Extraction and Classification of Handwritten Annotations

Andrea Mazzei
CRAFT - EPFL
Rolex Learning Center
Station 20
CH-1015 Lausanne
andrea.mazzei@epfl.ch

Frédéric Kaplan
CRAFT - EPFL
Rolex Learning Center
Station 20
CH-1015 Lausanne
frederic.kaplan@epfl.ch

Pierre Dillenbourg
CRAFT - EPFL
Rolex Learning Center
Station 20
CH-1015 Lausanne
pierre.dillenbourg@epfl.ch

ABSTRACT
This article describes a method for extracting and classifying handwritten annotations on printed documents using a simple camera integrated in a lamp or a mobile phone. The ambition of such a research is to offer a seamless integration of notes taken on printed paper in our daily interactions with digital documents. Existing studies propose a classification of annotations based on their form and function. We demonstrate a method for automating such a classification and report experimental results showing the classification accuracy.

Author Keywords
machine-printed and handwritten text separation, Document processing, annotation classification

ACM Classification Keywords
H.5.2 Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Algorithms, Experimentation, Human Factors, Languages, Measurement, Performance, Reliability, Theory

INTRODUCTION

Annotating and taking notes on paper is a very common practice in our daily routines. Readers write comments in the margins of papers, underline important passages and use other various marking strategies. These practices help them to understand better what they read and, at a later stage, find back easier relevant passages. It plays also an important role for associative thinking and linking the content with other ideas and documents. Despite the efforts to transfer annotating practices to digital documents, annotating on paper has many advantages compared to any electronic equivalent (Kawase et al. [6]).

This article describes a method for extracting and classifying handwritten annotations on printed documents using a simple camera integrated in a lamp or a mobile phone. The ambition of such a research is to offer a seamless integration of notes, taken on printed paper in our daily interactions with digital documents. Handwritten annotations have different forms and functions (Marshall [8]). We highlight or underline words as attentional landmarks. We write short notes within the margins or between lines of text as interpretation cues. We use longer notes in blank spaces or near figures to elaborate with complementary information. The system, described in this article, aims at not only extracting handwritten annotations, but also classifying them in one of these three categories based on their spatial and colourimetric properties. This automatically generated classification could then be used to sort, organize and share annotations, for instance, in the context of collaborative reading applications.

The first part of the article reviews a number of systems that have been investigated in the last 20 years to tackle this issue. The second part presents our own contribution as original combination of a technique for extracting annotations, a clustering algorithm and a classification approach. To the best of our knowledge the method herein described has not been applied to this problem beforehand. We report the results of a preliminary study showing that handwritten annotations can be extracted and classified in a satisfactory manner using this technique.

MACHINE-PRINTED AND HANDWRITTEN TEXT CLASSIFICATION: A SHORT REVIEW

Discriminating machine-printed and handwritten text in textual images is a problem that has been intensely investigated in the last two decades. In 1990 Umeda and Kasuya [14] described their discriminator of English characters. Their patented invention is based on the strong assumption of uniformity of each block. The discrimination is performed by calculating the ratio between the number of slanted strokes and the sum of horizontal, vertical and slanted ones and by imposing a predetermined static threshold. Under these conditions they achieved a recognition rate of 95%.

Few years later two works focused on the classification at character level. Kunuke et al. [7] proposed a classification methodology based on the extraction of scale and rotation invariant features: the straightness of vertical and horizontal lines and the symmetry relative to the centre of gravity of the character. Their results showed a recognition rate of

UbiComp '10, Sep 26-Sep 29, 2010, Copenhagen, Denmark.
ACM 978-1-60558-843-8/10/09.

96.8% on a training set of 3632 and 78.5% on a test set of 1068 images; Fan et al. [2] used instead the character block layout variance. They reported a correctness rate above 85% tested on English and Japanese textual images: 25 images containing machine printed text and 25 containing handwritten ones. In 2000 Pal et al. [11] presented their method for Bangla and Devnagari; it relies on the analysis of some structural regularities of the alphabetic characters of these languages. Their method uses a hierarchy of three different features to perform the discrimination. The head line is the predominant feature, in fact it forms a peak in the horizontal projection profile of machine-printed text. Their recognition rate is attested on 98.6%. Guo et al. [3] suggested a method based on a hidden Markov model to classify typewritten and handwritten words based on vertical projection profiles of the word. They tested the algorithm on a test-set of 187 words, reaching a precision rate of 92.86% for the typewritten words and 72.19% for the handwritten ones.

