

提示型ユーザーインターフェースにおける視線計測器を用いた適合フィードバック

Viktors Garkavijs, Mayumi Toshima, Noriko Kando†

本研究は、画像検索のユーザーインターフェースを用いて、レバンス・フィードバックのための方法を比較するものである。視線計測器の使用により入力を行う新しいユーザーインターフェースを開発し、視線の停留時間の長さをもとに適合度のランキングをすることで画像の入れ替えを行うアルゴリズムを開発し、その有効性を問うための実験を行った。実験の参加者は6人（男性6人、女性2人）の日本人大学生・大学院生であった。画像検索の入力装置として視線を計測することで、一定の時間視線が停留した画像から適合性のある画像を抽出し、モニタースクリーン上に提示するシステムを使い、実験を行った。また、参加者が自分で適合度の高い画像を選び、マウスで画像をクリックするシステムを使用する比較実験も行った。実験の結果、どの参加者もマウスを使用した検索よりも視線計測器を使用したシステムの方がより迅速にレバントな画像を獲得することができた。

Using eye-trackers as relevance feedback input device in ostensive image search user interfaces

Viktors Garkavijs, Mayumi Toshima, Noriko Kando

This research is focused on comparing methods for relevance feedback in image search user interfaces. We have developed UI that uses eye-tracker as an input device and employed a relevance re-ranking algorithm based on the gaze length. The experiment results showed that there is a potential of using the eye-tracker as an input device for relevance feedback and it also reduces the time required to perform search tasks.

† National Institute of Informatics, The Graduate University of Advanced Studies [SOKENDAI]

1. はじめに

Web の普及は、これまでにないほど人間の生活に「情報検索」の活況をもたらし、今や Web を用いて何かが検索されない日はない。そして、Web の創成期には想定されていなかった多様な職種や年齢層の人々が、日夜、Web 上で情報検索を行っている。

現在の Web 検索では、利用者が検索窓に文字列をクエリとしてタイプし、検索結果一覧ページ (SERP: Search Engine Result Page) から、求める情報に関連するページへのリンクをクリックし、その情報を含むページにたどり着くのをめざすという方式が一般的になっている。

しかしながら、利用者は、必ずしも、常に、検索開始時に、どのような情報が欲しいかを明確に自覚しているわけではない。システムとのインタラクションを通じて、次第に、何が欲しいかが明確になったり、興味あることがらに気がつかされたりする。特に、画像の検索では、欲しいものの漠然としたイメージがあっても、それを的確に言葉のクエリとして表現することは難しい。そのため、画像の検索には、本質的に探索的な過程が内在し、検索システムがいかに利用者とのインタラクションを支援して、素早く、かつ、的確に、利用者が何を求めているかを捉えることが重要である。

そこで、本研究では、画像検索において、より直感的かつ迅速に、提示された画像への関心や適合性判断を利用者がシステムにフィードバックする手段として、視線を用いることを提案し、プロトタイプシステム GLASE0.1 を用いた利用者実験の結果を報告する。

視線による適合フィードバックは、マウスクリックによる方法と比べて、1) クリックするかしないかの二値ではなく、連続値として関心の強さを表現できる、2) 多数の適合情報のフィードバックが容易、3) マウスクリックによってポップアップした画像の表示時間で重み付けをして適合フィードバックをするよりも、positive abandon や multi-tasking による情報の混在を避けることができるため、より正確、4) 操作が容易、5) 情報量が多く、より高度な検索モデルの適用が可能といったメリットがある。また、このように視線を入力として用いる検索システムでは、システムの操作に不慣れであったり、操作に支障があったりする、初心者・子供・高齢者・身体的あるいは学習や知覚における何らかの障害を持つ利用者などの検索を支援するといった応用も期待される。

本稿では、言葉によるクエリ入力後に、視線を用いて適合フィードバックを行う方法を提案しているが、「The query is dead」[2]で知られる提示型モデル (Ostensive model) に基づく検索システムのように、初期画面で代表的な複数の画像を提示し、利用者に関心に応じて選択したものの類似の画像を探していくことで検索を進める、言葉によるクエリの入力を必要としない検索手法への応用も可能である。言葉によるクエリが

必要でなければ、何らかの理由で文字を読むことができなかつたり、考えを現在の検索エンジンが求めるような仕方で端的に言葉によるクエリとして表現することに問題がある利用者でも、イメージだけで自分の欲しい情報にたどりつく機会が増え、より多様な利用者が利用できる「ユニバーサル」な検索インタフェースとなると期待できる。

