Cognitive and social effects of handwritten annotations

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

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

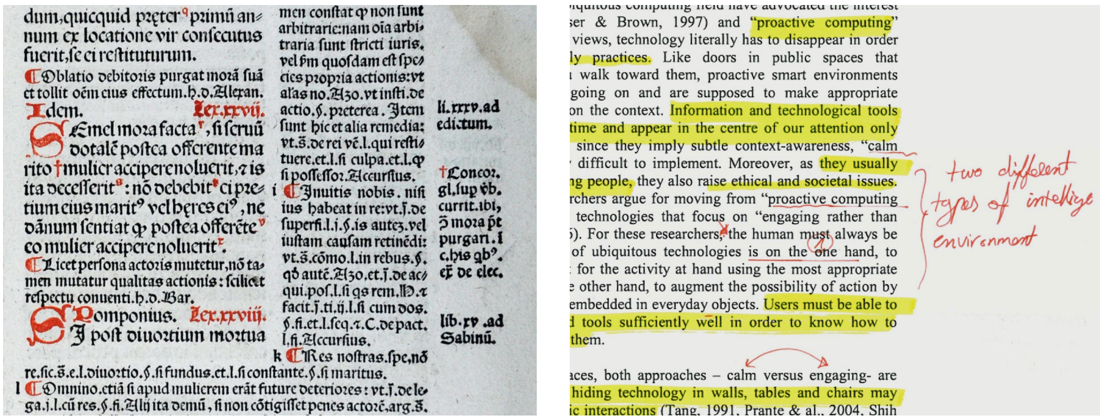
Pierre Dillenbourg
CRAFT - EPFL
Rolex Learning Center
CH-1015 Lausanne
pierre.dillenbourg@epfl.ch

Abstract

This article first describes a method for extracting and classifying handwritten annotations on printed documents using a simple camera integrated in a lamp. 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. In the second part of the article we provide a road map for conducting user-centered studies using eye-tracking systems aiming to investigate the cognitive roles and social effects of annotations. Based on our understanding of some research questions arising from this experiment, in the last part of the article we describe a social learning environment that facilitates knowledge sharing across a class of students or a group of colleagues through shared annotations.

1 Introduction

Annotations played an important role in the history of the book. Already in the early middle ages, the annotations, at that time known as glosses, started to appear on the manuscripts. They were born with a social vocation and on a scholarly need for elucidation and reinterpretation of the obscure passages of the medieval manuscripts. Therefore the glosses became widely considered as precious reading support. For example, one thinks to the adoption of the Justinian Codes in many law schools. The Infortiatum 1(a), the second volume of the Digest of Justinian was reconstructed with additional glosses aiming at re-contextualizing the ancient Roman norms in the current literature. Later, with the invention and spread of the printing press, the book lost its uniqueness becoming cheaper and largely accessible to people. Therefore less official and more individual forms of annotations naturally emerged and became the common practice still widely used nowadays. Textbook annotations have been recently formalized as set of different forms and functions [9] directly involved in the active reading process [1]. 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.



(a) XVI Century (b) July 2010

Figure 1: Printed textbook: a jump through centuries

Figure 1(a) shows a passage of the Infortiatum accompanied by some official glosses (XVI Century) and figure 1(b) shows part of a recent scientific article annotated by a Master's student. Although a comparison between these two elements is out of the focus of this paper, one notices that after centuries and despite the numerous recent digital reincarnations of the traditional printed book, paper remains the preferred medium for reading. In defense of this last statement there is a consistent body of literature comparing reading activities on paper and online documents. Some of the major findings have been summarized by O'Hara et al. [11]. Paper documents offer better legibility and better orientation and location. Physical tangibility facilitates handwriting and concurrent reading on multiple pages. In addition annotating on paper has many well known advantages compared to any digital equivalent [7]. Readers write comments in the margins of documents, underline important passages and use other various marking strategies. These practices help them to better understand what they read and, at a later stage, find back relevant passages. They also play an important role for associative thinking and linking the content with other ideas and documents. One ambition of such a research is to offer a seamless integration of notes, taken on printed paper in our daily interactions with digital documents. Such system inspires one first set of questions around the effects of personal annotations on reading, understanding and learning from texts.

