

Navigation 3D pour Explorer une Vaste Surface 2D : Les Promesses de la Vue en Perspective

Yves Guiard² Olivier Chapuis¹ Yangzhou Du² Michel Beaudouin-Lafon¹

(1) *Laboratoire de recherche en informatique (LRI)*
Bat 490 Université Paris-Sud
91405 Orsay Cedex - France

(2) *Laboratoire Mouvement et Perception (LMP)*
CNRS (UMR 6152) et Université de la Méditerranée
Faculté des sciences du sport
163 avenue de Luminy, CP 910
13288 Marseille cedex 9 - France

Résumé

La plupart de nos interfaces graphiques implémentent un modèle de caméra relevant de la réalité virtuelle, la caméra en question étant une abstraction informationnelle dont les degrés de liberté sont offerts au contrôle de l'utilisateur. L'état de l'art repose essentiellement sur le principe d'une caméra virtuelle qui, surplombant la surface d'un document, peut être translatée dans les trois dimensions, l'orientation de la caméra restant fixe, orthogonale au document.

L'espace offert aux mouvements de la caméra est 3D, mais la scène survolée est un document électronique, exhibant du texte ou du graphique fixe ou animé, donc une surface plane. D'autre part, nous déclarons explicitement dans le modèle l'existence d'un véhicule virtuel, responsable des déplacements de la caméra, et dans lequel l'utilisateur, doté de la télécommande, se sent virtuellement transporté. Ces deux hypothèses de départ nous invitent à un rapprochement fructueux avec la question classique du contrôle visuel de la locomotion humaine. Nous analysons ainsi les déplacements du point de vue au dessus du document avec les outils conceptuels de l'optique écologique gibsonnienne, en mettant l'accent sur les flux optiques occasionnés par une exploration active du document. Nous dressons un inventaire des avantages nombreux de la vue perspective pour la visualisation interactive des documents de vastes dimensions auxquels nous sommes confrontés quotidiennement.

Nous avons testé un substantiel échantillon d'utilisateurs soumis à une tâche d'atteinte de cible très éloignées dans un premier prototype d'interface leur permettant de faire défiler un grand document en contrôlant librement l'inclinaison de leur caméra virtuelle. Les résultats, tant au plan de la performance objective que de la satisfaction subjective, confirment pleinement l'intérêt du nouveau mode de visualisation. Ce travail débouche sur une large palette de suggestions concernant l'enrichissement de l'interface graphique de base. Nous avons entrepris d'évaluer quelques unes des possibilités les plus prometteuses.

Abstract

Most of your current graphical user interfaces implement a virtual-camera model that pertains to virtual reality, the camera in question being an information abstraction whose degrees of freedom are under the user's control. The state of the art essentially relies on a camera that hovers over the surface of a document. It can be translated in three dimensions, but its orientation, perpendicular to the document plane, is fixed.

Even though the camera is moved in 3D space, the visualized scene — a document that exhibits text or graphics, the latter stationary or animated — is a planar 2D surface. We make explicit another important feature of the model: the camera is transported aboard a remote-controlled virtual vehicle in which the user, who holds the commands, feels self-transported. These two starting assumptions are an invitation to link electronic document navigation to a classic theme of psychology, visually-guided locomotion in humans. We tackle the problem from the viewpoint of ecological optics, placing emphasis on the optical flow fields that are elicited by active exploration. We draw a list of the many advantages of the perspective view for the interactive visualization of the vast documents users have to handle today.

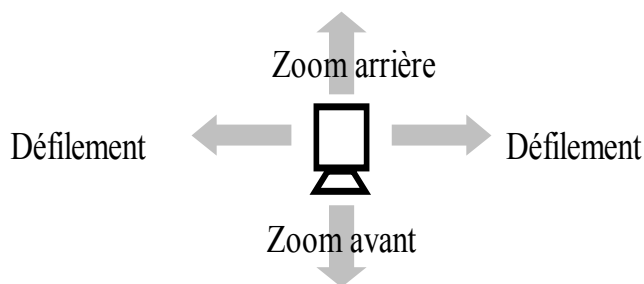
We have tested a substantial sample of users asked to reach and click remotely located targets in an interface prototype that allowed them to scroll the document while freely controlling the tilt of their virtual camera. Result support our expectations, showing that perspective visualization is both more efficient objectively and preferred subjectively. This research leads to a large palette of suggestions concerning the enrichment of the basic graphical user interface. We have started evaluating some of the most promising possibilities.

1. Introduction

Associant l'analyse théorique et l'expérimentation, cette étude explore l'espace de conception qui s'ouvre dès lors que, dans le modèle de la caméra virtuelle qui sous-tend la plupart de nos interfaces graphiques, on abroge la contrainte d'une orientation perpendiculaire fixe du point de vue sur les documents électroniques auxquels nos travaux nous exposent quotidiennement .

2. Le modèle de la caméra virtuelle

La plupart de nos interfaces graphiques implémentent un modèle implicite de caméra virtuelle, celui par exemple qu'utilise la programmation OpenGL. Cette caméra est une pure abstraction computationnelle dont l'utilisateur, toutefois, contrôle réellement les degrés de liberté grâce à aux dispositifs d'entrée (par exemple une souris) et de sortie (un écran), physiquement présents, qu'on met à sa disposition. Cette caméra virtuelle surplombe un document, qui peut être une page de texte, de code, une feuille de calcul, un atlas géographique, etc. Nos interfaces sont ainsi faites que la caméra, typiquement, peut être translaturée dans les trois directions de l'espace euclidien tandis que son orientation, perpendiculaire au document, reste fixe (Figure 1). Le changement d'altitude correspond au zoom, le changement de position dans un plan parallèle au document permet le défilement (*scrolling, panning*).



Plan du
document

Figure 1. Le modèle de caméra virtuelle à la base de la plupart de nos interfaces graphiques actuelle.

2.1. Une navigation 3D pour explorer une surface 2D

Même si c'est un espace 3D qui s'offre aux mouvements de la caméra, il n'en reste pas moins que la scène à explorer avec la caméra, un document électronique exhibant du matériel textuel ou graphique sur un seul plan, est une *surface*. Nous devons donc décrire avec notre modèle une navigation à trois dimensions au dessus d'une scène qui n'en comporte que deux.

