A proposal of a document image reading-life log based on document image retrieval and eyetracking

Olivier Augereau, Koichi Kise, Kensuke Hoshika
Osaka Prefecture University, Japan
Email: augereau@m.cs.osakafu-u.ac.jp

***Abstract*—Instead of analyzing directly the document images, analyzing the document reading can offer new perspectives for extracting information about both the reader and the document. Analyzing how people read texts can help to understand the cognitive process of the reading and might lead to new approaches and new solutions for pattern recognition and document image analysis. It can also lead to create smart documents that can measure reading information, provide feedback and adapt themselves depending on the behavior of the readers. As a step towards document reading analysis, the authors propose in this paper a solution for extracting the reading information and creating a "reading-life log". This reading-life log contains basic features that can be used for many different kinds of applications. A tag cloud evolving according to the reading is presented as a first application of the reading-life log.**

I. Introduction

Reading documents is one of the main sources of knowledge. We are daily spending a lot of time for reading paper or digital documents such as newspapers, novels, emails, websites, *etc*. Most of us do not live a single day without reading. This indicates the ubiquity and the importance of reading. Despite this importance, no general public system is yet available for recording and make use of our reading. This paper is concerned with the analysis of the reading activities. Analyzing the reading activity is important to know about the readers. In addition, we can know more about the documents from the viewpoint of reading. How a document has been read by users. In other words, mutual analysis of documents and readers is considered.

We believe that the mutual analysis can open a new era of the field of document analysis. In the field, we have mainly focused on the analysis of documents themselves. We have not considered the readers who are the recipients of the information. By considering the reading behavior, we can analyze documents from the readers' viewpoint. For example, the reading analysis can show which parts of a document are difficult to understand or interest readers. This information about the document cannot be found by a traditional document image processing.

Some previous work has been done about reading analysis, especially by psychological and cognitive researchers [13]. But almost all of them focus on studying and measuring the reading behaviors and thus experiments were done under carefully controlled settings. No work has been done to utilize the result of analysis of reading behavior for any concrete applications as we suggest in this paper. The eye movement is studied but it is not associated with words. We have already started to analyze the reading activities for estimating the number of read words [8] or to classify documents [9]. In addition, we have already done some aspects of estimating the level of understanding of documents [7]. All of them employ the reader's eye gaze data for analysis of the reading activity. The whole project is called "reading-life log".

Nevertheless, we haven't been successful to log all the reading behavior yet. This is because of the inaccuracy of the eye gaze estimation. It is not possible to estimate the word a reader is looking at by using the-state-of-the-art devices of eye tracking and especially if the environment is not fully controlled. The main contribution of this paper is to realize the reading-life log by taking into account this inaccuracy. A simple error estimator is employed to log the read words probabilistically. In addition, we discuss possible applications of reading-life log in addition to the examples mentioned above (wordometer, reading detection, document type estimation). This includes visualization of the read words by using a tag cloud.

The rest of the paper is organized as follows. In the next section some related work about reading analysis and its applications are presented. In section III, our method is developed in two steps: 1) measuring the error distribution of the eye gaze and 2) using it for approximating the read words and creating the reading-life log. Then, in the section IV, we will put the focus on one application that can be made by using the reading-life log: building a tag cloud evolving with the reading time. It will be compare to a standard tag cloud based on the whole document text. The section V will conclude the paper.

II. Related Work

There are many algorithms for analyzing, processing or synthesizing document images, but very few focus on how documents are read. Document reading analysis is an alternative to document image analysis for extracting information about the documents, but also about the reader. Recently, some work about reading activity has been proposed in order to recognize document types (textbook, novel, manga comic and a newspaper...) [9]. Some educational applications for helping to learn a foreign language [10], to infer English language skill and spotting difficult words in the reading activity [7] or to estimate the number of read words during daily life [8] have also been presented. These methods were mainly based on the analysis of the eye gaze *i.e.* the position where our eyes are looking at. We would like to go a step forward by adding to the eye gaze, the information of the read words and creating a reading-life log.

