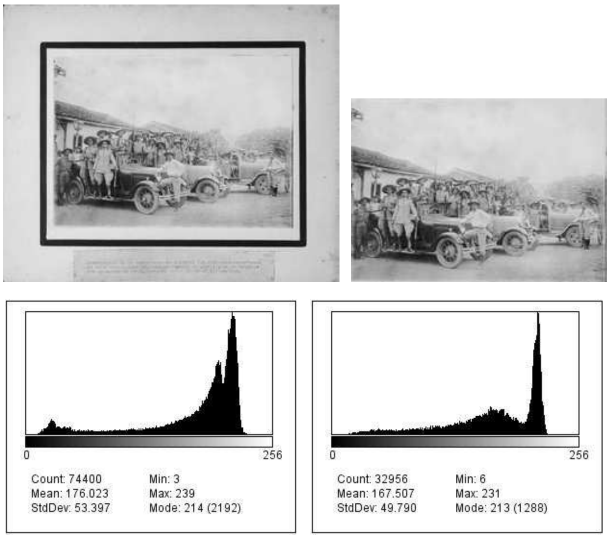
1. INTRODUCTION

The digitization of documents related to human historical and cultural heritage has emerged as a solution for at least two important issues, namely preservation and access. Moreover, storage, indexing and retrieval of such images, which have some unique characteristics when compared to contemporary ones, have been pointed out as important research topics [3].

When working with historical photographs it is common to find frames around them, as in the examples shown in Figure 1. As can be seen, these frames can serve decorative purposes or be caused by flaws in the digitization process. In tasks such as content-based indexing and retrieval [13], these frames can be highly noisy, tampering with the image extracted features. This has occurred for example in [11], which has firstly motivated this work. That work used Content-Based Image Retrieval techniques to identify different types of photographic prints, which can reveal important historical details. There, the frames had been extracted manually, which is a time consuming task and can become impractical depending upon the database size [2].

The effect of such frames can be observed, for example, by verifying the differences in the gray level histograms of an image with and without its frame. Figure 2 illustrates this: there, it can be seen that the gray level mean value of the image is reduced from 176 to 167 when the frame is removed. In Section 3 of this paper it is sh-

Figure 1: Examples of historical photographs with frames.



(a) Image with frame. (b) Image without frame.

Figure 2: The effect of frame in the gray level histogram of an image.

own that this difference is statistically significant in most images analyzed. Therefore, if one is interested in processing and analyzing image content, the removal of frames is an important preprocessing step.

An algorithm for removing noisy borders in monochromatic images of digitized documents was proposed by [1]. Their border removal algorithm is based on Flood Fill, Component Labelling and Region Adjacency Graphs and its quality is argued to be better in comparison to similar commercial tools whose algorithms seem to be proprietary and unpublished. In that case, however, the borders were introduced only by the digitization process using automatically fed scanners, having homogeneous color and texture, instead of the heterogeneous decorative or noisy frame borders we propose to remove.

Most frames have a rectangular shape, which allows a rectangle or parallelogram detector to be used to identify these frames. Because these images are very noisy, Hough transform based methods are the most suitable ones [5]. This paper proposes a method for automatically finding and extracting frames from digitized historical photograph images which is based on a parallelogram detector presented in [10], but extended to combine the detected parallelograms into a resulting frame.

That detector has the advantage of using tiles to reduce computational complexity due to Hough transform. The proposed method was applied to a sample of 633 images randomly selected from Minas Gerais State Public Archive historical photograph collection[1]. The collection is made up of over 80,000 photographs, out of which about 6,500 have already been digitized. The sample was extracted from this last set.

The main contributions of this paper are (1) to develop a new effective method for removing rectangular frames from historical photographs aiming at they are best suited for content-based image processing and (2) the use of a set of statistical and experimental design techniques to tune the method and to prove its effectiveness.

Also, this work provides an independent reassessment of the parallelogram detector proposed in [10] against a real and larger image database.

The proposed method is discussed in Section 2. The experimental results are described in Section 3. Finally, some conclusions are derived in Section 4.

2. PROPOSED FRAME EXTRACTION METHOD

This section describes the four main steps of the method for detecting and extracting photograph frames: a) detection of parallelograms, b) selection of candidate parallelograms, c) merging of candidate parallelograms and d) frame extraction. Before removing frames, though, the sample photographs were analyzed and preclassified in order to be possible to evaluate the effectiveness of the method. The overall process is summarized on Figure 3.

