

Enhancement of Layout-based Identification of Low-resolution Documents using Geometrical Color Distribution

Ardhendu Behera, Denis Lalanne, Rolf Ingold

*Department of Informatics, Chemin du Musée 3, University of Fribourg, CH-1700, Switzerland
{ardhendu.behera, denis.lalanne, rolf.ingold}@unifr.ch*

Abstract

This paper proposes a multi-signature document identification method that works robustly with low-resolution documents captured from handheld devices. The proposed method is based on the extraction of a visual signature containing both (a) the color content distribution in the image plane of the document, i.e. the color signature, and (b) the shallow layout structure of the document, i.e. the layout signature. The color distribution is first considered, in order to filter documents with very dissimilar colors, and the identification is finally done on the remaining set using the layout signature. An evaluation, that compares our color and layout-based method with the layout signature alone, is finally presented.

1. Introduction

Many systems use printed documents as an interface to access other media using image analysis [1], bar-codes [2], document/speech alignment [3] or slideshow temporal segmentation [4]. In all these systems, documents are associated with the captured multimedia streams and play a central role since they help structure the access to multimedia archives. In this paper, we propose a document image-based retrieval method. The goal is to identify low-resolution documents, often distorted and associate them with their original electronic form. Once identified, electronic documents could then be associated with all the media in which they have been visible.

Often, documents are captured using low-resolution mobile handheld devices. Such images could then be queried on a system in order to retrieve associated multimedia documents. The identification corresponds to the maximum similarity between the captured documents and the original electronics documents. The similarity is generally measured either based on the layout and/or content of the documents. The layout-based matching considers geometrical features, logical structure, or both [5][6]. The texture-based approach is

also considered for identifying the various components of a document (e.g. text, images, graphics, etc.) [7]. However, in such cases, the input documents need to be of high quality, i.e. at least 300 *dpi*.

In this article, we present an identification method for documents captured from low-resolution handheld devices. Color is a low-level feature that has been rarely incorporated along with the layout features for document identification and could be used efficiently for the classification of form-type documents [5]. The proposed method targets slides identification, i.e. documents with a limited textual content and with a variable layout. Identified slides could later be used for retrieving the content of a conference talk, a meeting, a lecture, etc. Slide identification has been already tackled using OCR, text layout and pixel-by-pixel matching of binary images [8][4]. However, such systems need at least some textual content and a uniform background which are not always the case in a slide. The proposed method uses both the geometrical distribution of colors and the layout structure information in the document.

2. Preprocessing of captured documents

The snapshot of the projection screen taken from a handheld device not only contains the projected documents but also the background. It is thus necessary to remove the background and to rectify the skewing of the remaining document image. The handheld capture devices are assumed to have low radial distortion. Therefore, one needs to consider the four corners of the quadrangle of the projected part and is mapped to a rectangle of common resolution. This is done using a perspective transformation matrix and then bi-cubic interpolation [4]. Currently, the corners of the projected part are selected manually with a graphical user interface. If the capture device remains at the same position, the calibration has to be done only once and then the same quadrangle corners and transformation matrix could be used for all the captured images. Finally, the noise in the rectified image is removed using *Weiner filter* [9] with a



Figure 1. Original, captured and rectified image of the slide document (left-to-right)

window size of 10×10 applied to each of the RGB channel.

3. Document Color Signature

Each captured document as well as each original electronic document is represented by a visual signature which facilitates their matching. Once the preprocessing has been done for a captured document, then the color distribution is first computed and followed by its layout structure. The procedure is similar for original electronic documents but without any preprocessing. The document color signature corresponds to the geometrical distribution of colors in the 2-D plane of the document image.

3.1 Geometrical distributions

Although the colors of the captured images are distorted due to changes in the lighting environment or even of capture devices, nevertheless, the geometrical distributions of the color in the image plane remain preserved. Therefore, the geometrical color distribution is given a higher priority than the spatial distributions in any one of the color space. First, the pixels of similar color in the RGB color space are grouped using the K-Means clustering algorithm with use of the *Mahalanobis* distance [10]. The assignment of the number of cluster K is necessary before starting the clustering. This value is derived from the number of predominant peaks in the RGB color histograms. An adaptive assignment of K could be used, starting with a minimum value of $K = 2$ and adaptively increase the number of cluster till the improvement in error falls below a threshold or a maximum number of clusters is reached. The drawback of this adaptive clustering is the processing time. Furthermore, the clustering time for a given data set and the number of clusters are dependent on the initialization of the clusters centroids. Often, this initialization is done with random seeding. In our case, both the number of clusters and the clusters centroids are derived from the color histogram and thus, the processing time is extremely fast in comparison with random seeding or adaptive clustering.

