Eye-tracking Analysis
for Automatic Documents Eye-catching Layout Retrieval

Véronique EGLIN
LIRIS, INSA Lyon
Veronique.eglin@liris.cnrs.fr

Jean CAELEN
CLIPS, IMAG Grenoble
Jean.caelen@imag.fr

Abstract

In this paper we present a synthesis of experiments of eye movement pursuit that have been applied to documents structure retrieval. The aim of this work is to propose a representation of structured documents content (the physical layout) through the simulation of a possible human inspired scan path. The research project which is presented here is based on the hypotheses that the analysis and the comprehension of real human trajectories are necessary to design a realistic automatic self-governing system that simulates eye-catching information retrieval whatever are the page designs.

Keywords: Eye-tracking, scan path simulation, documents layout retrieval.

1 Cognitive approach of information retrieval

The recordings of ocular movements which are measured on observers during the documents exploration give two kinds of information. Firstly, they give evidence on documents legibility: for example, they make easier the comparison between two documents which propose the same content. On the other hand, scan-path recordings underline different visual exploration strategies for a same investigation task, [10]. We have exploited two different human behaviors (inspection and skimming) and we have focused on typical strategies of structured documents reading so as to propose an image processing automated system dedicated to documents layout retrieval. The system has been designed so as to simulate human visual behavior for a global page scan. This work has been initially supported by the SHIVA[1] project and a PhD in the LIRIS, [6].

This work lies on the hypotheses that the scan-paths attest cognitive processes which are implied in the observation task, [7,8,10]. In the reading task for example, the analysis of fixations duration, their amplitudes, their locations and their variations gives evidence of mental brain operations which are automatic or under the observer control. The scan-path recording is a directly measurable sign of the observer's care and interest. In this paper, we present the analysis of different scan-paths applied to the evaluation of Web pages design. The conclusions of the study have been supported by well known human psychovisual behaviors ([10]) to design the architecture of our system of automated layout extraction. In this paper, we do not detail any results and we report the reader to complete them with different references and bibliographies which are cited through the paper.

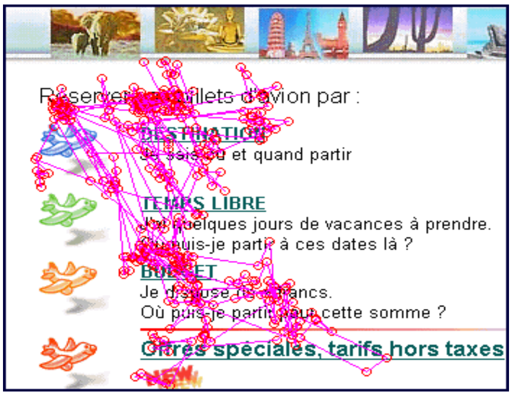
2 Retrieval and evaluation of documents visual content

2.1 Different exploratory situations

We will not present here the eye-movements recorded device (the eye-tracker), which has been used in our experiments. We only report the reader to complementary information concerning points measures exploitation and guidance software which has have been developed by the CLIPS laboratory, see http://www-clips.imag.fr. The eye-tracker device allows recording the position of each eye of a subject during a scan on an image. With the information relative to the succession of gazing points, it's possible to study the gaze

[1] SHIVA: Site Hypermédia et Inspection Visuelle Automatique (projet Emergence Rhône-Alpes)

movements of an observer during his exploration. Figure 1.1 represents a real time gazes recording with 250 ms time elapsed fixation points. The figure 1.2 illustrates a simplified recorded scan-path obtained on a scientific paper: a fixation area is represented in the scan by a direction change and corresponds to the average of a series of quasi stationary eyes positions during a time laps always inferior than 2s. The scan path is characterized by three aspects: a *temporal* component with fixation time in different parts of the documents; a *spatial* component with interest regions areas, fixation points localization ; a *strategic* component with fixations number and backtracking.