More recently Zheng et al. [16] reported a work on a robust printed and handwritten text segmentation from extremely noisy document images. They used different classifiers such as k-nearest neighbours, support vector machine (SVM) and Fischer and different features such as pixel density, aspect ratio and Gabor filter. They achieved a segmentation accuracy of 78%. In the meanwhile Jang et al. [4] described an approach, specific for Korean text, based on the extraction of geometric features. They employed a multilayer perceptron classifier reaching an accuracy rate of 98.9% on a test-set of 3,147 images. On the other hand Kavallieratou [5] showed that a simple discriminant analysis on the vertical projection profiles performs comparably to many robust approaches.

One interesting application is the detection and matching of signatures proposed by Zhu et al. [17], a robust multilingual approach, in an unconstrained setting of translation, scale, and rotation invariant nonrigid shape matching. Peng et al. [12] suggested a novel approach based on three categories of word level feature and a k-means classifier associated with a relabelling post-procedure using Markov random field models; they achieved an overall recall of 96.33%. And finally in a more general scenario of sparse data and arbitrary rotation Chanda et al. [1] recently described their approach based on the SVM classifier and obtaining an accuracy of 96.9% on a set of 3958 images.

METHOD

We here present our approach for extracting and classifying handwritten annotations on machine printed documents. Figure 1 provides an overview of the processing pipeline. It consists of four steps. The first step takes in input the image containing the already extracted annotations and proceeds by clustering the pixels. Parallely the retrieved digital source of the document is processed in order to acquire an accurate estimation of the bounding boxes around the main text blocks present in the image. The set of classified annotations and the estimated bounding box are given in input to a decision tree classifier. A final step is responsible for evaluating the accuracy of the classification by comparing the average colour of each annotation with the predetermined ones.

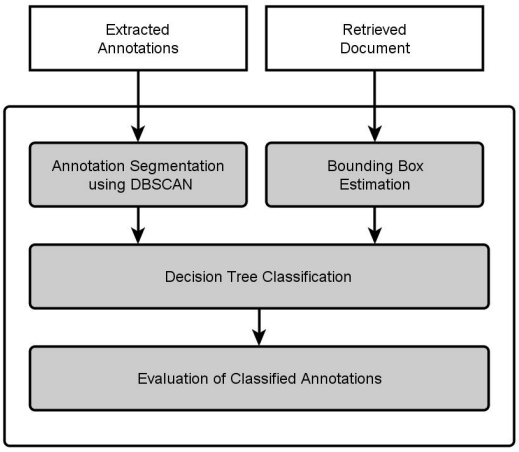


Figure 1. Processing pipeline

Annotation Extraction using Background Subtraction

A novel approach to separate handwritten annotations from machine-printed text is described by Nakai et al. [10]: they realized a method able to extract colour annotations from colour documents. Their method is based on two tasks: fast matching of document images based on local arrangement of features points and flexible background subtraction resistant to moderate misalignment. This method is more general than the above-mentioned ones, since it deals with any kind of annotation and printed document. Later improvements by the same authors [9] showed an accuracy rate of 85.59%. These results encouraged us to adopt their method.

Annotation Segmentation using DBSCAN

This module is responsible for grouping the colour pixels constituting the image containing the extracted annotations. To address this issue we decided to adopt the well known clustering algorithm DBSCAN (Density-Based Spatial Clustering of Application with Noise) for the following reasons:

- the pixels forming an annotation are subject to the conditions of spatial adjacency and colourimetric proximity
- the number of clusters is not known a priori: the number of annotations contained in a page is not predictable
- position, orientation, size and colour of an annotation are variable
- the algorithm should not have a bias toward a particular cluster shape and it should handle noise: the form of an annotation can vary from the rectangular highlighted region to the arbitrary handwritten mark
- the algorithm should distinguish adjacent or even self containing clusters: for instance the highlighted comments

Wu et al. recently reported significant improvements of the original DBSCAN algorithm in terms of time complexity [15]; they removed the original inadequacy in dealing with large-scale data. This allows us not to be bound up with low resolution images.