Le contrôle réel d'une caméra virtuelle pose un authentique problème de réalité virtuelle, mais la navigation au dessus d'un plan a peu retenu l'attention des chercheurs se réclamant de ce domaine thématique, focalisés sur le cas de la scène intrinsèquement 3D dans laquelle, de surcroît, le cahier des charges impose typiquement une immersion de l'observateur. De fait ce sont sur des documents plans, et visualisés sans immersion, que les utilisateurs des interfaces graphiques génériques doivent travailler quotidiennement. Ceci justifie amplement que l'on travaille à comprendre ce cas de figure, même s'il ne pose plus de problèmes notables au plan de la technologie informatique. En fait, sa relative simplicité est un atout pour débusquer et clarifier quelques problèmes fondamentaux relevant de l'interaction homme-machine (IHM), à l'intersection de la psychologie et de l'informatique.

2.2. Véhicule virtuel

Il convient de déclarer d'entrée de jeu une composante tacite et pourtant capitale de notre modèle de caméra virtuelle: le *véhicule* transporteur de la caméra. Une caméra ne se déplaçant pas toute seule, il nous faut bien admettre qu'elle se trouve à bord d'un véhicule virtuel. C'est bel et bien ce véhicule que l'on télécommande quand on zoome et fait défiler le document. Et puisque déplacer une caméra en boucle fermée revient, pour l'utilisateur, à déplacer son propre point de vue, tout se passe pour lui comme s'il se mouvait *lui-même* au dessus du document. Le modèle informatique de la caméra virtuelle nous renvoie ainsi à un thème classique de la psychologie scientifique, celui du *contrôle visuel de la locomotion*.

Plus précisément, nous sommes en présence d'un modèle qui simule la locomotion humaine à bord d'un véhicule d'observation survolant la surface terrestre. La Figure 2 illustre de manière plus exacte le principe de visualisation interactive mise en œuvre dans les interfaces de l'état de l'art : ces interfaces nous confient la télécommande d'un véhicule d'observation inhabité capable de n'importe quelle translation au dessus du sol (le document) mais dénué de toute capacité de rotation, et dont la caméra fixe, arrimée sous le plancher, scrute le document perpendiculairement.

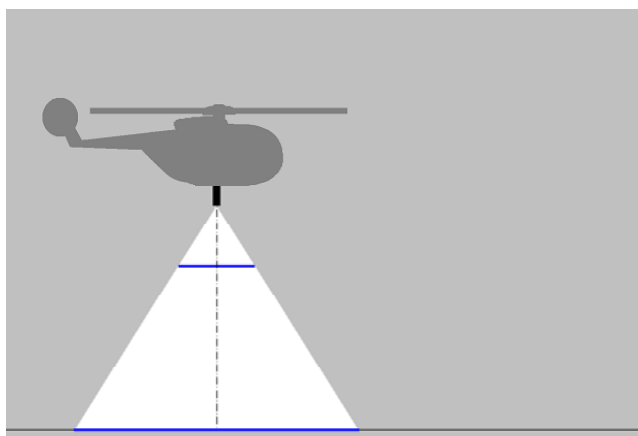


Figure 2. Le point de vue orthogonal que la technologie courante assigne au véhicule virtuel télécommandé de nos interfaces graphiques.

Il suffit de songer aux véhicules que nous connaissons dans le monde physique pour éprouver un doute quant à l'optimalité d'un tel arrangement. Une règle évidente qui guide la conception des véhicules automobiles, aussi bien terrestres qu'aériens, est de configurer l'habitacle — de placer le pare-brise, un terme révélateur dans notre langue — de telle manière que le pilote dispose d'une vue orientée, non pas vers le sol, mais *vers l'avant*, dans la direction du déplacement. En se conformant à cette règle, les concepteurs de ces véhicules ne font d'ailleurs rien d'autre qu'appliquer dans la sphère technologique un principe beaucoup plus général qui prévaut à l'évidence en biologie : chez les organismes vivants doués de capacités locomotrices, l'appareil visuel est, par défaut, orienté vers l'avant. Nous verrons plus loin combien l'arrangement de la Figure 2 est problématique quand nous examinerons, avec les outils de l'optique écologique gibsonnienne, les structures de flux optique occasionnées par les diverses orientations de point de vue lors de l'exploration d'un document.

2.3. Compléter le modèle: translations et rotations

La Figure 3 décrit le modèle de la caméra virtuelle dans le cas plus général permettant non seulement les translations du point de vue de l'utilisateur, mais aussi les variations de son orientation, ce qui implique d'introduire des inclinaisons de caméra. Cette figure (dans

laquelle le véhicule porteur, dûment enregistré, est redevenu implicite) illustre en quelque sorte le ‘débridement’ angulaire du modèle de caméra qu’implémentent nos interfaces graphiques courantes.

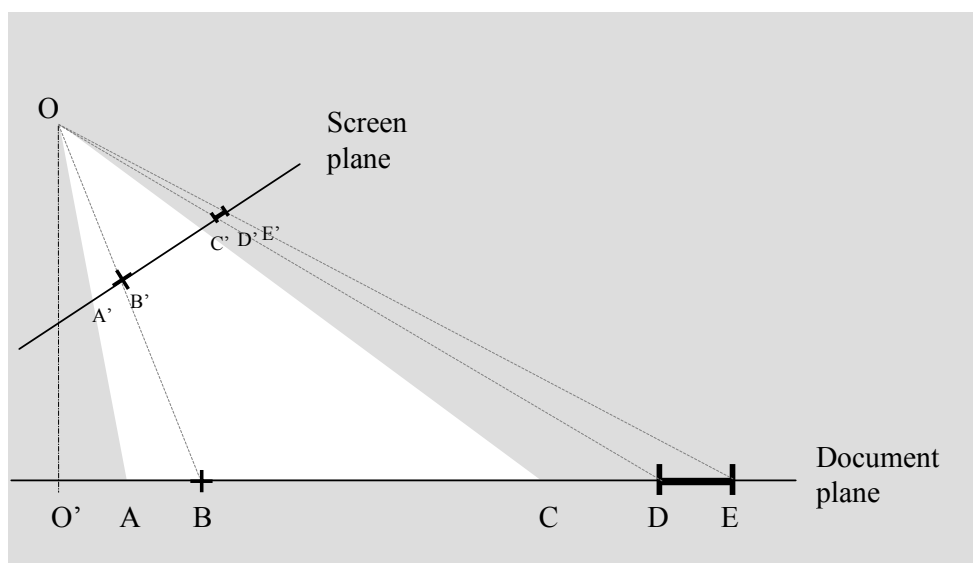


Figure 3. Le modèle de la camera virtuelle translatable et inclinable.