本稿の構成は、次節で画像検索インタフェースに関する既存の研究を概観し、視線計測器の主な用途や開発のあり方について簡単に説明する。3 節では、提案システムとして、視線計測器を用いた新しい画像検索システム GLASE を紹介説明する。4 節と 5 節で、GLASE0.1 を用いた検索実験とその結果について説明し、6、7 章では、実験から得た考察と今後の方針をまとめる。

2. 提案システム GLASE 0.1

2.1 データ

MIRFLICKR-1M のイメージコレクションから 786126 個の画像を抽出した。このコレクションはもともと 1,000,000 個の画像とその画像に対してユーザーから提供されたタグによって成り立っている。このコレクションでは 2 種類のタグデータがあり [1]、我々は「tags_raw」のファイルセットを使った。この tags_raw セットはユーザーが提供した形のままで使用されているタグのセットである。

2.1.1 データクリーニング Data-cleanup

ユーザーが提供した生のタグデータにはノイズがあり、ファイルの一部にはタグがついていないものもある。我々の目的はインタラクティブでなおかつ応答の速いインタフェースを作ることもあったので、速いインデックスを作ることが必要であった。ノイズを削減するため、またインデックス木の幅を縮小するため、データに対してクリーニングの処理をとった。タグセットの項目を ispell-enwl-3.1.20 の辞書ファイル（全 13 個のファイル）の項目と比較し、一致しなかったものをタグセットから削除した。

データクリーニング手続きの前とデータクリーニング手続きの後のタグデータの統計は表 1 に示されている。

Property	クリーニング前	クリーニング後
タグデータの総サイズ (バイト数)	117 977 254	45 313 089
タグデータのファイル数	1 000 000	786 126
ユニーク・タグ数	862 114	219 390
1000 回以上現れるタグ数	1 326	753

表 1. タグデータに対するクリーニング手続きの影響

2.1.2 インデックス生成

画像はタグによってインデックスされている。画像 ID の抽出手続きを加速させるためにインデックスをトライ木の様な構造に保存した。トライ木の枝はタグの文字であり、葉ノードでは画像 ID の一覧及びそのタグのスコアの値を保存している。すべてのタグの初期値は 0 である。

2.2 システム

2.2.1 スコア付け

パフォーマンスを比較するために、「表示時間」と「2 値」の 2 つのスコア付けの手法を実装した。表示時間の手法は、ポップアップの表示時間に基づいてタグにスコアを付ける。ポップアップが表示されていた時間のミリ秒数は各タグに均等に足される。

$$S_i = \frac{l_i}{|T_x|} \quad (1)$$

(1) 場合、

- i - イテレーション,
- s - スコア,
- l - ポップアップの表示時間,
- T_x - 画像 x のタグセット

とする。

2 値手法ではポップアップされた画像のタグに付けられるスコアは $\frac{1}{|T_x|}$ となる。

$$S_i = \frac{1}{|T_x|} \quad (2)$$

ユーザーの相互作用に適応させるために、重複のスライディング・ウィンドウを導入した。予備実験の結果に基づいてスライディング・ウィンドウのサイズを最適な値にした。最適な値としては、5 を選定した。(ただし、スライディング・ウィンドウの最適な値やそのパーソナライズ化についての研究は今後の課題とする)

i 番目のイテレーションの後のタグのスコアを下記の式の様に計算することができる。

$$\begin{cases} S_i = \sum_{j=1}^i S_j, & \text{if } i \leq w \\ S_i = S_{i-1} + S_i - S_{i-w}, & \text{if } i > w \end{cases} \quad (3)$$

この場合は、

w - ウィンドウのサイズ

S_i - i 番目のイテレーションの後のタグのスコアとする。

上記のスコアに基づいて適合度を下記の式の様に計算する:

$$R_x = \sum_q \sum_{t \in T_{X_q}} S_t \quad (4)$$

この場合は、

R - 適合度

x - 画像

q - クエリ中の語

X_q - タグ q を含む画像の集合

T_x - 画像の集合 X のタグの和集合

とする。

各 X_q の画像の適合度を計算した後、画像を適合度の高い順でソートし、SERP に表示する。

2.2.2 視線計測器の UI

このインターフェースの開発にあたっては、Tobii SDK 3.0 RC1 を使用した。このリリースは前のリリースと違って、Microsoft Windows の環境だけでしか使えない Microsoft COM の技術のサポートをやめて、他の OS とも互換性がある。API は 4 種類 (.NET, Python 2.6, C++, Cocoa) ある。我々は RAD (迅速なアプリケーション開発) のリッチな機能及びクロスプラットフォームの移植性を持つ .NET (C#) の API を選んだ。