Réserver vos billets d'avion par :

DESTINATION
Je sais où et quand partir

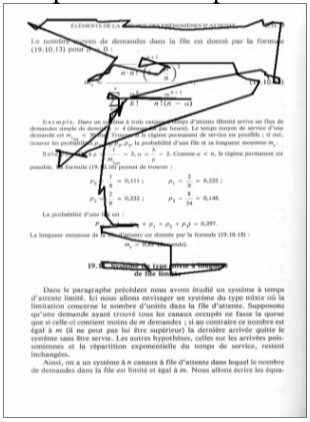
TEMPS LIBRE
J'ai quelques jours de vacances à prendre.
Où puis-je partir à ces dates là ?

BUDGET
Je dispose de x francs.
Où puis-je partir pour cette somme ?

Offres spéciales, tarifs hors taxes

NEW

Figure 1.1: Complete scan-path with 250ms time elapsed fixation points.



ÉLÉMENTS DE LA [illegible] DES PHÉNOMÈNES D'ATTENTE

Le nombre moyen de demandes dans la file est donné par la formule (19.10.13) pour [illegible] 0 :

Exemple. Dans un système à trois canaux [illegible] temps d'attente illimité arrive un flux de demandes simple de densité λ = 4 (demandes par heure). Le temps moyen de service d'une demande est [illegible] Trouver si le régime permanent de service est possible ; si oui, trouver les probabilités [illegible] la probabilité d'une file et sa longueur moyenne [illegible].

possible. La formule (19.[illegible].16) permet de trouver :

La probabilité d'une file est :

La longueur moyenne de la file d'attente est donnée par la formule (19.10.18) :

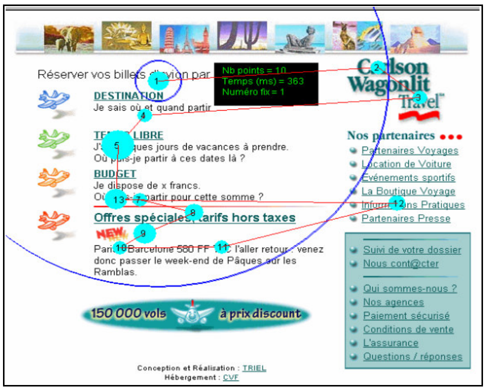
19.[illegible] Système du type mixte à longueur de file limitée

Dans le paragraphe précédent nous avons étudié un système à temps d'attente limité. Ici nous allons envisager un système du type mixte où la limitation concerne le nombre d'unités dans la file d'attente. Supposons qu'une demande ayant trouvé tous les canaux occupés ne fasse la queue que si celle-ci contient moins de *m* demandes ; si au contraire ce nombre est égal à *m* (il ne peut pas lui être supérieur) la dernière arrivée quitte le système sans être servie. Les autres hypothèses, celles sur les arrivées poissoniennes et la répartition exponentielle du temps de service, restent inchangées.

Ainsi, on a un système à *n* canaux à file d'attente dans lequel le nombre de demandes dans la file est limité et égal à *m*. Nous allons écrire les équa-

Figure 1.2: Simplified scan-path with average fixation regions.

Those three main characteristics depend on the kind of reading document: during a reading process, for example, the scan-path gives information on the way people construct the content signification, [9, 10]. It has been shown that a long fixation time[2] on the same text area or regular backtrackings actions reveal a difficulty for the observer to understand the content, [6]. The initial fixation seems to be sufficient to bring out global information on the sense of the scene. In [10], the author has proved that it must be considered as a determining starting point to determine interest areas in a document. On the other hand, we can notice two complementary processes: a peripheral one which allows colors and coarse shapes perception and a central one which is connected to cognitive processes with memory access, semantic comprehension and interpretation, see figure 3.



Réserver vos billets d'avion par

Nb points = 10
Temps (ms) = 363
Numéro fix = 1

DESTINATION
Je sais où et quand partir

TEMPS LIBRE
J'ai quelques jours de vacances à prendre.
Où puis-je partir à ces dates là ?

BUDGET
Je dispose de x francs.
Où puis-je partir pour cette somme ?

Offres spéciales, tarifs hors taxes

NEW

Paris Barcelone 580 FF TTC l'aller retour : venez donc passer le week-end de Pâques sur les Ramblas.