Le point d’observation est le point O et l’angle de vue est représenté par un cône clair sur fond sombre. La ligne continue horizontale représente l’espace du document, en le réduisant à une dimension. La ligne continue inclinée définit le plan de projection, qui s’identifie avec celui de l’écran. Il convient de distinguer ce que nous appelons la *sélection* (le segment AC), que nous définissons comme le sous-ensemble du document qui est visualisé, et la *vue* (le segment $A'C'$), définie comme le sous-ensemble de l’écran effectivement mis au service de la visualisation du document. La croix figurant au point B' représente le curseur, dont la raison d’être est de spécifier un point (B) dans l’espace du document, à l’intérieur de la sélection ; ce curseur appartient à l’espace de la vue, et il est enfermé dans l’intervalle $A'C'$. On a représenté sur le document une possible cible pour la navigation, l’intervalle DE , qui se projette en $D'E'$, donc hors de la vue dans cet exemple : pour atteindre cet intervalle cible, une navigation de caméra est nécessaire. La caméra est à une distance OO' du document (son altitude), et à une distance $O'D$ de la cible (dans les cas qui nous intéressent, on a typiquement $O'D \gg OO'$). La longueur focale f est la distance séparant le plan de projection du point d’observation. Nous dirons que les angles $OA O'$, $OB O'$, $OC O'$, etc., définissent les inclinaisons de vue pour les points A , B , C , etc. Un point singulier à noter est ce que nous appelons le *point de fixation*, le point du document désigné par la bissectrice de l’angle de vue AOC ; dans certaines applications ce point peut être spécifié avec une certaine précision grâce à mire fixe située au centre de la vue, au milieu de l’intervalle $A'C'$.

Il est utile de distinguer les deux manières possibles, complémentaires, de penser la mise en correspondance de l’espace (représenté) du document avec l’espace (représentant) de la vue, chacun de ces espaces pouvant servir de système de coordonnées pour décrire l’autre. La figure 4 illustre ce double système de description, en montrant l’effet dans les deux espaces, celui de la vue (à gauche) et celui du document (à droite), d’un mouvement de la caméra virtuelle analogue à une révolution ‘lunaire’ (voir la légende de la figure). La taille relative de la *sélection* (la portion du document qui est visible) par rapport au document entier ne peut évidemment être illustrée que dans l’espace du document. On observera de même que

c'est seulement dans le référentiel du document que l'on peut illustrer graphiquement quelle partie du document se trouve *exclue* de la sélection courante, une partie par définition non représentée dans l'espace de la vue. Or c'est au moment précis où l'attention de l'utilisateur quitte la sélection pour se porter vers une autre région, non visualisée, du document que commence à se poser le problème de la navigation documentaire (voir la Section 4.3).

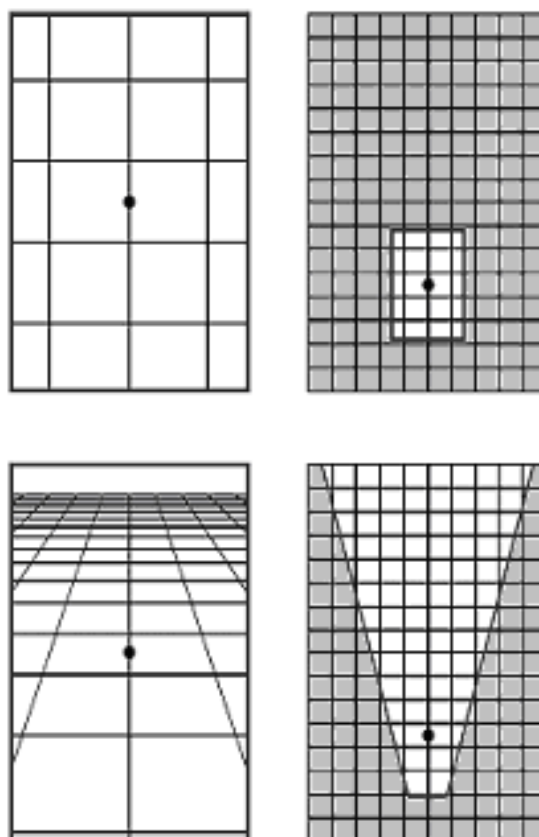


Figure 4. Effet d'une révolution 'lunaire' de la caméra (combinant rotation et translation de manière à toujours conserver le même point du document au centre de la vue) illustré dans l'espace de la vue (à gauche) et dans l'espace du document (à droite) où la sélection apparaît en blanc sur fond sombre.

2.4. La littérature : quelques prototypes d'interfaces avec inclinaisons de caméra

La littérature IHM rapporte un certain nombre d'études sur la navigation avec vue en perspective, mais à notre connaissance la question de l'inclinaison de caméra n'a nulle part été sérieusement traitée dans le cas du survol d'un document plan. Que peut-on gagner, à quelles difficultés peut-on s'attendre si l'on met à la disposition de l'utilisateur du véhicule virtuel un ou plusieurs degré(s) de liberté supplémentaire(s) pour l'autoriser à contrôler librement l'orientation de son point de vue sur le document ? Voilà des questions qui semblent avoir été seulement effleurées dans la littérature.

Ware et Fleet (1997) ont par exemple étudié un prototype de navigation à bord d'un véhicule volant virtuel, mais l'environnement qu'ils ont considéré était l'environnement 3D classique de la réalité virtuelle, un domaine de recherche où les rotations de caméra ont fait l'objet de nombreuses études (par exemple, REFS). Même si la réalité virtuelle constitue un vaste domaine de recherche, on s'y focalise sur la navigation dans les espaces 3D. On y laisse ainsi dans l'ombre le cas, pourtant particulièrement important dans nos interfaces de travail,

où l'environnement à explorer est une surface plane, ce qui est le cas quand on travaille sur du code, des nombres, du texte, ou des images.