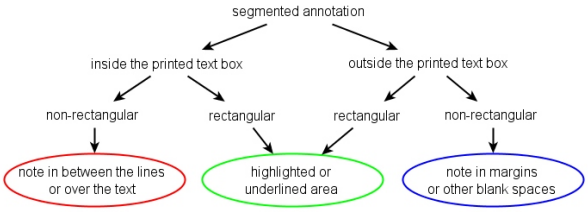


Figure 2. Decision tree classification

The input image containing the pre-extracted annotations is reprocessed. Each pixel is specified by 5 components:

$$p_i = (x_i, y_i, r_i, g_i, b_i) \quad (1)$$

the local position x_i and y_i, used as indexing terms, and the colour information r_i, g_i and b_i, which yields additional discriminative power. The output is obtained by partitioning this set of n pixels into a set of k clusters:

$$A = (A_1, A_2, ..., A_k) \quad (2)$$

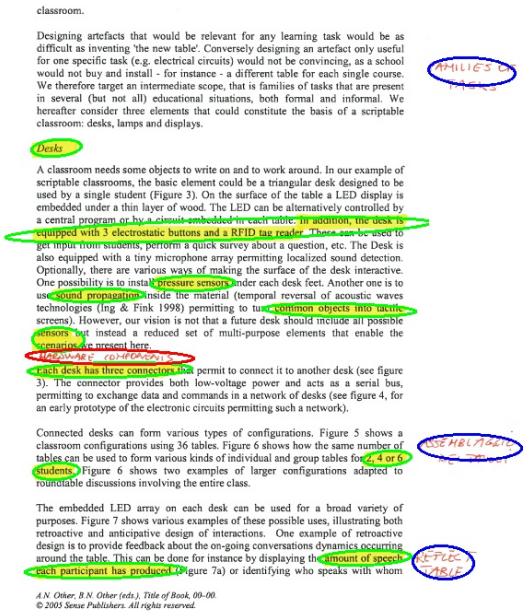
Each cluster corresponds to a correctly segmented annotation. The centroid contains the position of the centre of mass and the mean colour of the annotation. The algorithm is initialized by setting two radiuses, ϵ_{pos} for the spatial domain and ϵ_{rgb} for the colourimetric one and a minimum density $MinPts$ to discriminate all the pixels in core, density reachable and noise points.

Decision Tree Classification

A classification of different forms of annotation is analyzed by Marshall [8]; we regroup the discussed marking strategies by functionality: *memory recall* for underlined or highlighted elements, *interpretation cues* for symbols and short notes in between the lines or over the text, *contents elaboration* for notes in margins or other blank spaces.

We use a decision-tree-based classifier to map the clustered annotations into these categories. Figure 2 illustrates the structure of the decision tree and defines the annotation types in the leaf nodes. In the first level all the annotations are discriminated according to their local position on the page: annotations in between the lines or over text and annotations in the margins or other blank spaces. In the second level all the annotations are separated according to their rectangularity; some methods to compute this derived feature are proposed by Rosin [13]; these methods have desirable properties for our scenario such as rotation invariance and robustness to noise.

The rectangularity is calculated using the minimum bounding rectangle (MBR). More precisely the MBR can be calculated on the elliptical approximation of the shape of interest. Each value of rectangularity is then thresholded to separate more compact annotations such as highlighted areas from others with more complex boundaries such as notes and symbols. Figure 3 shows a satisfactory classification result. In this figure the red, green and blue ellipses contain the notes between the lines or over the text, highlighted passages and notes in the blank spaces respectively.


classroom.