More related work can be found in psychological, cognitive and neurological papers. Since the early 1970's to nowadays, Keith Rayner have conducted researches about eye movement analysis [13]. The basic characteristics are measuring from fixations and saccades [13]. The fixations correspond to the place where our gaze pause and focus, whereas the saccades are the transitions between the fixations. For example, the fixation duration, the length of saccades, *etc.* are used to distinguish the reading behavior of a child and an adult, to estimate the difficulty of a text or the language comprehension of a reader [14].

Some application have been done also about dyslexia. Schuett [17] analyzed reading performance and eye-movements in hemianopic dyslexia and then proposed some solutions for the rehabilitation by re-learning eye-movement control in reading. Music reading have also been studied, and especially the sight-reading - reading a score while playing an instrument [5]. Penttinen and Huovinen [11] analyzed the effects of skill development on the eye movements. This lead to understanding better the problems encountered by novice sight-readers and to advancements in the pedagogy of music reading education. The eye-tracking technology had also been used in in the WWW community for optimizing web pages advertisement [3]

Understanding eye movement and how we read text can certainly lead to major educational applications. Creating a reading-life log will allow researchers to use not only the eye movement but also the words that are read as a basic feature for analyzing the reading activity and proposing new applications.

III. Proposed system

One of the main problems that arises for analyzing the reading is that the natural reading behavior does not consist in stopping and staring to each individually word one by one. Small words are perceived but often skimmed. The reading activity is not fluid; it is decomposed into fixations and saccades. As described in Fig. 1, our eye get the major information from the fovea but also from the characters surrounding located in the parafovea. Our brain is able to anticipate and skip small or predictable words [17]. On average, around 7 to 9 letters separate two saccades and the perceptual span is from 4 letters to the left of fixation to 15 letters to the right of fixation [16]. Depending on the reading skill, one word can have sometimes more than one fixation - if the word is rare or the reader skill is beginner, and one word can have no fixation - if the word is short or the reading skill is high.

As illustrated in Fig. 2, the proposed system is composed of a camera, recording what the reader see and an eye tracker that compute the eye gaze coordinates. We assumed that the read document is contained in a database in a digital version. The content and position of all words are known in advance thanks to the digital version. The video images are used for two different things: 1) locating the eye gaze and 2) applying document image retrieval. If the document is retrieved, the homography between the document image of the video and the corresponding document image in the database is computed.

The eye gaze detected in the video image is then projected in the digital image and the read words can be estimated. We propose to estimate which word is read from the eye gaze by using a 2D Gaussian model.

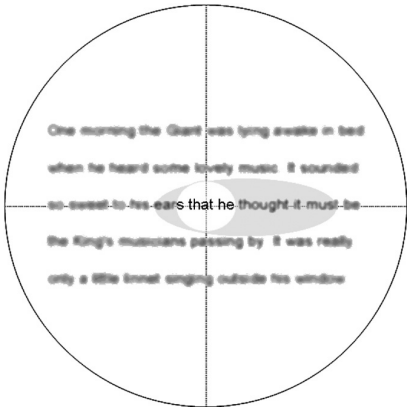

Fig. 1. The white circle represents fovea perception, this is where eye is focusing. The gray ellipse represents parafovea perception. The characters in the gray area are perceptible with less acuity by the eye, but help to understand the read word. If a small or a predictable word is located in the parafovea, it might be skimmed. This figure is extracted from [17].

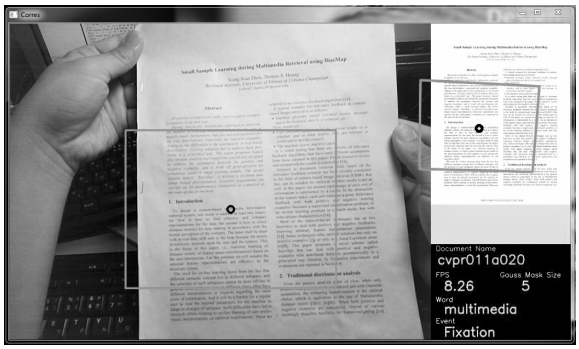

Fig. 2. The image in the left part come from the video recorded by SMi eye tracker. The circle correspond to the eye gaze fixation. The rectangle is the subsection of the whole image selected for applying document image retrieval. The right part represents the digital version of the document image. LLAH is used for retrieving the document image in the database and computing the transformation between the video image and the database image. The position of the eye gaze into the digital document image can be estimated thanks to this transformation.