Dans une étude récente, Chapuis et Roussel (2005) ont développé un gestionnaire de fenêtres graphiques, METISSE, qui rend possibles un grand nombre de manipulations de fenêtres, et notamment leur présentation en perspective (Figure 7). Conçu pour enrichir l'environnement graphique multi-fenêtres, ce prototype se situe, en termes de niveau de complexité, au dessus de notre modèle élémentaire de la Figure 3. Ici, en effet, la caméra virtuelle est implémentée en autant d'exemplaires qu'il y a de fenêtres affichées à l'écran. L'originalité de METISSE est en quelque sorte de mettre en scène une 'caméra-maître' virtuelle permettant d'obtenir sur l'écran maître — l'écran réel — une vue sur un ensemble d'écrans virtuels qui offrent chacun sa vue sur son document spécifique. Ces écrans virtuels étant disposés dans un espace 3D, un grand nombre de manipulations de fenêtres deviennent possibles. Dans la version courante du prototype, toutefois, le point de vue qu'offre sur son document chacune des caméras virtuelles composantes reste perpendiculaire.

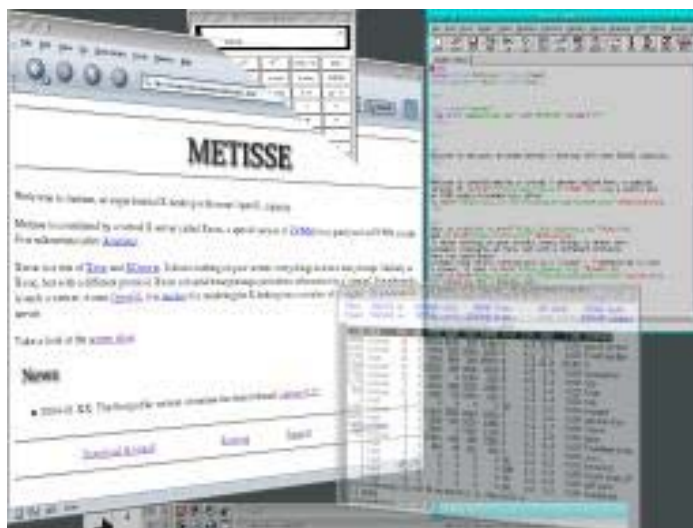
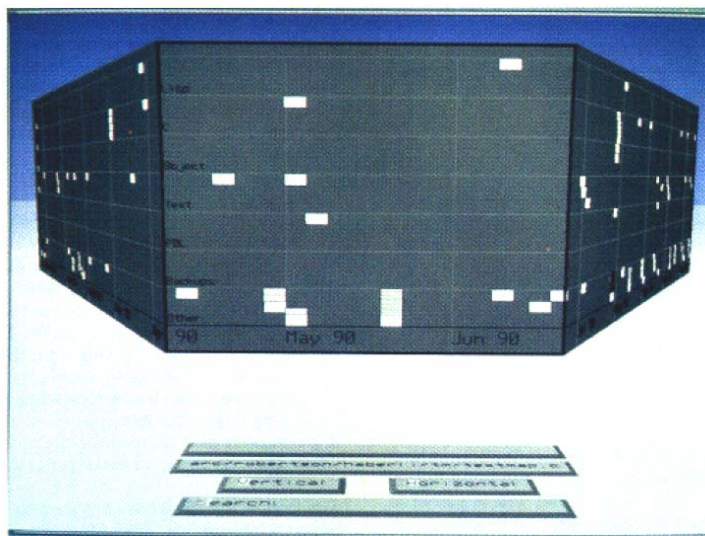


Figure 7. Un exemple de configuration de fenêtres dans METISSE (Chapuis et Roussel, 2005).

Mackinlay, Robertson et Card (1991) ont proposé le *Perspective Wall*, une technique de visualisation dans laquelle le document, plié en deux endroits, présente une large région centrale vue de face encadrée par deux régions plus étroites vues en perspective (Figure 5). Cette solution implique le sacrifice, non commenté par les auteurs, d'une propriété importante du document, sa configuration plane, ce qui de fait nous redirige vers l'attracteur de la réalité virtuelle à trois dimensions. D'autre part, alors que ces auteurs font valoir les avantages de la vue en perspective, leur technique, si l'on en juge par les surfaces respectives assignées aux trois régions du document, continue de faire la part belle à la visualisation orthogonale.



Mackinlay Plate 1

Figure 5. Le Perspective Wall de Mackinlay et al. (1991)

Rekimoto (1996) a décrit un prototype d'interface conçu pour les ordinateurs miniaturisés du type assistants numériques personnels dans lequel l'inclinaison de l'appareil, muni des détecteurs internes appropriés, est identifiée par le système comme une commande d'entrée (Figure 6). Parmi les diverses applications de cette invention (comme le pilotage d'un curseur au sein de menus), l'auteur mentionne incidemment une variante de l'interface où l'écran de l'assistant numérique fonctionne comme un fenêtre mobile ouverte sur une carte horizontale stationnaire : en faisant varier l'inclinaison de l'appareil, l'utilisateur obtient ainsi sur sa carte diverses vues en perspective. Toutefois, le bref compte-rendu de Rekimoto (1996), focalisé sur l'idée d'un nouveau type de commande adapté aux ordinateurs miniatures (incliner l'appareil, au lieu d'actionner un bouton physique), ne discute pas de l'intérêt potentiel d'une orientation non orthogonale du point de vue dans les interfaces graphiques en général.