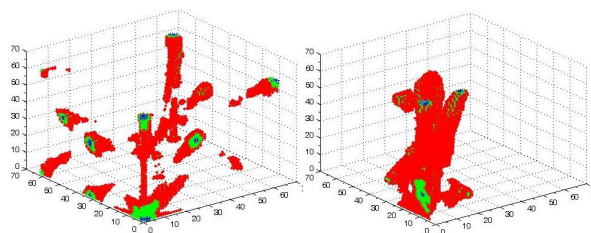


Figure 2. Histogram for peaks detection of original (left) and captured slides of Figure 1 in reduced RGB space

The K value is derived from the reduced color histogram. The standard method for creating an RGB color histogram is to consider the m higher order bits of each channel and then accumulate the pixels having similar color values in the 2^{3m} bins. If human vision is considered, then there is a large amount of redundancy in 24-bit RGB representation of color images. Wang *et al.* have reported that representing each of the RGB channel with only 4 bits introduced little or even no perceptible visual degradation [11]. Furthermore, the aim here is to estimate the number of dominant colors (peaks). For this reason, a $64 \times 64 \times 64$ ($m = 6$) color image is considered for generating the histogram rather than a true color image $256 \times 256 \times 256$ ($m = 8$). This is achieved by simply performing a 2-bit right-shifting on each RGB channel. The resulting histogram is smoothed using the 3-D *Gaussian* window. Subsequently, the numbers of predominant peaks are located and those peaks are chosen if the distance between two consecutive peaks is superior to certain threshold (Figure 2). Once the number of effective peaks are decided then the centroids of each cluster is initialized with the average RGB values of the pixels closest to the respective peak. Then, pixels are grouped using the K-Means. With this initialization, we observed that the K-means took much less time than with a random initialization. For an image of size 720×540 and $K = 5$, the clustering takes less than 3 iterations for a convergence rate of 99%, while the random initialization takes more than 15 iterations to converge. Furthermore, judging from the distances between the clusters' centroid positions in the captured and original image, we observed that the histogram-based initialization gives smaller distances than the random initialization for the same number of clusters and convergence rate.

3.2 Color features extraction

The color features corresponding to the geometrical distribution of each cluster i in the 2-D image plane are computed as follows (center of cluster $C_{x,i}$ and $C_{y,i}$, variance $V_{x,i}$ and $V_{y,i}$):

$$\left. \begin{aligned} C_{x,i} &= \frac{1}{N_i} \sum_{\forall p \in i} X_{pos}(p), C_{y,i} = \frac{1}{N_i} \sum_{\forall p \in i} Y_{pos}(p) \\ V_{x,i} &= \frac{1}{N_i} \sum_{\forall p \in i} (X_{pos}(p) - C_{x,i})^2 \\ V_{y,i} &= \frac{1}{N_i} \sum_{\forall p \in i} (Y_{pos}(p) - C_{y,i})^2 \end{aligned} \right\} i = 1 \dots K$$

The $X_{pos}(p)$ and $Y_{pos}(p)$ are the horizontal and vertical location of pixel p , respectively and N_i is the number of pixels per cluster i . The other features included in the color signature are the cluster density $d_i = N_i / total\ pixels$, the statistical properties of the color distribution, such as the mean ($M_{r,i}, M_{g,i}, M_{b,i}$) and variance values ($V_{r,i}, V_{g,i}, V_{b,i}$) for each channel of the RGB space per cluster. The peak in the histogram (Figure 2) having the highest value is considered as being the background color since the number of pixels in a uniform background is generally significantly greater than that of the foreground, which is obvious in documents such as slides.