2.1 Pre-classification of the images sample

As illustrated in Figure 1, the frames found in the target image collection are of varied types. In particular, some of them are very large when compared to the whole image area, while others are very thin. Although it is intuitively obvious that a larger frame can alter the image histogram more than a thinner one, an objective *criterion* is needed to separate significant frames from non-significant ones. Although failures in detecting non-significant frames are irrelevant for practical purposes, if these images are computed together with the significant ones they can hinder the analysis of the algorithm effectiveness, since most images of the database have thin frames, as can be seen in Figure 4.

To find such separation *criterion*, we proceeded as follows: it was found that from 373 photographs with frames, 104 had frames that occupied up to 10% of the image area. These photographs correspond to the first bin of the histogram in Figure 4. Then, five of these photographs were randomly selected and had their gray level mean values before and after frame extraction compared. The values found are in Table 1 and from these it was found that – with 95% of confidence – the difference between them is not statistically significant. Thus, we expect the effect of the frame on the gray level histogram of images in this category to be negligible.

The 127 photographs in which frames occupy between 10 and 20% of the image area (second bin of the histogram) were analyzed by the same process, but in this case the mean difference resulted statistically significant. With these results, shown in Table 2 all 633 photographs from the sample set (with and without frames) were categorized into three classes: images without frames (41% of the whole sample), images with negligible frames, taken as the images where the frame was under 10% of image area, (16% of the sample) and images with significant frames (43% of the sample).

2.2 Detection of parallelograms

It is a well-established result that methods based on Hough transform are more robust to detect shapes in noisy environments when compared to methods that work directly on the image space [5]. However those methods suffer with a high time and space complex-

[1]Arquivo Público Mineiro (APM),
http://www.cultura.mg.gov.br/?task=home&sec=5

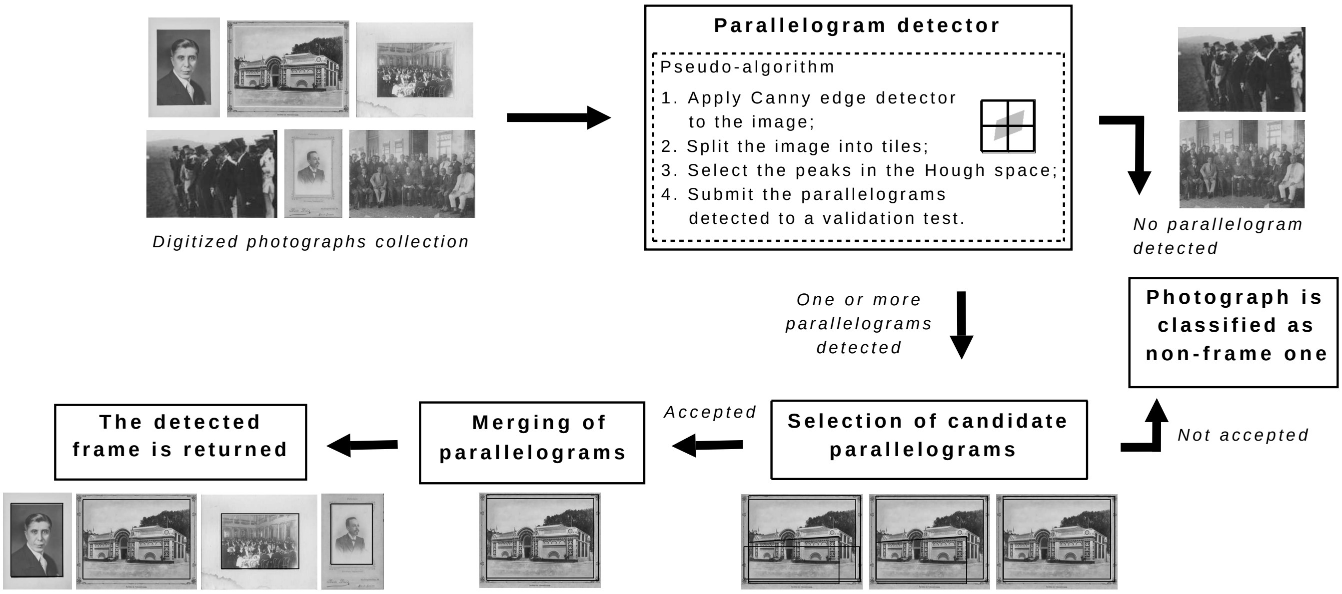


Figure 3: Frame detection steps.