150 000 vols à prix discount

Carlson Wagonlit Travel

Nos partenaires ●●●

- Partenaires Voyages
- Location de Voiture
- Événements sportifs
- La Boutique Voyage
- Informations Pratiques
- Partenaires Presse

- Suivi de votre dossier
- Nous cont@cter

- Qui sommes-nous ?
- Nos agences
- Paiement sécurisé
- Conditions de vente
- L'assurance
- Questions / réponses

Conception et Réalisation : TRIEL
Hébergement : CVF

Figure 2: Peripheral perception for the next fixation point location.

This cognitive activity is fundamental for human beings for their content understanding. It has been shown in [11] that it can be introduced at a time when a fixation of at least *250* ms is reached. In [9], that author has shown that this activity is strongly influenced by the page content layout. In that context, the typographical and typo-dispositional effects in the page are considered as essential for the reader's understanding. In that sense, this study states some responses to basic questions of information retrieval: "How does the reader proceed to extract the substantive page information. What kind of visual strategies are implied during the information retrieval process? Are there any rules for an ideal page layout configuration? Within all possible responses of those questions, is it possible to automatically retrieve the main page organization?" Some responses to this list of

[2] A fixation lasts on average between 100 et 500 ms. The average duration is 250 ms: it corresponds to the activation of cognitive processes, [12]

questions can be obtained by the exploitation of different visual exploratory strategies: the complete reading, (letters, novels…), the skimming (partial reading of magazines or newspapers) and the fine inspection for an intentional information retrieval (catalogue or dictionary reading), [3]. Here, we will focus our interest on web documents inspection (and the evaluation of their layout in time constraint delay) to produce an automatic system for structure capturing applied on complex pages images. Experimental protocols for Web pages diagnostic are roughly presented here (participants' number, recruitment modes, age and individual experience…). In opposite, we focus here on the conception of the automatic reading system more than on the measurements exploitation.

2.2 Web pages design evaluation using for automatic scan-path simulation

The layout of a document must be as legible as possible to be useful for the reader with a minimum of efforts. An interesting application domain to quantify this legibility and the importance of the structure is the evaluation of web documents. This evaluation consists in measuring the observer's scan path and the delay he uses to retrieve all requested information. This evaluation is considered as a fundamental assistance tools for web pages designers and for the conceiving of an automated scanning system.

The evaluation must be understood here as a visual diagnostic in regard with the customer objectives: if he's looking for precise information, he will necessary be guided by precise regions of the page (moving icons, eye-catching colors, regularly updated data…). So, the evaluation consists in verifying that some criteria (in the page layout and in the overall design) are well respected. By taking the problem in the opposite side, we can assume that a printed heterogeneous document that verifies some basic ergonomic rules (presence and predictable location of titles, existence of legends beside the figures…) can be automatically analyzed and retro-converted by an automatic document exploration.

The research project which is presented here is based on the hypotheses that the analysis and the comprehension of real human trajectories are necessary to design a relevant document structure retrieval system. In that context, we will show that the scan-path of an observer is a powerful indicator of the page hierarchical organization in both physical and logical levels.

3 Synthesis of the ergonomic diagnostic of web pages, [6], [9]

3.1 Reading schemes in context

We can remember that there exist reading schemes which can be adapted to the document content exploration. Existing works have shown that there exist four main scan path strategies, which are implied in all kind of documents exploration (on paper document but also on web pages), [7,10,11]. Here we have focused our interest on commercial web documents. If we combine the four different possible itemized layouts to the three different scan path strategies, we obtain 16 different categories of page layout. The protocol of this study has been presented in [2]. We only give the main points here. The test pages have been chosen for their great structure diversity (different layouts and content design). The experiment relates to a group of 24 voluntary Net surfers and 5 ergonomics experts. The parameters determining the goal of the exploration have been limited in number, [2]. The subjects examine all the sites, in a random order. For each exploration, we have considered visual measures as independent variables: fixations number and duration, distribution of the durations of fixings, the sum of times of glance on the various page components. Scan-paths are studied as sequences of points with duration superior to 250ms[3].