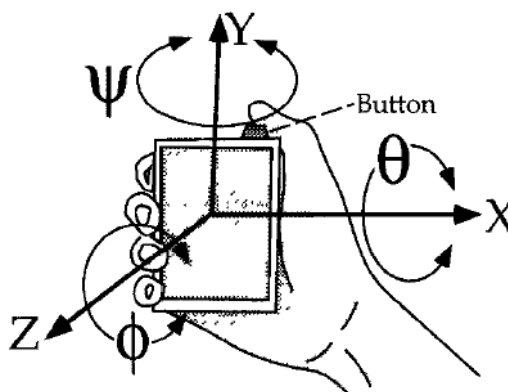


Figure 6. Le prototype de Rekimoto (1996)

De manière générale, il semble que l'on se soit peu soucié jusqu'à ce jour en IHM du fait que notre interface graphique standard, celle que nous utilisons quotidiennement dans nos activités professionnelles, continue d'exclure la possibilité d'une inclinaison de la caméra virtuelle sur les documents. Compte tenu de la puissance de calcul dont nous disposons

couramment aujourd'hui, il est évident que la contrainte d'orthogonalité du point de vue ne se justifie plus par des considérations de faisabilité technologique. A lire la littérature IHM, il ne semble pas qu'on ait une claire anticipations des propriétés de la vue en perspective. Peut-être se méfie-t-on aussi de ce genre de degrés de liberté : non seulement les rotations de caméra virtuelle recèlent-elles certaines chausse-trappes géométriques pour les concepteurs (e.g., Poupyrev et al., 2000), mais on souligne volontiers qu'elles sont souvent, pour l'utilisateur, source de désorientation spatiale voire de nausée, d'où un défaut de curiosité et un déficit de recherche.

3. Une approche écologique de la navigation électronique

Qu'est-ce qu'un document, et quel est le statut de ces documents électroniques dont nous disons qu'ils se prêtent à la navigation ?

3.1. De la feuille au document immatériel : des artefacts sur mesure pour la vision

Appréhendés en tant qu'objets matériels, les divers types de documents que nous connaissons dans le monde physique exploitent systématiquement la propriété fondamentale de la feuille : celle d'être un objet qui maximise son rapport surface/volume et donc son rapport surface/quantité de matière. Un livre est un empilement de feuilles; une carte routière revêt typiquement la forme d'une feuille de grande dimension qui se replie sur elle-même. La morphologie singulière de la feuille s'explique comme une réponse, naturelle ou artificielle, à un problème de lumière : la feuille végétale maximise, à masse constante, la capture de l'énergie radiante nécessaire à la photosynthèse ; quant à la feuille de papier de l'industrie humaine, elle maximise, à masse constante, la surface disponible pour afficher de l'information textuelle et graphique à l'intention du système visuel.

L'histoire de la fabrication des documents, depuis les ancêtres comme le papyrus et le parchemin jusqu'au livre en papier-bible, peut se raconter comme l'histoire d'une progression monotone de la quantité d'information transmise par gramme de matière. Supposons que l'on imprime 2 à 3000 caractères par feuille (une densité de texte assez ordinaire), soit environ 5000 caractères recto-verso, et que l'on utilise du papier ultra-fin (la limite supérieure aujourd'hui accessible est d'environ 50 gr/m²); puisque dans les langues naturelles la quantité moyenne d'information par caractère est de l'ordre de 1 bit (Shannon et Weaver, 1949), on parvient avec la feuille de papier moderne à un rendement informationnel de 1 à 2 kilo-bits par gramme de matière.

Évalué à cette aune, le document électroniques peut s'interpréter comme un nouveau pas en avant, et celui-là tout à fait décisif puisque, la masse du support devenant rigoureusement nulle, le rendement informationnel de l'objet — si l'on admet qu'il reste défini — atteint l'infini. Au lieu de posséder deux surfaces, un recto et un verso visibles en alternance, le document électronique se confond avec son apparence visuelle, il *est* sa propre surface, une surface à face unique dont la seule réalité est optique.

3.2. Importance des surfaces dans la vision naturelle

La notion de surface est essentielle à la compréhension de la vision. Comme le psychologue J. J. Gibson (1966, 1979) a été le premier à le remarquer, seules sont visibles les surfaces du monde : la structure du champ de lumière ambiante réverbérée que tout système visuel a vocation à échantillonner est essentiellement déterminée par l'arrangement spatial (*layout*) et les propriétés d'absorption et de filtrage des surfaces réfléchissantes qui composent le 'mobilier' de l'environnement. La notion gibsonnienne de surface relève de la topologie

davantage que de la géométrie : les surfaces que nous voyons, ce sont les interfaces entre les états de la matière, entre les états solide et gazeux (c'est le cas de la surface d'une feuille de papier), entre les états liquide et gazeux (la surface d'un lac) ou entre les états liquide et solide (le fond d'un lac). Puisque le document, nous l'avons vu, est essentiellement une surface, demandons-nous à présent comment sont vues les surfaces dans le monde réel.

3.3. Ubiquité du point de vue oblique sur les surfaces du monde réel

Les surfaces de nos environnements, depuis un point d'observation donné, ne sont visibles qu'en perspective. Il ne peut en aller autrement, quelles que soient les variations de la direction du regard, qui ne font que faire varier l'échantillonnage de la scène optique considérée (*the optical array*, selon la terminologie de Gibson). Dans le cas spécial d'une surface plane de petite dimension (par exemple, une photographie ou une pochette de CD), un observateur peut se donner une vue à peu près homogène de l'objet en l'orientant frontalement et symétriquement par rapport à lui (jusqu'à former un triangle isocèle du point d'observation aux bords opposés de la surface examinée). On peut alors parler d'une orientation orthogonale du point de vue, mais même dans ce cas limite, tous les points de la surface, sauf un, sont vus obliquement. L'ubiquité de la vision perspective est d'autant plus évidente que la surface considérée est de plus grande dimension : à l'échelle d'une rue ou d'un paysage, les effets de la perspective deviennent considérables et omniprésents. On peut donc dire que le point de vue oblique est la norme dans la vision naturelle.

L'effet de perspective, bien sûr, est négligeable face à une feuille de papier placée à une distance usuelle de lecture — un document au format A4, par exemple, peut même être lu sans difficulté en vision oblique. Mais la question du positionnement et de l'orientation du document par rapport à l'observateur devient évidemment beaucoup plus sensible dans le cas d'un document étendu sur plusieurs m² ou davantage, comme ceux que nous donnent à voir aujourd'hui les technologies informatiques.