Designing artefacts that would be relevant for any learning task would be as difficult as inventing 'the new table'. Conversely designing an artefact only useful for one specific task (e.g. electrical circuits) would not be convincing, as a school would not buy and install - for instance - a different table for each single course. We therefore target an intermediate scope, that is families of tasks that are present in several (but not all) educational situations, both formal and informal. We hereafter consider three elements that could constitute the basis of a scriptable classroom: desks, lamps and displays.

Desks

A classroom needs some objects to write on and to work around. In our example of scriptable classrooms, the basic element could be a triangular desk designed to be used by a single student (Figure 3). On the surface of the table a LED display is embedded under a thin layer of wood. The LED can be alternatively controlled by a central program or by a circuit embedded in each table. In addition, the desk is equipped with 3 electrostatic buttons and a RFID tag reader. These can be used to get input from students, perform a quick survey about a question, etc. The Desk is also equipped with a tiny microphone array permitting localized sound detection. Optionally, there are various ways of making the surface of the desk interactive. One possibility is to install pressure sensors under each desk feet. Another one is to use sound propagation inside the material (temporal reversal of acoustic waves technologies (Ing & Fink 1998) permitting to turn common objects into tactile screens). However, our vision is not that a future desk should include all possible sensors but instead a reduced set of multi-purpose elements that enable the scenarios we present here.

Each desk has three connectors that permit to connect it to another desk (see figure 3). The connector provides both low-voltage power and acts as a serial bus, permitting to exchange data and commands in a network of desks (see figure 4, for an early prototype of the electronic circuits permitting such a network).

Connected desks can form various types of configurations. Figure 5 shows a classroom configurations using 36 tables. Figure 6 shows how the same number of tables can be used to form various kinds of individual and group tables for 2, 4 or 6 students. Figure 6 shows two examples of larger configurations adapted to roundtable discussions involving the entire class.

The embedded LED array on each desk can be used for a broad variety of purposes. Figure 7 shows various examples of these possible uses, illustrating both retroactive and anticipative design of interactions. One example of retroactive design is to provide feedback about the on-going conversations dynamics occurring around the table. This can be done for instance by displaying the amount of speech each participant has produced (Figure 7a) or identifying who speaks with whom

A.N. Other, B.N. Other (eds.), Title of Book, 00-00.

Figure 3. Annotation classification output

EXPERIMENTAL RESULTS

We have collected 33 annotated pages of scientific articles containing a total of 571 annotations produced by a culturally heterogeneous group of Master and PhD students. They produced the annotations in their own native languages and using their personal style. We set only one constraint: we asked them to use the same colours for each type of annotation within one page. This constraint is imposed only to automatically and objectively evaluate the accuracy of our approach. For each page we supervised the last step of the pipeline (Figure 1) indicating the corresponding function of each colour used for annotating. The experimental results show a classification accuracy of 84.47%.

Strengths and Weaknesses

We here report the observed strengths and weaknesses. The adopted method for extracting annotations from printed documents and the ones discussed in the literature review introduce noise in the separation. DBSCAN effectively identifies and handles these noise pixels. We now report some relevant cases of correct and robust classification and cases of failures. Figure 4(a) shows a difficult scenario in which our approach correctly classifies the annotations. An interline comment is between two highlighted words: in this specific case the spatial information is not discriminative enough to distinguish them: the colour information is determinant to perform the separation. Another strength is that our approach does not depend on a specific language. Figure 4(b) shows a case of correct classification of a note written in Iranian. Figure 4(c) shows a case of correct clusterization but incorrect classification. The big red ellipse contains a chain of bordering highlighted regions. This region is clustered as a set of homogeneous annotations but wrongly classified as interline note because of a wrong value of rectangularity.

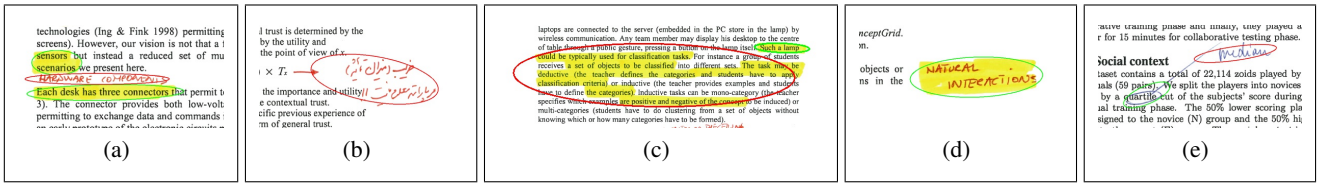


Figure 4. Cases of robust, poor and wrong classification

Figure 4(d) shows a case of self contained annotations. In this case the red ink diffuses into the highlighter ink creating a colour transition between them. This leads to a rough clusterization result. Lastly our approach is not well-suited to capture the notion of linking as shown in Figure 4(e).