3.2 Synthesized observations

Two great families of behaviors emerge from the recorded scan paths. The firsts correspond to a selective and fast inspection which guides the reader to the precise location and identification of the target. The reader only picks up some partial information of the page content. The

[3] On the methodological level, the one minimal duration choice has been fixed to 250ms so as to ensure that the subject has time to partially process the data captured by its glance and then to direct the following fixations in the desired target. Some studies have shown that 100ms is the best threshold to assure a satisfying image recognition, [8].

second behaviors correspond to an exhaustive page inspection: the time spent to explore the document in depth is generally very long. With this kind of readers, we have an objective and complete description of the page layout with a semantic exam of relevant textual regions (in titles, legends, notes…). The analysis of the first 10 fixations shows that the reader attention is generally guided by the parafoveal region of current fixation. That reinforces the hypotheses that the perception of low resolution peripheral information is fundamental in a process of target detection, [4, 8, 11].

On the other hand, this study has shown that the analysis of the number of fixations which is necessary to the pages content provides significant differences between observers expertise. Then, we have observed that the fixations number is sensitive to the kind of perceived information, [2]. The *overall time* which is required for the page exploration seems to be a good revealing of the information density and the informational complexity of the page. Finally, we have also used a third conclusion which shows the beginning of the exploration is decisive for the page evaluation, [2].

- The initial perception of informative area is supported by the presence of a limited number of elements and by the presence of well identified objects (like hyper links).
- Within the space which is occupied on the screen (by the web page), the spatial distribution or the information density plays an important part.
- Finally, a last factor must also be considered: the task which is assigned for a dedicated exploration. Some more cognitive resultants have been developed

This study intended to show that scan-path is a realistic manner to estimate interest areas for the observer. It has especially show that the 100 ms fixations are visually pregnant for the target retrieval. It is the *perceptual* threshold for interest regions capturing. The 250 ms fixations seem to be the *cognitive* minimal threshold to know which objects or which information are identified as being interesting for the task and for the net surfer. The exploration duration seems to reveal very efficiently the density of information or the layout complexity of the page. For a given task, the *best* page will be the one with the shorter exploration duration. Total fixations number for an exploration with goal is sensitive to present information. It is necessary to aim at the headings semantic precision to reduce the total fixations number. Finally, the fixations distribution correlated to the fixation rank points out the evolution of the observer's interest during his exploration.

The results which are summarized here have been compared to general results which have already been demonstrated in the context of visual inspection, [8], [11]. Here, they have been used to design our simulation system which reproduces basic psycho-visual behaviors of observers in a situation of skimming of composite and complex printed documents. The model takes into account the approaches of page hierarchical scanning. The simulation process is based on the interest regions classification with emergent features, like salient contrasts and frequencies. Image processing is the tool that we have used for the realization of the cartography of salient interest areas which are successively selected to produce a balanced page scan.

4 Application to the simulation of complex pages layout retrieval

4.1 Principle of attention based selection

To evaluate the structural organization of the document, we have developed an exploratory strategy which is based on the previous behavioral analysis in a context of salient information capture. The system which is described here simulates the document skimming with a proposition of scan-path that captures the main structure elements of the page. The system has been tested on printed pages structure retrieval. We have essentially worked on newspapers, scientific papers, and advertisings. The structure retrieval is expressed by a reconstruction of informative blocks shapes and highlights the interest regions distribution on the overall page area. The result of the process can be compared to attention based segmentation. The successive stages of the process for blocks extraction are based on perception based rules for guiding the selection of the « next » fixation point.

The rules which have been used for this representation are essentially derived from the Gestalt Theory and from some more recent approaches developed in the 90's, [8, 11]. In our application, the main elements, which can be compared to invariants in perceptual organization, can be summed up in the properties of *symmetry, closure* (as a law of *perceptual stability) and the complexity* (as a factor of *good shape*). Thus, those psychological properties confirm the existence of active processes linking elements across visual space. The next section presents the overall experimental protocol and the main scientific and technical points for the realization of the system.