3.4. Une limite à la quantité d'information visualisable par scène optique

Considérons le cas d'une personne placée face à une page de texte. Le format ordinaire de la feuille de papier, qui contient quelques milliers de caractères — et donc, selon Shannon, porteuse de quelques kilo-bits d'information textuelle¹ — est tel que normalement la page est intégralement lisible grâce au seul balayage saccadique du regard, sans déplacement du point d'observation. L'information contenue dans cette page est assimilable alors que la page reste contenue dans une unique scène optique, ce que Gibson appelle *the optical array*.

Deux facteurs déterminent la charge informationnelle de cette scène, la densité du texte (taille des caractères, interligne) et la surface de la page. Mais la distance d'observation de la page ne peut excéder un certain *maximum* parce que la résolution de la vision fovéale est physiologiquement finie. La distance d'observation se heurte d'autre part à un *minimum* puisque la qualité de la vision va se dégrader, cette fois pour des raisons optiques, avec l'accroissement accéléré de l'obliquité du regard vers les régions périphériques de la page (que l'on suppose plane). Ainsi la marge de liberté de l'observateur est-elle étroite quant au choix de sa distance d'observation, une distance qui est encadrée par un maximum (contrainte physiologique) et un minimum (contrainte optique) proches l'un de l'autre.

Il ne servirait à rien d'accroître inconsidérément la densité du texte ou la surface de la page : il nous faut reconnaître chez l'humain l'existence d'une limite supérieure, le *maximum*

¹ Selon l'estimation de Shannon et Weaver (1949), la quantité moyenne d'information véhiculée par caractère est dans les langues naturelles de l'ordre de 1 bit (loin du maximum $\log_2 27$) du fait du niveau élevé de redondance.

d'information assimilable par scène optique (MIVSO), cette limite correspondant au maximum d'information que peut offrir un document depuis un point d'observation donné. Ce simple fait paraît expliquer à lui seul la nécessité de ce que désignons métaphoriquement comme la « navigation » dans les documents de grande taille. La valeur réelle du seuil informationnel à partir duquel la navigation est nécessaire, une valeur que l'on doit supposer très variable inter-individuellement, reste à évaluer, mais on peut en donner un ordre de grandeur. Si l'on se rappelle les divers stratagèmes de pliage auxquels ont recours la plupart des lecteurs de journaux quotidiens, on peut supposer que, pour un sujet en pleine possession de ses moyens visuels, cette capacité maximale d'absorption informationnelle propre à la vision stationnaire ne se situe pas très loin de l'information contenue dans une page au format A4 (ou *US Letter*) efficacement mise en page.

3.5. Echelle des surfaces et mouvement : manipulation et locomotion

Il est utile d'introduire ici la notion d'échelle, que nous définirons de manière relative, par rapport aux dimensions de la personne. On déplace par rapport à soi les objets de taille inférieure à la sienne (suivant l'usage introduit par Gibson, 1979, nous pouvons parler génériquement de *manipulations*), mais on se déplace soi-même par rapport aux objets de dimensions nettement supérieures à soi (c'est la notion classique de *locomotion*). Par exemple, on peut saisir et lancer un caillou, un mouvement qui se décrit commodément dans un système de coordonnées ego-centré comme un geste corporel ; mais on doit contourner ou escalader un rocher, ce qui implique un déplacement du corps, considéré comme un tout, par rapport à l'environnement. Ainsi le continuum physique de l'échelle présente une certaine région critique du point de vue de la psychologie: de part et d'autre du niveau d'échelle correspondant à la taille de l'organisme qui nous intéresse, on observe de la manipulation et de la locomotion, deux types tout à fait distincts de comportements moteurs.

Les documents, en tant qu'objets matériels, ne font pas exception. Si le contenu d'un livre s'explore par des manipulations (on le saisit, on le compulse, ce qui n'empêche en rien de rester assis), il devient en revanche difficile de manipuler une carte routière déployée sur une table : pour explorer un tel objet, on se tiendra plutôt debout et l'on se déplacera sans doute soi-même, au minimum avec des mouvements de la tête et du tronc ; face à une information affichée, la lecture devient un exercice de marche. Ainsi, on peut dire que le document, selon sa taille, sera quelque chose que l'on manipule ou bien au contraire un élément de l'environnement par rapport auquel on se déplace.

4. Potentialités de la vue en perspective

4.1. Familiarité de la vue en perspective

La vue en perspective, nous l'avons vu, correspond au cas de figure le plus courant dans la vision naturelle. Les documents de la réalité physique paraissent constituer un cas particulier, puisqu'on tend à les examiner en vue orthogonale. Mais ceci n'est possible que grâce à une série d'inventions techniques qui ont permis de ramener aux dimensions d'un objet manipulable une surface documentaire parfois considérable — par exemple, 50 pages au format A4 imprimées recto-verso représentent plus de 6 m²). Il est important de réaliser que des inventions comme celles du rouleau puis du *codex* répondent à des contraintes du réel que l'on ne retrouve pas nécessairement dans les mondes électroniques. Ainsi, la locomotion humaine est énergétiquement coûteuse, lente et terrestre, et d'autre part les humains ont le souci de limiter l'encombrement de leurs habitations. Mais une caméra virtuelle peut se

déplacer à n'importe quelle vitesse concevable et voler librement, pour un coût énergétique nul, et l'espace d'information électronique est d'étendue potentiellement infinie.

Une interface qui permette de naviguer une vaste surface documentaire déployée vue en perspective, placera l'utilisateur dans une situation à la fois familière et inédite : familière parce que c'est par rapport à des surfaces vues en perspective que se déroule normalement la locomotion, inédite parce que les ressources disponibles pour se déplacer sont sans limites et que l'environnement qui lui est proposé est une surface documentaire.

Il n'y a pas de raison qu'une telle interface modifie les conditions de lecture des documents. Aussi longtemps que l'utilisateur est engagé dans une tâche spécifique (lecture, écriture, édition etc.) il ne navigue pas, et l'on peut supposer que le point de vue orthogonal classique convient tout à fait. C'est à l'évidence pour les actions du type navigation — l'ensemble des opérations permettant des déplacements du point de vue sur l'espace documentaire, soit dans un but exploratoire soit pour atteindre une cible définie située hors de la vue courante que notre essai d'enrichissement de l'interface courante semble se justifier.