A. Eye tracking

The eye tracker used is the SMI iViewX ETG (Eye Tracking Glasses). This eye tracker has a sampling rate of 30 Hz and the scene video is recorded with a resolution of 1280x960 pixels. Before using the eye tracker, a calibration step is needed. The raw output data is processed in order to filter microsaccades. The eye gaze fixations and saccades are obtained by using the Busher and Dengel method [2]. The coordinates of the fixations are obtained from the video image as shown in the left part of Fig. 2.

B. Document image retrieval

Locally Likely Arrangement Hashing - LLAH, is a document image retrieval technique [18]. It is based on interest point extraction and a hash table. The retrieval is robust against perspective distortion. Several million pages can be stored in a database and a document image can be retrieved in real time. If

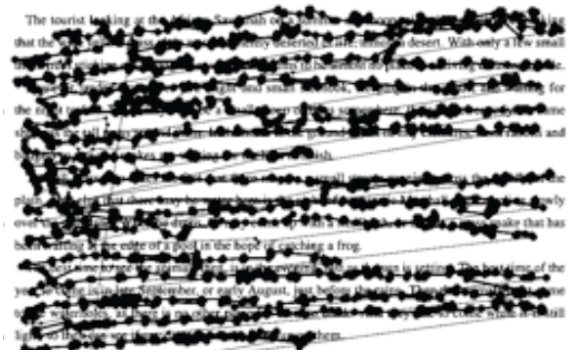

Fig. 3. Eye gazes displaying on the digital document image after applying the homography transformation. The dots correspond to the fixations and the dashes correspond to the saccades. As we can see, the position of the eye gaze and the words are not exactly aligned because of the lack of accuracy of the eye tracker.

the document is retrieved, the homography between the video image and the document image contained in the database can be computed. We use this homography in order to compute the coordinates of the eye gaze in the document image. Figure 3 shows the eye gaze projected in the digital document image.

C. Measuring the eye tracking precision

We would like to compute the difference between the eye gaze position recorded by the eye tracker and the real position we are looking at. As we can see in Fig. 3, the eye gaze fixation coordinates correspond roughly to the position of the read words. The idea is to create a model that takes into account this difference in order to approximate which words might have been read for a given eye gaze fixation. Evaluating the error is difficult because no ground truth is available. In order to know which word should be associate with which eye gaze, one text is read out loud. Then we manually associate word position - the center of the word bounding box, with the eye gaze.

Based on the assumption that the distribution is symmetric along the x and y axis, the 2D Gaussian model of errors distribution can be computed and represented as follow:

$$f(x,y) = \frac{1}{2\sigma_x\sigma_y\pi}\exp(-\frac{1}{2})\{\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2}\},$$

where μ_x and μ_y are the expected values of the differences along x and y, and σ_x and σ_y are the standard deviation along x and y.

The obtained model is dependent to the user and to the eye tracker technology. If the eye tracker is very accurate, the shape of the Gaussian model will be narrower; and if the eye tracker is less accurate, the Gaussian model will be wider.

D. Reading-life log creation

When we read, long words can have several eye gaze and short words can have no eye gaze. It is not possible to associate exactly one word to one gaze. For every eye gaze, the 2D Gaussian model is centered to the eye gaze and the probability is computed for the surrounding words. So, the equation became:

$$f(x,y) = \frac{1}{2\sigma_x\sigma_y\pi}\exp(-\frac{1}{2})\{\frac{(W_x-G_x)^2}{2\sigma_x^2} + \frac{(W_y-G_y)^2}{2\sigma_y^2}\},$$

where G_x and G_y are the coordinate of the eye gaze; and W_x W_y the coordinate of the word. In order to reduce the processing time, the Gaussian function is thresholded by a rectangle centered to the Gaussian of $2\sigma_x$ width and $2\sigma_y$ height. If the word is farther, the weight is approximated to 0. A word closer to the eye gaze will have a greater value than distant ones.