4.2 Methodology of the simulation

The analysis that we propose is primarily based on the collection of indices which are related to ocular movements and ocular tactics (in relation with the overall acquisition of information). It is based for a great part on the results concerning the evaluation of the Web pages and on for another part on Treisman's specific cognitive studies, [11].

This section presents the simulation method which is based all previously detailed perception phenomenon. The system displays the unequal importance of information in the visual field. The access of information is directly linked to the search of attractive areas. This search is based on the idea to free oneself from an unbending physical structuring and from a uniform vertical and horizontal scanning of the document, so as to classify the data in order of importance and interest.

The control of ocular saccades has led us to define some general rules for visual data capturing. Especially, it has been experimentally proved that *gazing points* tend to gather around *salient angles, angular points, curved lines, contrasted lines,* like beginnings or ends of lines, thick features, edges of forms...; all in all, elements which present important local discontinuity and which have in their neighborhood a particular local relief, [1]. Generally, we detect very quickly all deviations from the standard value of things. The observer's eyes, once fixed on the attractive points, assess the nature of the neighboring information, more or less precisely, depending on its distance to the fovea. What is more, some findings in cognitive sciences have shown that relevant objects are extracted by independent cortical units by a pop-out effect. Those units deal with color (as a textural information), with orientation (with the location of edges and corners), and with size (as a sampling representation of the retinal image). Those entire phenomenons have been used as inspiration sources for the system development, see fig. 3.

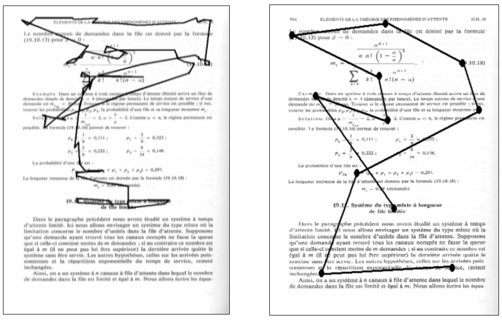


Figure 3 : Example of real scan-path (left) and simulated (right).

We have been inspired by this modular organization of the first stages of visual analysis, and have then decided to cut our process in as many units as necessary to recover texture, orientation and size-based information. Using a *space-variant geometry* for the block selection, the page image, instead of being represented by the bitmap format, can be abstractly represented by the block format. This space-variant geometry lays a sound basis for elaborating the kinetic of the ocular shifting on a document, which provides not only a meaningless document representation in blocks, but shows a unified view corresponding to the integration of time-variant representations of the same visual field, see figure 4. Thanks to the succession of representations of the same document, the resulting segmentation can be expressed by the integration of either low-resolution images or high-resolution images. The convergence to the finest segmentation consists in keeping for each pixel of the image its value corresponding to the higher resolution and displaying a detailed description of components. The dual operation (for the coarse page segmentation) only keeps the pixels with the lowest resolutions, so as to represent regions into coarse blocks. If the number of fixation points is not efficient, or if the successive points are not properly located, the coarse or fine description of blocks can not make the integration converge to a satisfying coarse or fine segmentation, [4]. The resulting drawing underlines the functional page structure, as it has been defined by Doermann in [3] and which corresponds to the visual page layout.