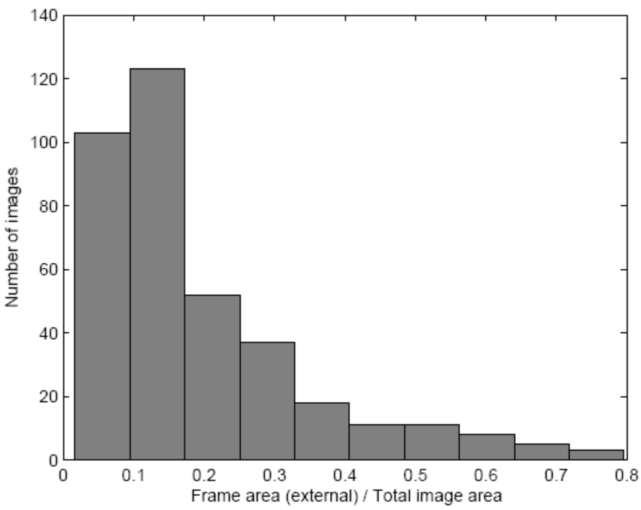


Figure 4: Distribution of the images with frames based on relative frame area.

ity. Thus, a number of strategies which introduces domain-guided modifications into Hough transform to detect specific shapes has been proposed. Some examples are [9] for detecting rectangles, [4] for ellipses and [6] for circular forms.

The parallelogram detector proposed in [10] is particularly suitable for the frame detection problem. Their algorithm uses a Hough transform variation called Tiled Hough transform, which is expected to be robust in detecting parallelograms in the noisy old photographs of the target image database. This detector is based on the fact that each straight line appears as a peak in the Hough space. Finding peaks that have certain geometric relations among them indicates the existence of a parallelogram in the image. So,

Table 1: Mean gray levels of five randomly chosen images with and without frames. These images were chosen among those with a frame area smaller than 10% of the whole image.

Image	Mean Gray Level		
	Frame	No Frame	Diff.
1	158.08	154.11	3.97
2	159.83	159.00	0.83
3	112.35	109.26	3.09
4	80.08	78.49	1.60
5	109.66	105.87	3.78
Conf. Interv. for the average diff.:			-0.29 to 2.65

Table 2: Mean gray levels of five randomly chosen images with and without frames. These images were chosen among those with a frame area between 10 and 20% of the whole image.

Image	Mean Gray Level		
	Frame	No Frame	Diff.
1	133.88	128.01	5.87
2	122.40	113.97	8.42
3	102.82	90.59	12.23
4	135.53	129.54	6.00
5	134.85	127.13	7.71
Conf. Interv. for the average diff.:			2.54 to 8.05

an important parameter of the detector is the threshold point for observing these peaks. If beyond the threshold point, a peak may become a side of a parallelogram. Parallelogram detection is applied after the edge map of the image is computed.

A remarkable feature of this parallelogram detector is that it splits the image into tiles, applying the Hough Transform to each tile. This tiled-feature has the advantage of reducing computational overload due Hough transform computation. Also, in contrast with the similar method described in [9], this one is able to detect parallelograms across up to four tiles, disposed two by two. So, by using exactly four tiles, it was possible to take advantage of the reduced complexity of the method of [10] even for finding the large rectangles that should indicate the presence of frames.

The parallelograms detected by this technique are submitted to a final validation test, which consists in comparing the estimated perimeter with the actual perimeter. The estimated perimeter is computed from summing up the sides estimated in the Hough space and the actual perimeter, as calculated by the distance between edges in the original image space. These measurements need to be approximately the same, which leads to another threshold that resulted important to frame detection.

2.3 Selection of candidate parallelograms

If the applied detector does not return any parallelogram, the photograph is classified as being a non-frame one. Yet, if one or more parallelograms are detected, their dimensions are compared to a threshold, avoiding false detections caused by small image artifacts. This threshold is established as a proportion 'p' of the image dimensions: this means that if any side of the parallelogram is smaller than the corresponding dimension of the image multiplied by 'p', the parallelogram is discarded as a frame candidate. The corresponding side can be the width or the height of the image, depending on the position of the parallelogram side.

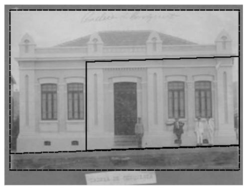
Finally, if the dimensions of any of the parallelograms are greatest than the dimension threshold, the photograph is classified as having a frame, and the selected parallelograms are considered as the candidates for being frames. These ones we called the *candidate parallelograms*.