An interesting difference between the medieval glosses and modern annotations is in the value that follows. For example Irnerius, the founder of the School of Glossators, collected and re-organized the original meager interlinear notes of the Infortiatum (figure 1(a)) into an integral part of the document, written in the margins of the page as reading support for everybody. On the other hand the annotations done in our everyday interactions with textbooks probably won't be read by anyone other than us. In addition the glossators acquired considerable authority in the Academic community. Compatibly to their reputation their glosses, acquired a certain social importance. Drawing inspiration from this scenario a second set of research questions focuses on the social effects of the annotations in a real learning context.

The first part of the article reviews a number of systems that have been investigated in the last 20 years to tackle the classification of machine-printed and handwritten text. 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. In the third part of the article we provide a road map for conducting user-centred studies using eye-tracking systems aiming to investigate the cognitive roles and social effects of annotations. Based on the hypothetical findings of this experiment we describe in the last part of the article a social learning environment that facilitates knowledge sharing across a class of students or a group of colleagues through shared annotations.

2 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 the early 90s two works focused on the classification at character level. Kunuke et al. [8] 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 96.8% on a training set of 3632 and 78.5% on a test set of 1068 images. Fan et al. [3] 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. [12] 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. [4] 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. [18] 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. [5] 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 [6] showed that a simple discriminant analysis on the vertical projection profiles performs comparably to many robust approaches.

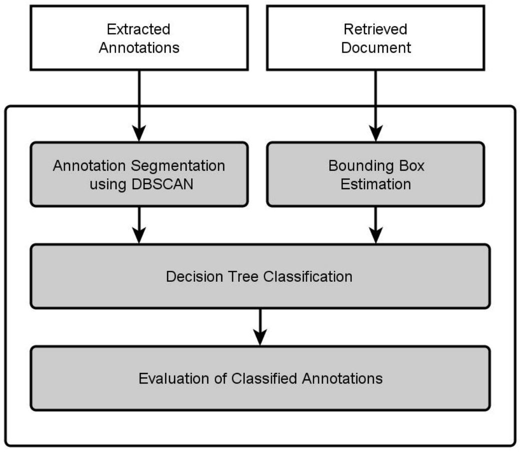


Figure 2: Processing pipeline

One interesting application is the detection and matching of signatures proposed by Zhu et al. [19], a robust multilingual approach, in an unconstrained setting of translation, scale, and rotation invariant non-rigid shape matching.

Peng et al. [13] suggested a novel approach based on three categories of word level feature and a k-means classifier associated with a re-labelling 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. [2] recently described their approach based on the SVM classifier and obtaining an accuracy of 96.9% on a set of 3958 images.

3 Method

We here present our assemblage of techniques for segmenting and classifying handwritten annotations on machine printed documents. Figure 2 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.

3.1 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 type and colour of annotation and any printed document. It can also extract handwritten annotations from handwritten documents. Later improvements by the same authors showed an accuracy rate of 85.59%. These results encouraged us to adopt their method.

3.2 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

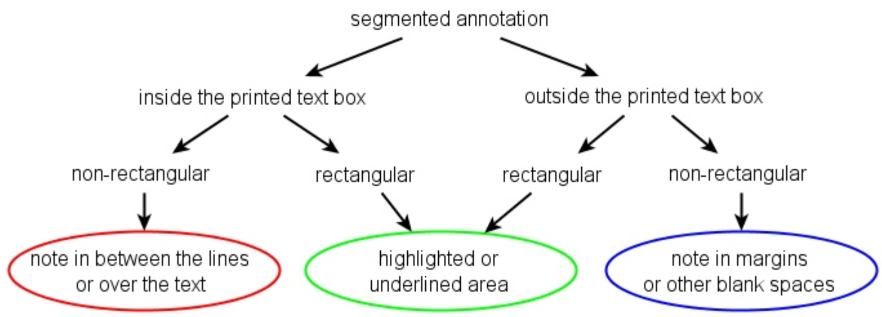


Figure 3: Decision tree classification

- 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 [17]; they removed the original inadequacy in dealing with large-scale data. This allows us not to be bound up with low resolution images.