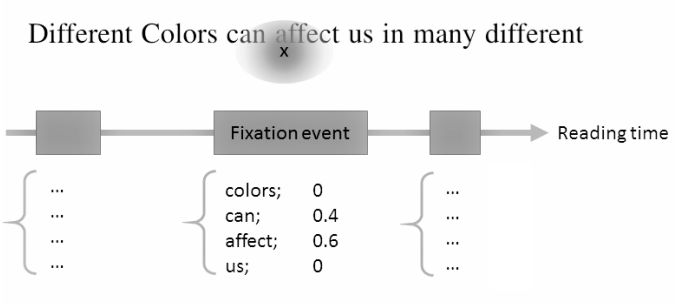

Fig. 4. In the upper part of the figure is represented the read text, the cross is the eye gaze and the ellipse represents the Gaussian model. This model is then used for estimating the weight of each surrounding words. The lower part represents the reading-life log. For each eye gaze fixation, the estimated read words are save with the corresponding weight.

As shown in Fig. 4, the reading-life log contains a list of estimated read words evolving with time. For each fixation event - represented as rectangles in the figure, the surrounding words are weighted with the estimated probability. We also save the fixation duration of the words which is directly obtained from the eye tracker.

This reading-life log is the corner stone for building the following reading analysis applications.

IV. Application

Our application consists in reading a paper document with a digital version available in a database. An image retrieval technique is used for retrieving the scene document image and to provide access to the document layout.

A. Tag cloud summarization

The tag cloud has been chosen as an application because we can easily show the difference between a standard document image text analysis providing a classical tag cloud and the tag cloud obtained based on the reading-life log. Even if all the document is read, the reading tag cloud might be different from the document tag cloud. For example, some parts of the document have been read several time or some parts have been skimmed. According to Rayner *et al.* [15] about 10 % to 15 % of time, the saccades go backward to the text, and a skilled readers fixate only about two third of the words in a text. The advantage of using the reading-life log is that it reflects the content of what we read instead of reflecting the content of the document and provide a different information than standard document image analysis.

TABLE I. ESTIMATED GAUSSIAN PARAMETERS. σ_x, σ_y AND μ_x AND μ_y ARE RECIPROCALLY THE STANDARD DEVIATION AND THE AVERAGE DISTANCE IN PIXELS BETWEEN THE GAZE AND THE READ WORDS ALONG X AND Y AXIS.

Estimated parameters	reader 1	reader 2
σ_x	11.59	33.44
σ_y	16.32	40.16
μ_x	21.79	45.37
μ_y	32.61	40.43

As an experiment, we propose to compare the tag cloud generated for two different readers and the standard tag cloud based on the whole document.

The experiment was conducted in 3 steps:

1) Eye tracking calibration. The calibration is done by using the software SMi BeGaze [1]. It consists of 3 dots alignment.
2) Model error estimation. One first text is read out loud in order to create the error model by associating eye gaze with words. One Gaussian model per user is created.
3) Reading-life log creation. Another text is then read in normal condition - not aloud. In this experiment the document is composed of four paragraphs. For each eye gaze, we estimate which word have been read.

The table I summarize the estimated parameters of the Gaussian model for the two different readers. The format of the digital document is A4 with a resolution of 200 dpi, being 1654 x 2339 pixels. The result from the table shows that the error of the estimation is roughly the size of few letters. In order to compare, the size of "e" is around 14 x 17 pixels. The estimation of the read text for the reader 2 will be less accurate as his standard deviation and average value are higher.

These two models were then used in order to estimate the read words while reading in normal condition another text. In order to compare the result obtained with the reading-life log and the document, the tag cloud based on reading information and the standard tag cloud based on all the text of the document are displayed in the Fig. 5. In the tag cloud based on the reading, the size of the font depends both on the frequency and the weight associated to the read words in the reading-life log. In the tag cloud based on the document text, the size of the font depends only on the occurrences of each word. Unlike to document analysis, reading analysis contain information of time and progress.

As we can see, the content of the tag words is very different for the two users. The tag clouds estimated after reading the whole document tend to be more similar each other because words are cumulated trough paragraphs. If every paragraph is read exactly one time, the tag cloud based on the full read text will tend to be similar to the tag cloud based on the text of the document. But, if some paragraphs are skipped or reread, the reading-life log and the tag cloud will be quite different. For example, if one reader stop after reading only the first paragraph, the tag cloud will be quite different than if he read all the document. Comparing to standard document analysis, using the reading-life log give an information about which part have been read or not.