2.4 Merging of candidate parallelograms

At this step, accepted parallelograms are combined into one by selecting the coordinates which are farthest from image borders. This comparison is made by relative position: in other words, the top-left coordinates of all the frame candidates are compared against each other, and the one with the greater values of x and y is selected, since it gives the most internal point between them. The same procedure is applied to the top-right, bottom-left and bottom-right corners, always selecting the most internal points. In most cases, the points selected this way correspond either to the smallest parallelogram or to the interception among them. This approach assumes that more external lines are frame decorations and thus the real frame is made up of the more internal detected lines. Since the smallest detected parallelograms are previously removed from the frame candidates set, it can be considered a sensible assumption. Figure 5 illustrates the whole process, showing how our approach is able to correctly identify a frame of irregular shape. This is possible because of the procedure for combining the frame candidates, selecting the points for each corner independently.

2.5 Frame extraction

The resulting shape of the previous step is returned as the detected frame. This shape is represented by its four corner points and indicates that pixels outside the area delimited by the lines of the parallelogram are considered frame pixels. Frame extraction is done by the withdrawal of these pixels of the image.



(a) 3 parallelograms are detected;

(b) Small one is removed;

(c) Final frame selected.

Figure 5: Frame detection overall process: (a) shows all the parallelograms found by the parallelogram detector; (b) shows the parallelograms that are greater than the dimension threshold; (c) the corners that are most distant from the image borders are selected for the frame.

3. EXPERIMENTAL RESULTS

The importance of experimentation is discussed in [12]. According to the author, although the usage of more refined experimentation techniques is not very widespread in Computer Science, it permits more meaningful validation of the results and to reduce uncertainties, building a reliable base of information. Also, statistical and experimental design techniques help to obtain more information from experiments and analyze them more throughly.

Thus, in this work we have used a set of well established experimental techniques in order to analyze the effectiveness of the proposed method. Firstly, a sample of 633 images was built and pre-classified as described in Subsection 2.1.

The original resolution of the images is 600 dpi. Preliminary tests with two different image sizes (300 pixels and 1200 pixels in width) showed no differences in the results, but a great difference in execution time. This was due to the computational complexity for computing the Hough transform. Hence, the experiments were based on a set with all images shrunk to a width of 300 pixels.

3.1 Used Metrics

A set of experiments was designed to answer the following questions: a) is the algorithm able to distinguish between images with frame and images without frame? b) when a detection is made, is the frame detected in the right position? c) when a false positive is detected, how much information is lost?

Four metrics are used for answering these questions: the *correct detection percentage* and the *false-positive percent* answer the first one. The correctness of the detected frame position is estimated computing the *quality ratio*, calculated from the areas of interception and union between the detected and the manually extracted frame, as follows:

$$quality\ ratio = \frac{area\ of\ interception\ between\ frames}{area\ of\ union\ of\ frames} \quad (1)$$

where the ‘frames’ above are 1) the manually identified one and 2) that automatically discovered by the method.

The quality ratio varies from 0 to 1. The closer it is to 1, the better the detection. Figure 6 illustrates the ideas behind this metric.

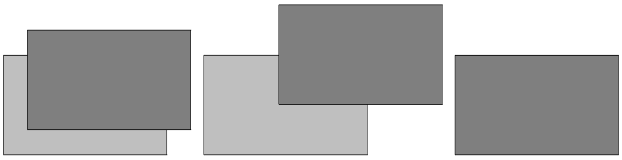


Figure 6: Schematic illustration of the quality ratio.

To assess the loss of information in case of false positives, the ratio between the external area removed as a frame and the total image area is used. This *loss ratio* gives the percentage of the image area that is lost from removing the false frame. More formally:

$$loss\ ratio = \frac{removed\ area}{total\ image\ area} \quad (2)$$

Figure 7 shows the loss ratio schematically.

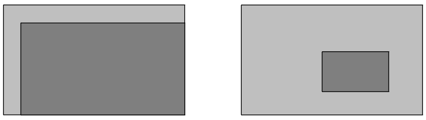


Figure 7: Schematic illustration of the loss ratio.

3.2 Algorithm Evaluation and Tuning

To compute the metrics described in Subsection 3.1, the method was applied to all the 633 sampled images, previously separated by the described categories. The quality ratio expressed by Equation 1 was computed for all images where a frame was detected and the loss ratio was computed from Equation 2 for all the false positives. All the detector parameters were kept with the default values defined by [10].