4.2. Une variation d'échelle simple, monotone, non-arbitraire, et de gamme étendue

Comme l'ont noté Mackinlay et al. (1991), la vue perspective est particulièrement attractive pour la visualisation des documents de grandes dimensions parce qu'elle fait varier l'échelle de visualisation dans l'espace de la vue selon une loi progressive et simple.

Le fait que l'échelle de visualisation varie dans l'espace de la vue en perspective, plutôt que dans le temps, permet de contraster cette technique avec le pan-zoom, la technique la plus souvent mise en œuvre dans la technologie courante : les différentes échelles, au lieu d'être disponibles successivement, sont dans la vue perspective disponibles *simultanément*.

La visualisation perspective possède d'autres propriétés spécifiques de nature à lui donner également l'avantage sur les autres techniques de multiplexion spatiale.

Alors que la vue multi-focale et le fish-eye introduisent dans la vue des discontinuités plus ou moins abruptes, la variation d'échelle dans une vue en perspective est *graduelle*.

Ensuite, la technique du fish-eye repose sur une variation non monotone et arbitraire de l'échelle, introduisant des phénomènes de distorsion optique dont on sait qu'ils rendent malaisée la sélection des objets graphiques avec un curseur d'écran (Gutwin, 200X). En visualisation perspective, au contraire, l'échelle suit une loi de variation qui est à la fois *simple, monotone et non-arbitraire*.

Enfin, la vue en perspective permet de couvrir une *gamme d'échelle* beaucoup plus étendue qu'avec un fish-eye ou une vue multi-focale, même si à cet égard la technique du pan-zoom paraît indépassable (Guiard & Beaudouin-Lafon, 2004).

4.3. Une information pertinente au regard de la distribution du champ d'attention

Un défaut majeur de la technique pan-zoom courante, et qui disparaît avec la visualisation perspective, est d'interdire à l'utilisateur une exploration *sélective* du document dans la direction qui l'intéresse. Examinons comment se pose le problème de navigation, et comment il se résout dans le temps, en comparant les techniques du pan-zoom et de la visualisation perspective.

L'intention de naviguer trouve son origine chez l'utilisateur dans le constat d'un décalage entre ce qu'il voit et ce qu'il désire voir : c'est à ce moment qu'est conçu le projet de modifier la sélection courante représentée dans la Figure 8 par le rectangle S_1 . L'utilisateur connaît, au moins approximativement, l'emplacement de son nouveau point d'intérêt, cible des opérations de navigation qu'il se prépare à effectuer — il y a toutes chances que la direction, au minimum, soit définie : dans la figure on suppose que cette cible est située quelque part vers le haut du document. Or, en dépit du fait que l'utilisateur désire 'aller voir' vers le haut, le zoom arrière de la technique courante va lui imposer d'étendre sa sélection, de manière indifférenciée, dans toutes les directions. La sélection S_2 ainsi obtenue après le zoom arrière inclut la cible, mais c'est pour la noyer au sein d'une vaste étendue presque entièrement dénuée de pertinence : non seulement cette nouvelle sélection inclut toute la partie inférieure du document, que l'utilisateur ne désirait aucunement explorer, mais de surcroît le nouvel espace visualisé est majoritairement vide.

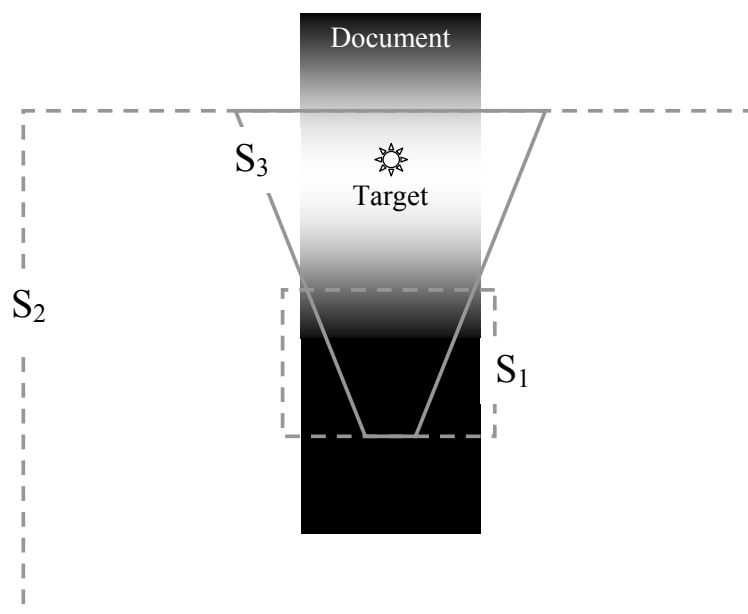


Figure 8. Changement de la sélection opérée au sein du document lors de d'un zoom arrière (de S_1 à S_2) et lors d'une inclinaison de caméra (de S_1 à S_3). Le document est le rectangle rempli avec des nuances de gris codant la distribution instantanée, dans l'espace documentaire, de l'intérêt de l'utilisateur.

Comme le montre la Figure 8, une inclinaison de caméra, qui correspond à la transition de S_1 à S_3 , nous épargne ces écueils. Au lieu de faire l'objet d'une homothétie indifférenciée, la sélection initiale se déforme en trapèzoïde, une forme qui semble bien adaptée au cahier des charges de l'utilisateur : en effet, d'une part la nouvelle sélection comprend sélectivement les régions situées dans la direction pertinente (à l'exclusion de toute la partie située en dessous de la sélection initiale), d'autre part elle n'inclut qu'une proportion raisonnable d'espace vide, enfin elle offre un gradient instantané d'échelle qui permet à l'utilisateur de continuer à voir, à un haut niveau d'échelle, son point de départ et, à un niveau plus bas d'échelle, la cible de son exploration, insérée dans son contexte.

La Figure 8 représente, avec des nuances de gris, ce que nous appellerons la distribution du *champ d'intérêt* de l'utilisateur dans l'espace du document. On doit imaginer ce champ d'intérêt hypothétique comme éminemment changeant dans le temps : la figure

schématise donc un instantané de la distribution de l'attention au moment où l'utilisateur conçoit le projet de rendre visite à une autre région du document. L'utilité de cette représentation est notamment de nous rappeler que l'intention de navigation est orientée, comme est orientée toute intention de déplacement locomoteur. Ainsi, nous pouvons distinguer une direction 'vers l'avant' et une direction 'vers l'arrière' chez l'utilisateur avant-même que ne commence ouvertement l'activité de navigation. Mais c'est l'analyse de la dynamique de la navigation qui va le mieux mettre en relief l'intérêt de ce type de distinction.