The similarity between two images is often measured by computing the distance between their respective histograms. The *Minkowski distance* [10] is generally used to measure this distance and is known to perform well for images of same size with negligible color distortions [12]. In this scenario of captured documents, one faces a problem of non-uniform color shifting in the captured image as compared to the original image. This shift of color is due to the presence of color cast, which is the predominant superimposed color and is due to variations in the lighting conditions or to the capture device properties. As illustrated on Figure 2, this causes a shift of the peaks and valleys in the histogram of the captured image in comparison to the original one and thus the standard histogram-based similarity distance would not perform efficiently.

4. Documents layout signature

The resolution of the captured documents is very low for the extraction of the complete layout structure, i.e. of both physical and logical structures. Indeed, the average size of the projected part is of 450×560 and a resolution of below $75\ dpi$. For this reason, the extracted layout structure is shallow and close to the perception of human vision. The physical structure is first extracted; physical blocks are then labeled as text, image, bullets, bars, etc. and hierarchically structured to form the *Layout Signature* of the documents. This signature is particularly appropriate for slideshows, since often the textual content of the slides is limited whereas the layout structure varies significantly in

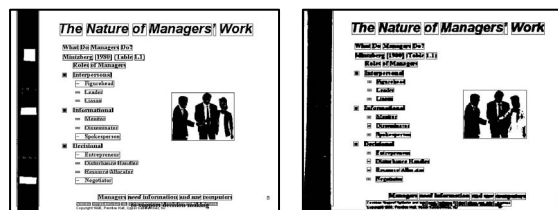


Figure 3. Layout signatures: bounding boxes for each feature of the original slide and its corresponding captured slide image.

comparison to other types of documents (e.g. newspaper, articles, etc.). The detailed extraction of this layout signature is explained by Behera *et al.* [13]. Figure 3 illustrates the layout signature of a document image, i.e. the bounding boxes of each layout feature, such as text blocks, lines, words, graphics, bullets, etc.

5. Matching of signatures

The similarity between two documents is computed using their corresponding signatures. Before starting the matching procedure, the visual signature of each original electronic document in the repository, is computed. The captured image is preprocessed and the corresponding signature is also computed. This signature is then queried for finding a matching one in the repository. The color signature is first used for filtering the space, and to get a set of the signatures having similar color distributions and finally the layout signature is used in order to find the original document.

5.1 Color features matching

The matching algorithm takes into account the number of clusters and the geometrical features of each cluster in both the original and queried signature. It has been mentioned earlier that the cluster of the highest density represents the background color and the others, the foreground.

5.1.1 Merging clusters to imitate color cast in the original image. In most cases, the number of clusters in the captured image is inferior to that of the original. This is due to the presence of color cast in the captured image, which provokes more convergence in the color histogram of the captured image, i.e. adjacent colors are often brought closer (Figure 2). The idea is to reproduce the geometrical distribution of the clusters of the captured image in the original image, by merging the clusters in the original signature so that the number of clusters and their respective geometrical centroids in both the capture and original images are brought closer. Separating the clusters of the captured

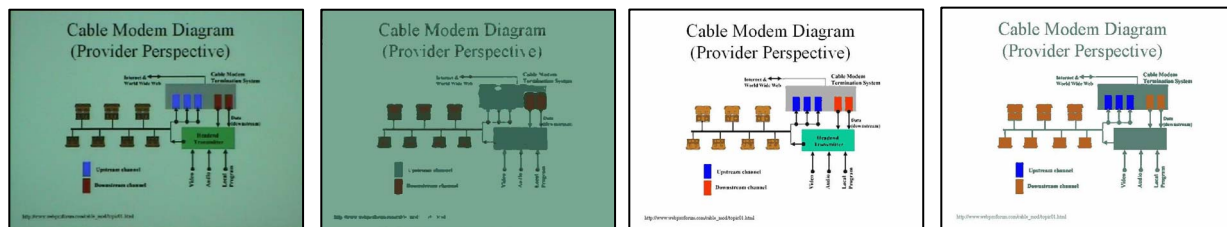


Figure 4. Left-to-right a) pre-processed captured slide, b) clustering (4 clusters), c) clustering of the original image (8 clusters) and d) merging of clusters of original image (4 clusters)

image rather than merging in the original image is not feasible, since the locations of pixels are not stored in the color signature. Let K and M be the number of clusters in the captured and the original image ($M > K$), two or more clusters in the original image are to be merged if and only if the following conditions are satisfied: a) the sum of the cluster densities of the merged clusters and b) the resulting geometrical centroid of the merged clusters is close to the respective cluster density and geometrical centroid of the captured image. Let the p^{th} and q^{th} cluster of the original image are to be compared with the i^{th} cluster of the captured image. The two clusters (p & q) are merged only if the conditions $\|d_i - d_j\| < T_d$ and $\|C_{x,i} - C_{x,j}\| + \|C_{y,i} - C_{y,j}\| < T_c$ are satisfied. Subsequently, the cluster properties are updated.