実験では Tobii T60 の視線計測器を使用した。SDK によってサポートされている視線計測器であればどのようなものでも使用が可能である。

3. 実験

3.1 参加者

実験の参加者は、日常的に Web ブラウザおよび検索エンジンを利用している都内近郊の複数の大学の学部生・大学院生 6 名 (男性 4 名、女性 2 名) であった。実験機器として使用した視線計測器を過去に使用した参加者は 2 名 (参加者 2 および 5) であった。6 名全員が視線計測器を用いて入力する画像検索システムを使用するのは初めてであった。

3.1.1 実験環境および実験手続き



図 1. 視線計測器を使用した
情報検索システム



図 2. モニター画面の例

参加者は、実験の目的とタスクの概要、実験参加は自発的意思によること、いつでも中断可能であることについて説明を受け、同意書サインをした。実験に同意した 6 名の参加者が本実験に参加した。

各参加者は、事前アンケートの後、以下の 3 つのシステムのうちの 2 つを用いて検索をし、事後アンケートに回答し、インタビューを受けた。

a) ET (視線計測器): 視線による適合フィードバック。注視した時間による重み付け

b) MB (マウス・バイナリ): マウスクリックによる適合フィードバック。選択したか否かの二値

c) MD (マウス・タイム): マウスクリックによる適合フィードバック。クリック後の表示時間による重み付け

参加者は、システムの説明を受け、練習セッションを行い、2つのタスクについて検索を実施し、事前アンケートに回答した。これを各々の参加者が2つのシステムを用い、計4タスクを行った。使用するシステムと順序はカウンターバランスを取った。いずれのシステムを用いた場合も、検索中の視線を視線計測器で記録した。タスクは、それぞれ画像を提示され、それと似た画像を見つけることであった。使用した画像は、図 x に示す。

実験は、一人ずつ個別に行った。

実験設定の割り当てを表 2 に示した。表 3 はタスクに用いた画像を示す。

被験者 ID	1	2	3	4	5	6
タスクセット 1	MD t1, t2	MB t1, t2	ET t1, t2	ET t1, t2	ET t2, t3	MD t2, t3
タスクセット 2	ET t3, t4	ET t3, t4	MD t3, t4	MB t3, t4	MB t4, t1	ET t4, t1

表 2. 実験設定の割り当て

タスク番号	画像	画像 ID	タグ
1.		85926	bee, animal, animals, nature, green, verde, garden, flower, flowers, insect, black, work, plant, plants, wings, alas, eat, food
2.		585623	europe, france, paris, tour, eiffel, blue, tower, tour eiffel, eiffel tower, blue eiffel, tower, tour eiffel, lightening, star, stars, european, union, monument, illumination, night, by night, paris by night, paris la, light
3.		926255	architecture, building, geometric, skyscraper, sky, cloud, blue, white, window, wall, curve, line, urban, city, rectangle
4.		781859	san francisco bay area, california, usa, america, march, urban area, winter, west coast, san francisco, urban, light stream, moving, movement, light, lights, color, black, dark, night, low light, long exposure, above, over, view, northern california, purple, red, road, street, roadway, highway, interstate, freeway, hill, late night, image, city

表 3. タスクに用いた画像

#	英語	日本語	Score
1.	nature	自然	34218
2.	canon	キャノン (カメラの機種)	31320
3.	sky	空	29490
4.	blue	青	28887
5.	macro	マクロ	27070
6.	flower	花	25564
7.	water	水	25112
8.	red	赤	23733
9.	portrait	肖像画	22740
10.	green	緑	22529
11.	art	芸術	21032
12.	light	光	20082
13.	night	夜	19227
14.	white	白	19209
15.	film	フィルム	18484
16.	sunset	夕日	17809
17.	clouds	雲	17514
18.	street	通り	17069
19.	winter	冬	16127
20.	yellow	黄色	15858
21.	beach	ビーチ	15806
22.	architecture	建築	15800
23.	people	人	15709
24.	city	街	15517
25.	landscape	風景、見晴らし	14976