CONCLUSION

In this paper we propose a system for clustering and classifying handwritten annotations extracted using already existing techniques. Although there is room for improvements using this approach, the results are promising enough to extend the investigation to a more accurate and granular classification.

REFERENCES

1. S. Chanda, K. Franke, and U. Pal. Structural handwritten and machine print classification for sparse content and arbitrary oriented document fragments. In *SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 18–22. ACM, 2010.

2. K.-C. Fan, L.-S. Wang, and Y.-T. Tu. Classification of machine-printed and handwritten texts using character block layout variance. *Pattern Recognition*, 31(9):1275 – 1284, 1998.

3. J. K. Guo and M. Y. Ma. Separating handwritten material from machine printed text using hidden markov models. In *ICDAR '01: Proceedings of the Sixth International Conference on Document Analysis and Recognition*, page 439. IEEE Computer Society, 2001.

4. S. I. Jang, S. H. Jeong, and Y.-S. Nam. Classification of machine-printed and handwritten addresses on korean mail piece images using geometric features. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2*, pages 383–386. IEEE Computer Society, 2004.

5. E. Kavallieratou, S. Stamatatos, and H. Antonopoulou. Machine-printed from handwritten text discrimination. *Frontiers in Handwriting Recognition, International Workshop on*, 0:312–316, 2004.

6. R. Kawase, E. Herder, and W. Nejdl. A comparison of paper-based and online annotations in the workplace. In *EC-TEL '09: Proceedings of the 4th European Conference on Technology Enhanced Learning*, pages 240–253. Springer-Verlag, 2009.

7. K. Kuhnke, L. Simoncini, and Z. M. Kovacs-V. A system for machine-written and hand-written character distinction. In *ICDAR '95: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 2)*, page 811, Washington, DC, USA, 1995. IEEE Computer Society.

8. C. C. Marshall. Annotation: from paper books to the digital library. In *DL '97: Proceedings of the second ACM international conference on Digital libraries*, pages 131–140. ACM, 1997.

9. T. Nakai, K. Iwata, and K. Kise. Accuracy improvement and objective evaluation of annotation extraction from printed documents. In *Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on*, pages 329–336, 2008.

10. T. Nakai, K. Kise, and M. Iwamura. A method of annotation extraction from paper documents using alignment based on local arrangements of feature points. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 1, pages 23–27, 2007.

11. U. Pal and B. B. Chaudhuri. Automatic separation of machine-printed and hand-written text lines. *Document Analysis and Recognition, International Conference on*, 0:645, 1999.

12. X. Peng, S. Setlur, V. Govindaraju, R. Sitaram, and K. Bhuvanagiri. Markov random field based text identification from annotated machine printed documents. In *ICDAR '09: Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, pages 431–435. IEEE Computer Society, 2009.

13. P. L. Rosin. Measuring shape: ellipticity, rectangularity, and triangularity. *Mach. Vision Appl.*, 14(3):172–184, 2003.

14. T. Umeda and S. Kasuya. Discriminator between handwritten and machine-printed characters, 1990.

15. Y.-P. Wu, J.-J. Guo, and X.-J. Zhang. A linear dbscan algorithm based on lsh. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 5, pages 2608 –2614, 2007.

16. Y. Zheng, S. Member, H. Li, and D. Doermann. Machine printed text and handwriting identification in noisy document images. *IEEE Trans. Pattern Analysis Machine Intelligence*, 26:2003, 2004.

17. G. Zhu, Y. Zheng, D. Doermann, and S. Jaeger. Signature detection and matching for document image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:2015–2031, 2009.