$$d_j = (N_p + N_q) / N, \quad 1 \leq p, q \leq M, \quad N = \# \text{total pixels}$$

$$C_{x,j} = (N_p C_{x,p} + N_q C_{x,q}) / (N_p + N_q)$$

$$C_{y,j} = (N_p C_{y,p} + N_q C_{y,q}) / (N_p + N_q)$$

$$V_{x,j} = (N_p V_{x,p} + N_q V_{x,q}) / (N_p + N_q)$$

$$V_{y,j} = (N_p V_{y,p} + N_q V_{y,q}) / (N_p + N_q)$$

Figure 4 is an example of such merging of clusters. Before merging, there are 8 clusters in the original image and only 4 in the captured image. The clusters in the original image are then merged to 4 clusters for the comparison with the captured images. While looking at Figure 4, it is clear that after merging, the geometrical distribution of colors in the original image is much closer to that in the captured image than before.

5.1.2 Color signature matching: comparing clusters properties. The clusters are compared in an ascending order of their densities ($d_1 \leq d_2 \leq \dots \leq d_K$). The similarity distance, between each cluster geometrical properties, in the original and in the captured image, is then computed. If the sum of the distances is inferior to a certain threshold, then the original image is kept in a set of eligible solutions, i.e. original documents that will be further compared using the layout signature. The threshold value is set conservatively, so that the rejection rate is null, i.e. the document to be identified is not rejected. The final identification is ultimately done using the layout signature.

5.2 Layout features matching

The layout feature-based matching is performed on the remaining set of signatures after matching of the color signature. The score is computed at each of the feature node (text, image, bullets, bars, etc.) of the layout signature by comparing the number of elements and their geometrical properties. The features are prioritized according to their frequency of appearance in the documents. The features with higher priority are most reliable, which require less analysis. The weighted sums of the features are computed and compared. The signature having the highest score is picked up and corresponds to the identified document. The detailed procedure for the assignment of weight to the features and the computation is explained in [13].

6. Evaluation and results

The evaluation of the proposed method has been performed by querying 355 slides from 16 different slideshows on a repository containing 2000 slides from 60 different slideshows. In this evaluation, we assume that the queried documents are already present in the repository. The evaluation metric used are the identification rate, I and the rejection rate, R :

$$I = \frac{\# \text{ correct documents retrieved}}{\# \text{ total documents queried}}$$

$$R = \frac{\# \text{ documents rejected}}{\# \text{ total documents queried}}$$

In Table 1, the average identification result for each slideshow is shown. The last row of the table represents the average values of all 16 slideshows. Table 1 shows that not only the identification rate of the color and layout (89%) is better than the layout alone (79%) but also the search space is reduced, drastically (16%). This is due to the prior filtering with the color signature, which results in the reduction of the processing time (1.2 sec). In last two slideshows, the identification rate using the layout signature alone is the lowest (35% and 37%), which is due to the non-uniform (vertical color gradient) background. In the

layout signature, the dark part of the background is segmented as an image, which is the same for all slides in the slideshow. In the case of the combined method, the color signature approximated the color gradient with 3 different clusters. The centroids of these clusters are different due to the variation of the foreground objects, which helps in identification. This is a very encouraging result for the handling of the non-uniform background. Furthermore, for the slides having non-uniform (textured) or gradient variation of the background, often the layout features consider the whole document as a single image block, which considerably reduces the matching performance whereas this drawback is overcome by the color feature. The rejection rate is nearly 2% and upon analysis of the results for such unidentified documents, it is observed that those are too noisy for such comparison and is due to the non-uniform lighting condition. The evaluation has been performed on a 1.7 GHz Pentium 4 PC.