表 4. 実験時に使ったタグの辞書

4. 結果

1 タスクあたりの所要時間の平均と標準偏差を表 5 に示した。小数点以下は四捨五入してある。視線を使ったシステム (ET) では、マウスを使った他のシステムより、タスク完了時間が短かった。

	MB	MD	マウス	ET	All
Mean	137	125	131	113	122
St. dev.	60	38	48	24	38

表 5. 1 タスクあたりの所要時間の平均と標準偏差(秒)

このことは、検索結果一覧ページのヒートマップによっても説明できる

4.1 ヒートマップ

Tobii Studio によって生成されたヒートマップを分析した結果、検索者の eye-fixation points と刺激画像 (stimulus image) の相関が明らかである。

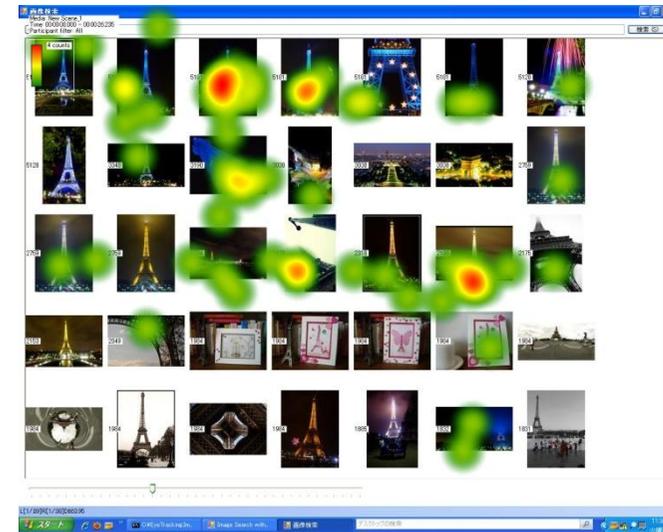


図 3. 被験者 3 タスク 2 のヒートマップ。この図においてはタスク (imageId=585623, 夜のライトアップのエッフェル塔の画像) に似たような画像と検索者の eye fixation points との強い相関が明らかである。

画像をクリックや視線の停留によってポップアップさせた回数とポップアップした画像の表示時間について、ヒートマップを作製した。(表 6 参照)。

	クリック数							ポップアップ表示時間						
MB n=128 t=153.114	3	4	9	3	3	5	2	2658	8169	9820	3441	2767	5720	2716
	4	10	4	6	5	2	2	3390	10298	3714	6173	4918	2458	2392
	2	5	5	2	3	2	2	1555	5959	8424	2502	2727	2308	2528
	2	3	1	4	9	4	3	3727	5245	738	4803	11128	6308	4708
	3	2	0	5	6	2	1	4252	1747	0	5622	6458	2447	1324
MD n=54 t=150.593	3	4	2	4	1	3	0	8646	28783	1696	13056	7279	11505	0
	3	2	2	2	2	1	2	8342	4709	1941	7710	6969	2102	2022
	0	0	0	1	0	0	1	0	0	0	3073	0	0	2796
	2	2	3	0	1	3	1	2284	4003	6802	0	508	2356	2590
	0	2	1	2	1	2	1	0	4581	645	2022	1134	11257	1782
マウス n=182 t=303.737	6	8	11	7	4	8	2	11304	36952	11516	16497	10046	17225	2716
	7	12	6	8	7	3	4	11732	15007	5655	13883	11887	4560	4414
	2	5	5	3	3	2	3	1555	5959	8424	5575	2727	2308	5324
	4	5	4	4	10	7	4	6011	9248	7540	4803	11636	8664	7298
	3	4	1	7	7	4	2	4252	6328	645	7644	7592	13704	3106
視線 n=245 t=245585	21	16	10	12	7	7	12	30256	17923	8936	4937	4241	11654	9778
	12	1	5	6	4	4	10	9747	599	6064	3183	2329	3404	6210
	6	4	3	5	7	9	4	6456	6314	5510	4700	8879	19519	3406
	3	4	1	9	8	9	4	3232	6302	535	7328	12700	8923	6074
	9	2	7	5	4	6	9	6534	1879	5654	2164	2880	2737	4598