[1] http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/begaze-analysis-software.html

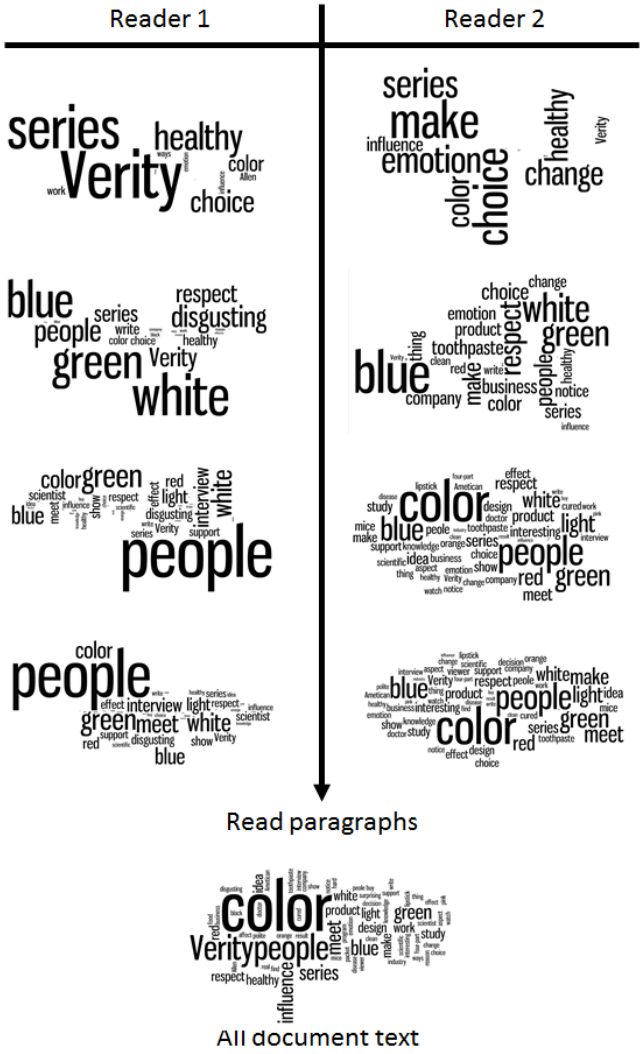

Fig. 5. Tag words based on the reading-life log. Two readers have read the same text composed of four paragraphs. From top to bottom, the tag words cumulate read words from 1st paragraph, 1st and 2nd paragraph, 1st, 2nd and 3rd paragraph and finally, the whole document. In the lower part of the figure the tag cloud computed from the text of the whole document - four paragraph, is displayed.

B. Others applications

We would like to show that the reading-life log we presented should be seen as a standard data structure that can be easily used for others applications.

1) Researching read information: The reading-life log contains information about the words we read and at what time we read them. A direct application is to use the reading-life log to search about something we read at a certain time.

2) Measuring user interest: By accumulating the reading-life log from many users for each paragraph, it is possible to see which paragraph is more read and where reader lose interest.

3) Understanding: Some cognitive researchers had shown that the level of understanding of a reader is directly link to the fixation time [4]. The reading-life log also contains the fixation duration of each word, so it is possible to estimate the level of understanding.

The reading-life log can improve our reading life by giving a feedback about what we read and how we read text. It will lead to have more knowledge about ourself and about document. Many studies have been done in order to predict the comprehension of the reader based only on eye movement [14], [1], [6]. But as said Perfetti *et al.* "for comprehension to succeed, readers must import knowledge from outside the text" [12]. Not only the eye movement but also the read text is important. By using a reading-life log in the daily life, it can be possible to save many information about all the texts we read and estimate the knowledge of the reader in order to provide him assistance.

V. Conclusion and perspectives

We presented in this paper a new system for logging the words we read. The creation of a reading-life log is quite challenging for two main reasons. Firstly, the current eye tracking technology is not accurate enough for determining exactly which word is read at which time. Depending on the used device, the error of the estimation will be more or less large. And secondly, the reading behavior is complex: some words are fixate but some other words are skimmed.