The results are shown in Table 3, where one can see a 72% correct detection rate for the significant frames. The quality ratio for these detections has an average value of 0.86, varying between 0.82 and 0.90 within a 95% confidence interval. An example of image with a quality index of 0.82 (worst case in the confidence interval) is shown in Figure 8.

These results also show that the separation into negligible and significant frames indeed leads to a more accurate analysis of the results, since there is a statistically significant difference between the detection ratios if they are considered together or not. Actually, the algorithm performs considerably better for the frames which are the most important to be removed.

Table 3: Detection and false positive ratios using default settings. The intervals between *min* and *max* are computed at a confidence of 95%. The quality ratio for the significant frames are 0.86, varying between 0.82 and 0.90. The loss ratio for the false positives is 21%, varying between 17% and 25%.

Image Category	Detec. Ratio	Min.	Max.
Significant frames	**72%**	67%	77%
Negligible frames	50%	41%	59%
Global (all frames)	65%	61%	70%
False Positives	16%	17%	19%

Figure 8: An image with a quality ratio of 0.82. The internal solid line indicates the expected frame border, the external dashed line is the frame detected automatically.

After this first evaluation of the method, a fine-tuning of its parameters was conducted as follows. First, some preliminary tests were made with about 15 images so as to identify detector parameters that have the strongest impact on detection results. From these tests, it was observed that the peak and the perimeter thresholds as explained in Section 2 were the most promising ones. Also, it was observed that applying a simple contrast expansion or an edge enhancement to the images before parallelogram detection allowed the algorithm to detect some difficult frames previously not detected.

From these results, a 2^{4-1} fractional factorial experiment [8] was designed to analyze the relative impact of those parameters and to devise a better set of values for them. For this experiment, a smaller random sample of 29 and 22 images was taken from the significant–frame and negligible–frame categories, respectively. These values were chosen so as the relative percentage for each category in the original sample was maintained. The negligible-frame images were not included in this step.

The results of this new experiment are presented in Table 4, where one can see that the 4th experiment configuration gives a good detection rate without increasing too much the false positive ratio.

However, the previously described experiments were not able to distinguish the most appropriate edge detector. Thus, a one-factor experiment [8] was designed to compare some alternatives for edge detectors. The detectors evaluated through this new experiment were: Canny, Log, Prewitt, Roberts, Sobel and Zerocross [7].

Table 4: Tuning experiment summarization. A stands for the peak threshold, B for contrast expansion, C for edge enhancement and D for perimeter threshold, while y_1 stands for the detection ratio, y_2 for the false-positive ratio, y_3 for the quality ratio and y_4 for the loss ratio.

Conf.	Parameters				Metrics			
	A	B	C	D	y_1	y_2	y_3	y_4
1	1	yes	yes	1	93%	72%	0.78	0.11
2	0.1	yes	no	0.1	79%	18%	0.85	0.14
3	1	no	yes	0.1	90%	59%	0.78	0.09
4	1	no	no	1	**86%**	**18%**	**0.82**	**0.14**
5	0.1	yes	yes	0.1	24%	9%	0.83	0.17
6	0.1	yes	no	1	41%	5%	0.90	0.29
7	1	no	no	1	62%	27%	0.79	0.08
8	0.1	no	no	0.1	28%	0%	0.9	0

It should be noted that the frame detector produces deterministic results, while one-factor experimental design relies on replications to compute experimental errors. To solve this, the workload was varied, i.e., the algorithm was applied onto three distinct samples similar to that used in the previous experiment.

The results showed that the edge detector explains 64% of the total variation for the correct detection ratio, and 48% of false positive ratio variation. Its influence on the quality and the loss ratios is negligible. The best detectors were Canny, Log and Zerocross. The effects of these detectors over the mean values are shown in Table 5, where one can see that the Canny detector gives the best relation between a high correct detection percent and a low false positive ratio.

The results of applying the algorithm with this parameter setting to the images are shown in Table 6, which gives a correct detection ratio of 87%. This represents an increase of 15% when compared to the previous result, while the increase of the false positive ratio was of 5%. The quality ratio and the loss ratio did not show statistically significant differences, thus the detection quality was maintained. All frame detection results for the aforesaid sample can be seen at http://wavelet.dcc.ufmg.br/framedetection.

4. CONCLUSION

Besides preservation purposes, digitization of historical and cultural artifacts is also aimed at providing access to this material for a wider audience. This is only feasible if adequate retrieval tools are available. Such retrieval is primarily made possible by manual indexing, but since the products of the digitization process are images, the usage of Content-Based Image Retrieval techniques is a natural path. Our group have been working in this direction, and the work described in [11] has risen the issue of the frames found in historical photographs. By means of simple a statistical analysis, we have demonstrated that these frames can indeed interfere with the image characterization and therefore degrade retrieval results.