表 6. SERP オブジェクトの位置のヒートマップ

視線を使ったユーザインタフェースは、ポップアップ回数が、マウスを使ったユーザインタフェースより、平均で 1.35 倍多かったが、ポップアップした画像を提示しておく時間は、他のユーザインタフェースより、平均で 1.24 倍短かった。この差は統計的に有意である。いわゆる「利用者が画像を見ただけで、クリックしないポジティブ・アバンドン ("positive abandonment")」は、視線を用いたユーザインタフェースでは捕捉することが可能であり、その情報を検索のための学習に用いることができるが、マウスを用いた方式ではこのような事象を捕捉することはできない。したがって、視線を用いたユーザインタフェースを有するシステムは、より短い時間で利用者の情報ニーズを満たすことが可能であった。

利用者の探索に必要な労力と時間は、視線を用いたユーザインタフェースの方が小さいが、より多くの情報がフィードバックされ、フィードバック回数も多いため、適合度の再計算の計算量は大きくなる。

5. 考察

5.1 実験の結果から見た視線を用いた画像検索の有効性

本稿では、視線を用いて適合フィードバックを行い、利用者が探している画像をより直感的かつ迅速に検索できるシステム GLAZE0.1 を提案した。このシステムを用いて、利用者が、所望の画像に類似した画像をどのくらい多く、かつ、素早く探せるかを調べるため、6 人の実験参加者を用いた検索実験を行った。その結果、以下のことが明らかになった。

1. マウスを用いた検索と視線を用いた検索とを比較したところ、視線を用いた検索の方がポップアップの回数が多かったが、利用者が満足して検索を終了するまでにかかる時間は短かった。すなわち、GLAZE0.1 の方がマウスを用いて適合フィードバックを行うシステムより、迅速に利用者の関心を満たすものを収集することができた。
2. ヒートマップを一定のエリアに区切って分析する手法を提案した。これにより、検索インタラクション中に利用者がスクリーン上で視線を配るエリアとマウスでクリックするエリアには偏りがあることがわかった。また、視線は停留したのにマウスではクリックしない例が多く、Positive Abandon が相当するあることが示唆された。
3. 1 回のポップアップあたりの平均の表示時間は、システムの種類によって有意な差があり、視線を用いたシステムが最少であった。

また、今回の実験の結果から示唆された課題としては、以下のようなものがある。

1. 探索エリアには偏りがある。今後は画像提示の場所の最適化を工夫する必要がある。
2. 1画面に提示する画像のサイズと配列については、予備実験では1ページに7×5枚の画像の提示が最適だとされたが、本実験ではウィンドウサイズと画像のボリュームが最適だったという報告はなく、1画面あたりの提示画像の枚数、レイアウト、画面構成については、今後の検討が必要である。

5.2 探索行動と時間から見た画像検索の特徴

今回の実験では、参加者6人に共通した探索行動を導き出すことはできなかった。実験前の仮説では、「視線計測器に抵抗感のある人は、マウス使用を好み、視線計測器で探索が速かった人は、視線計測器に満足感を感じる」としたが、そのどちらも検証されなかった。

参加者6人には、それぞれ探索行動に特徴があり、課題自体に難しさは感じなかったにもかかわらず、内観としては、「慣れていないので使いにくかった」という意見が多かった。しかし、過去に視線計測器を用いた実験に参加した経験が1回ずつある参加者は、本実験の4つのタスクを実施する間に、タスク完了時間が短縮されていく傾向がみられ、少しの使用経験が、タスクの習熟に好意的に働く可能性が示唆された。また、インタビューの場では全員が「今後、このシステムを使ってみたい」と答えている。さらに、「グループでイメージを共有、合意を取りながら検索をしたい場合」、「何かをしながら検索をする場合」など、具体的に視線を用いた検索が特に有用であると思われるケースを提案した参加者もあった。

実験中の姿勢については、視線を用いた検索であっても、マウスを置くあたりに手を置いている参加者もあったし、手を組んでゆったり椅子の背に持たれてリラックスして検索にのぞんでいる参加者もあった。

これらのことから、GLASE0.1は、いくつかの改善点を見出すことによって、よりユーザ・フレンドリーなインターフェースになる可能性があると思われる。

〈改善を検討す劇点〉

- 1 提示される画像のサイズと枚数の調整
- 2 提示される画像のレlevanceの検討
- 3 フィグゼーション・タイムの閾値の設定調整
- 4 利用者の個人差への対応（具体的には、個々の検索能力と検索状況の把握、学習との関係の考慮）