Table 1. Result of the proposed identifications

Slide Shows #Slides	Layout only (Average)				Color + Layout (Average)			
	Search space	I	R	Time Sec.	Search Space	I	R	Time Sec.
34	1.00	0.83	0.00	4.11	0.32	0.92	0.00	1.79
10	1.00	0.60	0.00	4.07	0.01	0.90	0.10	0.48
15	1.00	0.75	0.00	4.02	0.02	0.75	0.00	0.51
28	1.00	1.00	0.00	4.20	0.19	1.00	0.00	1.24
30	1.00	0.88	0.00	4.05	0.20	0.96	0.00	1.23
24	1.00	0.86	0.00	3.93	0.18	0.86	0.07	1.14
19	1.00	1.00	0.00	4.18	0.20	1.00	0.00	1.70
28	1.00	0.96	0.00	4.10	0.18	0.96	0.04	1.37
25	1.00	0.76	0.04	4.01	0.17	0.80	0.12	1.35
20	1.00	0.82	0.12	4.06	0.11	0.94	0.00	1.23
29	1.00	0.79	0.00	4.05	0.15	0.98	0.00	1.49
17	1.00	1.00	0.00	3.99	0.20	1.00	0.00	1.24
15	1.00	1.00	0.00	3.96	0.23	1.00	0.00	1.34
16	1.00	0.71	0.00	3.91	0.03	0.75	0.14	0.63
20	1.00	0.35	0.00	3.59	0.13	0.72	0.00	1.15
25	1.00	0.37	0.00	4.07	0.19	0.67	0.00	1.33
355	1.00	0.79	0.02	4.02	0.16	0.89	0.03	1.20

7. Conclusion and future works

The method proposed here, identifies documents captured from low-resolution handheld devices based on a visual signature composed of both (a) the geometrical distribution of the color, i.e. the color signature and (b) the shallow layout structure of the documents. Finally, an evaluation of this visual signature has been presented. The combination of both features improves recognition rate and saves a lot of computational time. In the near future, we plan to

develop a method in order to calibrate colors automatically and an identification method which considers only the color features for improvement in performance.

8. References

- [1] L. Nelson, S. Ichimura, *et al.*, "Palette: A paper interface for giving presentations", *Proc. of CHI*, 1999, pp.354-361.
- [2] J. Graham, and J. J. Hull, "A Paper-Based Interface for Video Browsing and Retrieval", *Proc. of ICME*, Baltimore, MD, July 6-9, 2003.
- [3] D. Lalanne, R. Ingold, D. Rotz, *et al.* "Using static documents as structured and thematic interfaces to multimedia meeting archives", *1st Int'l Workshop on MLMI*, 2004, Martigny, Switzerland, LCNS, vol. 3361, pp. 87-100.
- [4] S. Mukhopadhyay, and B. Smith, "Passive Capture and Structuring of Lectures", *Proc. of ACM Multimedia*, 1999, pp. 477-487.
- [5] P. Duygulu, and V. Atalay, "A hierarchical representation of form documents for identification and retrieval", *IJDAR*, 2002, Vol. 5, pp.17-27.
- [6] R. Haralick, "Document Image Understanding: geometric and logical layout", *Proc. IEEE Conf. CVPR*, 1994, vol. 8, pp. 385-390.
- [7] J. F. Cullen, J. J. Hull, and P. E. Hart, "Document Image Database Retrieval and Browsing using Texture Analysis", *Proc. of ICDAR*, 1997, pp. 718-721.
- [8] B. Erol, and J. Hull "Linking Presentation Documents Using Image Analysis", *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 9-12, 2003.
- [9] Lim, J. S., *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice Hall, 1990.
- [10] D. Zhang, and G. Lu, "Evaluation of Similarity measurement for image retrieval", *Intl. Conf. NNSP*, Nanjing, China, May 30-June3, 2003.
- [11] B. Wang, X. -F. Li, F. Liu, and F. -Q. Hu, "Color text image binarization based on binary texture analysis", *Proc. of ICASSP*, Montreal, Quebec, Canada, May 17-21, 2004.
- [12] M. Petkovic', "Content-based video retrieval", *7th Int'l Conf. on Extending Database Technology*, March 27-31, 2000, Konstanz, Germany, pp 74-77.
- [13] A. Behera, D. Lalanne and R. Ingold "Visual Signature based Identification of Low-resolution Document Images," *ACM Symposium on Document Engineering*, Milwaukee, Wisconsin, 2004, pp. 178-187.