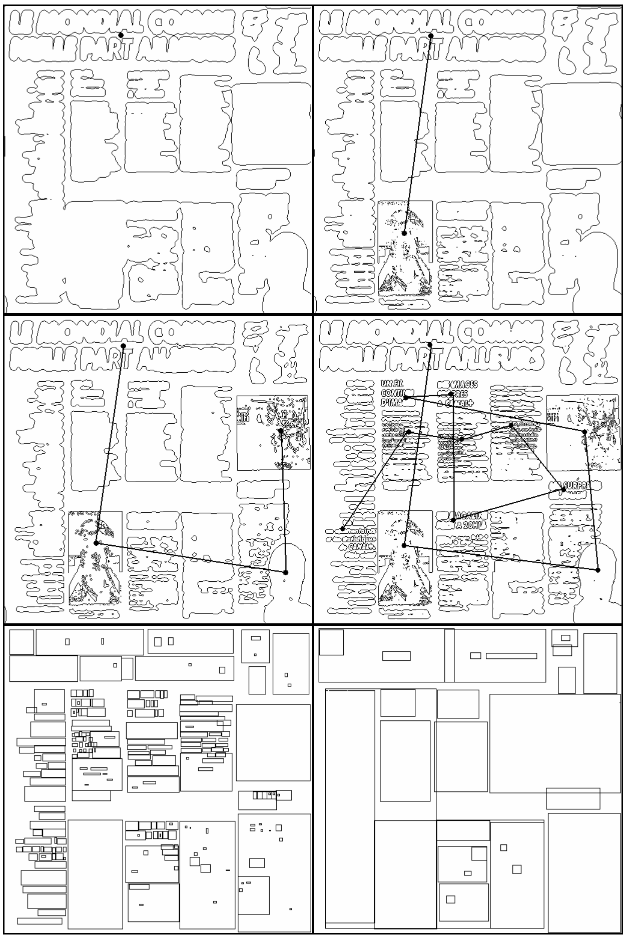


Figure 4 : Fixations convergence after 12 steps (only 4 represented here) to an unified representation where the succession of gazing points leads to the page reconstruction in low (right) and high resolution (left).

From this structure information, we can classify interest areas of the page. This first geometrical simulation has been recently improved by texture based primitives to classify the perceived regions by their own visual content, [4,5]. Finally, we will keep on improving those strategies by integrating more psychological human criteria. This part is currently under investigation.

5. Conclusion

In this work, we wanted to highlight the interest of scan-path analysis for self-governing documents images structure retrieval system. The system design has required specific protocols based the analysis on resulting scan-paths. To do that, we have focused our interest on salient areas which have shown that they are attractive for a reader. We have noticed that there does not exist an unique ideal scan-path during a document exploration: the readers' interest can be multiple and can be attracted by different visual salient areas. The simulation process gives a possible solution which is not unique but which brings relevant information of the analyzed page structure.

Acknowledgements: These studies had the support of the Rhone-Alps Region in the context of the SHIVA project (Emergence project).

6. References

[1] ADELSON, E.H., BERGEN J.R., *Early Vision,* Computational models of visual processing, Michael S.Landy, J.A.Movskon, p.3,1991.

[2] CAELEN, J., EGLIN, V., HOLLARD, S., MEILLON,B. Mouvements oculaires et évaluation de documents électroniques. CIDE pp.77-86 2003.

[3] DOERMANN, D., ROSENFELD, A., RIVLIN, E., «The function of documents», *Proceedings of the ICDAR-97*, vol.2, p. 1077-1081, 1997.

[4] EGLIN, V., Contributions à la structuration fonctionnelle des documents imprimés. Exploitation de la dynamique du regard dans le repérage de l'information, *Thèse INSA de Sciences Appliquées de Lyon,* 249p., 1998.

[5] EGLIN, V., GAGNEUX A., «Functional labeling and printed text featuring», *Proceedings of ICDAR-01*, p. 27-45, 2001.

[6] GAGNEUX, A. Structuration des documents par l'exploitation de la dynamique du regard chez l'homme - Application aux documents web, PhD Thesis, Nov. 2003.

[7] LEVY-SCHOEN A., « Les mouvements des yeux comme indicateurs des processus cognitifs », *Psychologie cognitive, modèles et méthodes,* P.U.G., p. 329-347, 1988.

[8]LECAS, J.C. *L'attention visuelle, de la conscience aux neurosciences : Problèmes fondamentaux et mécanismes de la perception visuelle.* Liège : Pierre Mardaga, 1992, 310p.

[9]NICOLAI M., Suivi du regard et ergonomie cognitive, DEA de sciences cognitives, Grenoble, 45p., 001

[10]RAYNER K., POLLATSEK A., «Eye Movements and Scene Perception», *Canadian Journal of Psychology*, 46, p. 342-376, 1992.

[11]TREISMAN, A. *L'attention, les traits et la perception des objets.*Paris : Folio Gallimard, 1992, 250p.