6. 将来の研究の方向性

本研究は、情報検索を行うユーザが、従来のキーボードとマウスを用いたシステムから、自分自身のアテンションを視線で表現することによって、探索のターゲットとなる画像を獲得することに対してユーザ自ら感じた有効性、抵抗感、満足度などを計測したものである。

手指を使用せずに視線だけで画像検索ができる方が、ユーザの満足度は高くなるのではないかという仮説を立て、この仮説を検証しようとしたが、実験の結果は参加者の「マウスの方が慣れているため、使いやすいと感じた」という意見が優位であった。しかし、視線計測器を以前使用したことのある参加者の意見は、必ずしもGLASE0.1に批判的ではなかったことから、

トレーニング・フェーズを活用することによって、ユーザが新しいシステムに慣れ、視線計測器を利用した方が、マウスよりも迅速かつ有効に画像検索ができるようになるということも考えられる。

特に、レポートや報告書を書きながらの検索は、手を使いながら作業をするため、GLASE0.1のハンズフリーという大きな特徴が活かせる可能性がある。たとえば、濡れた手で料理を作りながら視線の移動のみでレシピを検索する、機械を両手で組み立てながら、そのマニュアルを視線で探すなどが可能になる。また、高齢者や子供、そして障害を持つ人々など何らかの理由で既存の検索システムの利用が難しいユーザであっても、視線計測器のようなテクノロジーの改変による支援を受けることによって、最適な情報検索の利用可能性が考えられる。特に「イメージは分かっているけれど、言葉で検索を行うことが難しい」ものごとや運動行為などを探そうと思った際にGLASE0.1は最適な探索システムに発展する可能性がある。

さらに、今後テクノロジーの進化に伴い、今まで高価であった視線計測器が廉価になることも十分考えられ、現に医療研究用に使用されている視線計測器は、急速に価格が下がりつつある。安価なWebカメラを搭載した視線測定の研究も進みつつあることも踏まえると、一般の人々が近い将来デジタルカメラや携帯電話のような気楽さで購入できるようになることもあるだろう。

参考文献

- 1) Huiskes, M. J., Lew, M. S.: The MIR Flickr Retrieval Evaluation, ACM International Conference on Multimedia Information Retrieval (MIR'08), Vancouver, Canada. (2008).
- 2) Campbell, I.: Applying ostensive functionalism in the place of descriptive proceduralism: "the query is dead". In Workshop on Information Retrieval and Human Computer Interaction. University of Glasgow (1996).
- 3) Kelly, D.: Methods for Evaluating Interactive Information Retrieval Systems with Users, Foundations

- and Trends in Information Retrieval Vol. 3, Nos. 1–2, Now Publishers Inc. Hanover, MA, USA (2009).
- 4) Prasov, Z., Chai, J. Y.: What's in a Gaze? The Role of Eye-Gaze in Reference Resolution in Multimodal Conversational Interfaces, Proceedings of the 13th international conference on Intelligent user interfaces, ACM, New York, NY, USA (2008).
 - 5) Liu, T.: Learning to Rank for Information Retrieval, Foundations and Trends in Information Retrieval, Volume 3, Issue 3, Now Publishers Inc. Hanover, MA, USA (2009).
 - 6) Kumar, M.: Gaze-Enhanced User Interface Design, A Dissertation, Stanford University, Stanford, CA, USA (2007).
 - 7) Kumar, M., Paepcke A., Winograd, T.: EyePoint Practical Pointing and Selection Using Gaze and Keyboard. CHI '07 Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, NY, USA (2007).
 - 8) Oliveira F., Aula A., Russell D.: Discriminating the Relevance of Web Search Results with Measures of Pupil Size, CHI '09 Proceedings of the 27th international conference on Human factors in computing systems, ACM, New York, NY, USA (2009).
 - 9) Hansen D. W., Hansen, J. P.: Robustifying Eye Interaction. IT University, Copenhagen, Denmark (2006).
 - 10) Mollenbach, E., Stefansson T., Hansen, J. P.: All eyes on the monitor: gaze based interaction in zoomable, multi-scaled information-spaces. IUI '08 Proceedings of the 13th international conference on Intelligent user interfaces. ACM New York, NY, USA (2008).
 - 11) Porta, M., Ravarelli, A., Spagnoli, G.: ceCursor, a contextual eye cursor for general pointing in windows environments. ETRA '10 Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. ACM New York, NY, USA (2010).