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

$$p_i = (x_i, y_i, r_i, g_i, b_i) \tag{1}$$

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:

$$A = (A_1, A_2, ..., A_k) \tag{2}$$

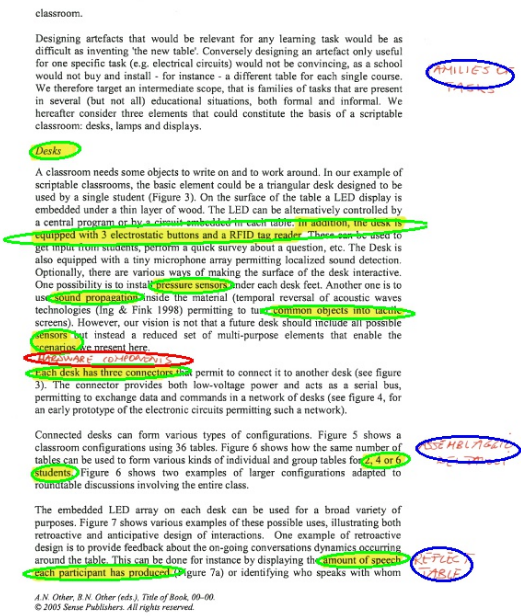
Each cluster corresponds to a correctly segmented annotation. The centroid contains the position of the centre of mass and the mean colour of the annotation. The algorithm is initialized by setting two radiuses, ϵ_{pos} for the spatial domain and ϵ_{rgb} for the colourimetric one and a minimum density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

3.3 Decision Tree Classification

A classification of different forms of annotation is analyzed by Marshall [9]; 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 3 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 a measure of rectangularity; some methods to compute this derived feature are proposed by Rosin [15]; these methods have desirable properties for our scenario such as position, scale, rotation invariance and robustness to noise.

The rectangularity measure that we compute on each annotation is based on the correspondent minimum bounding rectangle (MBR). 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 4 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.

HARDWARE COMPONENTS

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.
© 2005 Sense Publishers. All rights reserved.

Figure 4: Annotation classification result

3.4 Preliminary 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 2) indicating the corresponding function of each colour used for annotating.

The experimental results show a classification accuracy of 84.47%. 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.

4 Eye-gaze patterns to explain the cognitive roles of the annotations

The potentialities of the system presented in the previous section lead us to address some basic questions concerning the effects of personal/shared annotations on reading, understanding and learning from texts. We first pose two main general research questions:

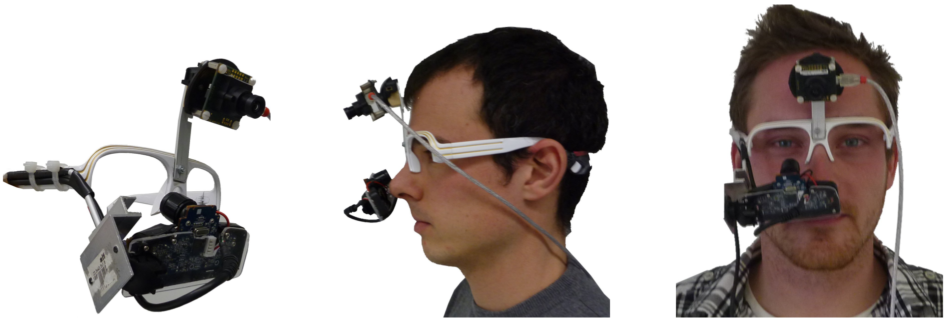
- If A reads a document annotated by B, is her/his reading pattern and learning gain different than the ones performed while reading a document annotated by her/himself?
- If B annotates a document, are the learning gain and the produced annotations affected by being aware that the document will be read by A?

To be able to answer these questions, we here design some explorative and experimental eye tracking studies.