Table 5: One-factor experiment. y_1 shows the effect of each detector over the mean detection ratio. Similar measures are y_2 for the false-positive ratio, y_3 for the quality ratio and y_4 for the loss ratio.

Detector	Metrics			
	y_1	y_2	y_3	y_4
Canny	**17%**	**20%**	**-0.03**	**0.07**
Log	11%	11%	-0.02	0.03
Zerocross	11%	11%	-0.02	0.03

Table 6: Detection and false positive ratios after tuning. Again, the intervals have a confidence of 95%. The quality ratio and loss ratio are 0.84 and 0.25 respectively, showing no statistically significant differences when compared to the first evaluation.

Image Category	Detec. Ratio	Min.	Max.
Significant frames	**87%**	83%	91%
Negligible frames	79%	71%	86%
Global (all frames)	84%	81%	88%
False Positives	21%	18%	24%

In this work, a method to find and extract those frames is proposed, then evaluated and tuned systematically. Before applying the method, a sample of the real target database is analyzed in detail, showing that about 59% of the images have frames, but 16% of them can be considered negligible from the gray level histogram point of view. This isolation of the negligible frames proved to be important for a more precise evaluation of the method.

The proposed method is comprised of four main steps: detection of parallelograms, selection of candidate parallelograms, merging of candidate parallelograms and frame extraction. Experimental results obtained show that the proposed method is able to retrieve correctly about 87% of the existing significant frames for this photographic database within a 95% confidence interval. In addition, frame coordinates are accurately identified, and the method is also able to find frames of irregular forms, due to the merging step.

The false positive ratio lies between 18% and 24%, with a loss ratio between 21% and 30%. At first sight, these values can be considered relatively high. However, it is important to observe that most of the high entropy piece of the image – the meaningful information – is located at the central portion of the picture, diminishing the impact of a false frame removal operation. Specifically, the loss of information is minimized in collections such as the one used in this study.

5. ACKNOWLEDGEMENTS

The authors are grateful to CNPq, CAPES and FAPEMIG for financial support to this work and thank Cláudio Jung for his valuable contribution.

6. REFERENCES

[1] B. T. Ávila and R. D. Lins. Efficient removal of noisy borders from monochromatic documents. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR)*, volume 3212 of *Lecture Notes in Computer Science*, pages 249–256. Springer, 2004.

[2] B. T. Ávila and R. D. Lins. A new algorithm for removing noisy borders from monochromatic documents. In *Proceedings of the ACM Symposium on Applied Computing (SAC)*, pages 1219–1225, New York, NY, USA, 2004. ACM.

[3] C. Chen, H. Wactlar, J. Z. Wang, and K. Kiernan. Digital imagery for significant cultural and historical materials.

International Journal on Digital Libraries, 5(4):275–286, 2005.

[4] A. Y. S. Chia, M. K. H. Leung, H. Eng, and S. Rahardja. Ellipse detection with Hough transform in one dimensional parametric space. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, volume 5, pages 333–336. IEEE Computer Society, 2007.

[5] E. R. Davies. *Machine Vision : Theory, Algorithms, Practicalities*. Morgan Kaufmann Publishers Inc., 2004.

[6] A. de A. Araújo, R. M. Hadad, and P. M. Jr. Identification of patterns in satellite imagery – circular forms. In *Proceedings of the IS&T/SPIE Electronic Imaging*, pages 25–34, San Jose, CA, USA, 2001. SPIE.

[7] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[8] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley and Sons, Inc., 1992.

[9] C. R. Jung and R. Schramm. Rectangle detection based on a windowed Hough transform. In *Proceedings of the XVII Brazilian Symposion on Computer Graphics and Image Processing (SIBGRAPI)*, pages 113–120, 2004.

[10] C. R. Jung and R. Schramm. Parallelogram detection using the Tiled Hough Transform. In *Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 177–180, 2006.

[11] C. J. S. Oliveira, A. de A. Araújo, J. L. S. Pio, and N. C. Batista. Photographic prints identification. In *Proceedings of the 13th International Conference on Systems, Signals and Image Processing*, pages 231–234, Budapest, Hungary, 2006.

[12] W. F. Tichy. Should computer scientists experiment more? *Computer*, 31(5):32–40, May 1998.

[13] R. S. Torres and A. X. Falcao. Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, 13(2):161–185, 2006.