In order to avoid the problem of text detection in scene image, we apply document image retrieval and use the digitized version of the document in order to access the text. A 2D Gaussian is used in order to model the gap between the eye gazes and the read words. The model is dependent to the user and the eye tracking technology. A reading-life log is then created based on this model. It represents basic information about the read text and can be used for different purposes. It is a first step toward new approaches for document and reading analysis.

The reading analysis can help us to obtain new information from the reader and the document that cannot be obtained by standard document image analysis. We showed an application for creating a tag cloud as it is a useful tool for summing up the read text, but we also see that it can open new possibilities such as researching something we read, or measuring the interest and understanding of a document.

As future work we consider some applications for educational purposes. The first one will be a feedback for professors to see which part of the lesson the students reread the most and need more explanation or to provide individual help. The second one will be for learning a language: estimate the vocabulary of the reader and to sum up new frequent words he read in a day in order to help him to remember them.

Another future work will be done in order to improve the reading-life log itself, especially for decreasing the error estimation of the read words by matching the eye gazes to the document layout. We also would like to combine our error estimation model with the cognitive model presented in Fig. 1, *i.e.* to consider the parafovea perception.

Acknowledgment

This work was supported in part by the JST CREST and the JSPS Kakenhi Grant Number 25240028.

References

[1] Nicola Ariasi and Lucia Mason. From covert processes to overt outcomes of refutation text reading: The interplay of science text structure and working memory capacity through eye fixations. *International Journal of Science and Mathematics Education*, 12(3):493–523, 2014.

[2] Georg Buscher and Andreas Dengel. Gaze-based filtering of relevant document segments. In *International World Wide Web Conference (WWW)*, pages 20–24, 2009.

[3] C.S. Campbell. Method and system for the recognition of reading skimming and scanning from eye-gaze patterns, March 29 2005. US Patent 6,873,314.

[4] Charles Clifton, Adrian Staub, and Keith Rayner. Eye movements in reading words and sentences. *Eye movements: A window on mind and brain*, pages 341–372, 2007.

[5] Helga Rut Gudmundsdottir. Advances in music-reading research. *Music Education Research*, 12(4):331–338, 2010.

[6] Haijun Kang. Understanding online reading through the eyes of first and second language readers: An exploratory study. *Computers & Education*, 73:1–8, 2014.

[7] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. Towards inferring language expertise using eye tracking. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 217–222. ACM, 2013.

[8] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. The wordometer–estimating the number of words read using document image retrieval and mobile eye tracking. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 25–29. IEEE, 2013.

[9] Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. I know what you are reading: recognition of document types using mobile eye tracking. In *Proceedings of the 17th annual international symposium on International symposium on wearable computers*, pages 113–116. ACM, 2013.

[10] Ayano Okoso, Kai Kunze, and Koichi Kise. Implicit gaze based annotations to support second language learning. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 143–146. ACM, 2014.

[11] Marjaana Penttinen and Erkki Huovinen. The early development of sight-reading skills in adulthood a study of eye movements. *Journal of Research in Music Education*, 59(2):196–220, 2011.

[12] Charles A Perfetti, Julie Van Dyke, and Lesley Hart. The psycholinguistics of basic literacy. *Annual Review of Applied Linguistics*, 21:127–149, 2001.

[13] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[14] Keith Rayner, Kathryn H Chace, Timothy J Slattery, and Jane Ashby. Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10(3):241–255, 2006.

[15] Keith Rayner, Barbara R Foorman, Charles A Perfetti, David Pesetsky, and Mark S Seidenberg. How psychological science informs the teaching of reading. *Psychological science in the public interest*, 2(2):31–74, 2001.

[16] Keith Rayner, Timothy J Slattery, and Nathalie N Bélanger. Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, 17(6):834–839, 2010.

[17] Susanne Schuett. The rehabilitation of hemianopic dyslexia. *Nature Reviews Neurology*, 5(8):427–437, 2009.

[18] Kazutaka Takeda, Koichi Kise, and Masakazu Iwamura. Real-time document image retrieval for a 10 million pages database with a memory efficient and stability improved llah. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 1054–1058. IEEE, 2011.