4.1 Eye-gaze data collection

Eye tracking systems have been frequently employed to study eye movements in reading [14]. In this experiment eye movements will be recorded using a head-mounted gaze tracking system (figure 5) specifically designed for analyzing reading tasks on paper material. The system is composed by:

- a colour camera, placed on the head of the reader, responsible for localizing and recognizing the paper document on the table and for processing the handwritten annotations (Section 3)



(a) Glasses (b) Side (c) Front

Figure 5: Wearable eye tracker

- an infrared led close to the reader's right eye, used to control the illumination and simplify the pupil tracking system
- a second camera, placed in front of the reader's right eye, equipped with an infrared pass filter, used to acquire the eye-gaze data [16]

4.2 Participants and instructional material

A corpus of University students will be recruited for our experiments. The instructional material will be created ad hoc. The prior knowledge on the instructional material will be properly determined through a pre-test.

4.3 Experiments

Annotating purpose affects reader's understanding. Producing annotations while reading plays an important role in the learning process. We intend to investigate the effect of annotating for different purposes on the reader's cognitive effort put during the reading task. The two conditions are: annotating for individual benefit and annotating for a social purpose. We intend measure the Relative Learning Gain (RLG) and to explain it through the reading gaze pattern. We also intend to explore whether the annotating purpose leads to a pattern of "visual quality" of the annotations produced or a specific proportion among the annotation types, already described in Section 3.3.

Annotations affect reader's understanding. A document that has been annotated for a social purpose should contain better quality and meaningful annotations. While annotations produced for an individual purpose will supposedly be less readable and understandable from anyone other than the author. We intend to explain the outcome of the measured RLG with the reading gaze pattern. In addition we investigate whether the scarcity of annotations has an impact on the reader's gaze patterns and understanding.

4.4 Dependent Variables

We will measure the Relative Learning Gain (RLG) at the end of each reading session using ad hoc post-tests. The RLG will be calculated as normalized difference between the post-test and pre-test score. It will be computed for each participant and for each portion of text involved in a question.

4.5 Explanatory Variables

Rayner [14] gives a very detailed overview on the eye-gaze reading features. We intend to find out explanatory behaviours of these featers regarding significant trends in the RLG.

5 Social effects of handwritten annotations: research questions

Teamwork is a skill that is often taught and encouraged in universities. Students typically work together in small informal, sometimes formal, groups in order to solve exercises, to discuss the course material and to prepare exams. Team working involves active participation, interaction and frequent sharing of ideas and annotated instructional material. Based on our understanding of the basic questions addressed in the previous Section, we will develop a social learning environment that facilitates knowledge sharing across a class of students or a group of colleagues through shared annotations.

By experimenting this environment, we will complement our understanding of

- the cognitive effects of annotations (*question 1*)

and acquire knowledge about

- their social effects in a real context (*question 2*)

To address this second question, the experimental results of our study will be embedded into a collaborative reading environment that captures, classifies and shares annotations. A simple example of application would be to show to students a hierarchical view on the most annotated pages of their lecture notes, right before the exam. Another application is to provide an automatic production of abstracts based on the annotations produced by a class of students. These tools will be developed for and tested in an authentic context. We believe that restoring the original social vocation of the annotating process through this micropublishing platform will encourage contributory and participatory behaviours among the students.

6 Conclusion

In the first part of this paper we describe a system for clustering and classifying handwritten annotations, extracted using already existing techniques, achieving the accuracy rate of 84.47%. In the second part of the article we provide a road map for conducting user-centred studies using eye-tracking systems aiming to investigate the cognitive roles and social effects of annotations. Based on the hypothetical findings of this experiment we describe in the last part of the article a social learning environment that facilitates knowledge sharing across a class of students or a group of colleagues through shared annotations.

7 Acknowledgments

I take advantage of this opportunity to thank my colleagues Quentin Bonnard and Youri Marko for their help with the implementation of the page recognition system and the wearable eye tracker and the other colleagues at CRAFT for their valuable advices and assistance.

References

[1] M. J. Adler and C. Van Doren. *How to read a book*. Touchstone Books, 1972.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[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] K. O'Hara and A. Sellen. A comparison of reading paper and on-line documents. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 335–342, New York, NY, USA, 1997. ACM.

[12] 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.

[13] 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.

[14] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372–422, November 1998.

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

[16] J. San Agustin, H. Skovsgaard, J. P. Hansen, and D. W. Hansen. Low-cost gaze interaction: ready to deliver the promises. In *CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4453–4458, New York, NY, USA, 2009. ACM.

[17] 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.

[18] 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.

[19] 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.