4.4. Des champs de flux optique plus pertinents

L'étude du contrôle visuel de la locomotion, un thème abondamment traité par la psychologie scientifique et les neurosciences, a été l'occasion pour Gibson (1966, 1979) de développer une discipline entièrement originale, l'optique écologique. Reléguant délibérément au second plan le thème de l'image rétinienne et les problèmes de physiologie nerveuse qui lui sont associée, l'optique écologique s'intéresse en priorité aux propriétés des champs de lumière réverbérée dans lesquels nous sommes immergés, des champs qu'échantillonne activement le système visuel — l'ensemble des organes sensoriels et moteurs co-impliqués dans la recherche active de l'information visuelle.

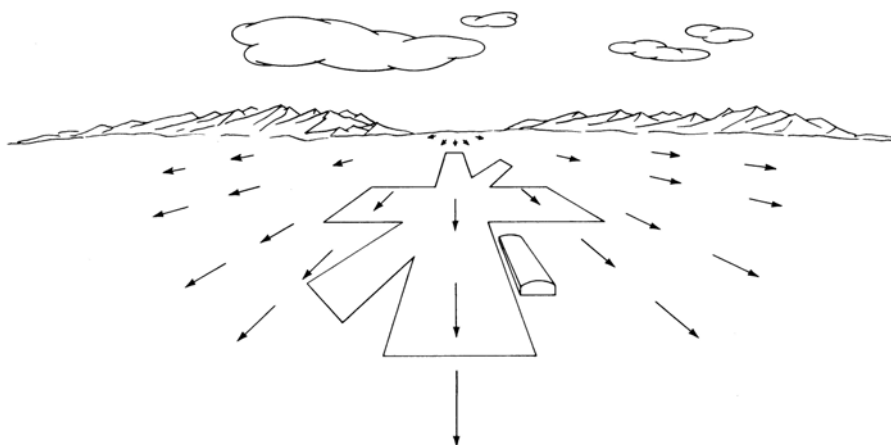


Figure 9. Le champ de flux optique occasionné par un vol parallèle au sol, selon le point de vue du pilote (Gibson, 1979, p. 124, Figure 7.4).

Un concept central de l'optique écologique gibsonnienne est celui de *champ de flux optique*. La Figure 9, empruntée à Gibson (1979), schématise le champ de flux optique accessible à un pilote d'avion se rapprochant d'une aire d'atterrissage, quand il regarde vers l'avant — une option pour un passager, une obligation pour le pilote. Chaque vecteur (dont il faut imaginer qu'il a été tracé, non dans l'environnement, mais sur l'écran) définit une valeur locale de vitesse de déplacement dans le plan de projection d'un quelconque grain de texture optique — n'importe quel élément visuellement discernable dans le champ. Ce schéma, qui montre une expansion radiale centripète de la vitesse optique, permet de comprendre un certain nombre de faits de première importance pour le guidage visuel de la locomotion. Par exemple le foyer de ce pattern radial centrifuge spécifie, dans l'environnement, la cible courante du déplacement de l'observateur : dans l'exemple, l'avion progresse essentiellement à l'horizontale, puisque le point futur de collision spécifié dans le flux correspond à la ligne d'horizon ; le pilote perçoit également que son appareil — si les conditions courantes du déplacement perdurent — va bientôt s'engager dans l'espace que l'on aperçoit entre les deux

massifs de collines. Ce pattern radial est fortement asymétrique, avec des vitesses angulaires qui s'accroissent de haut en bas dans le champ de flux optique — une parfaite symétrie radiale correspondrait au cas du piqué.

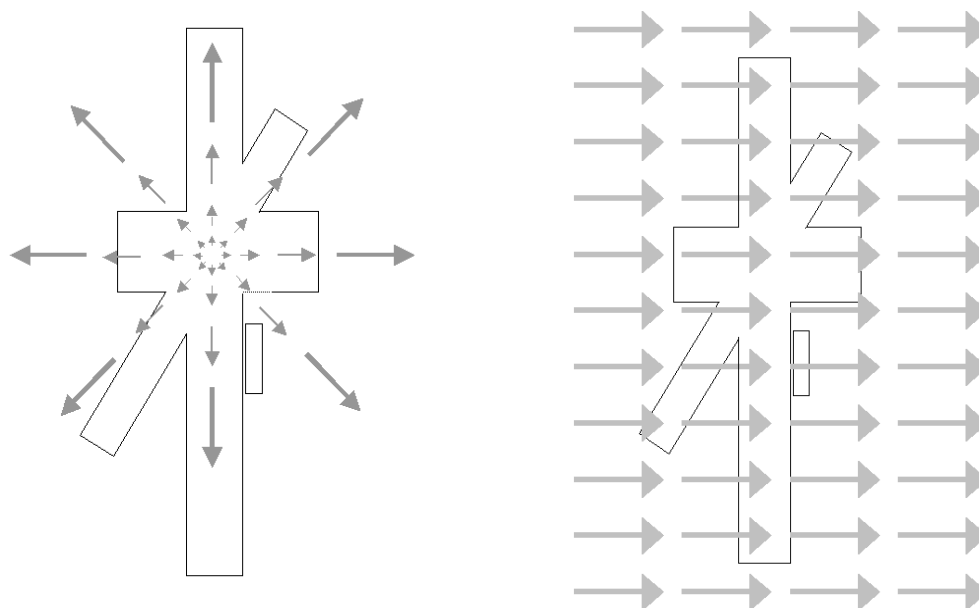


Figure 10. Les champs de flux optique délivrés sur l'écran lors d'un zoom avant (gauche) et d'un défilement (droite) dans l'interface de l'état de l'art.

La Figure 10 reprend la configuration au sol imaginée par Gibson pour montrer les champs de flux optiques auxquels est exposé l'utilisateur des interfaces zoomables de l'état de l'art lors d'un zoom avant (à gauche) et lors d'un défilement du document (à droite). Le zoom avant occasionne un flux radial centrifuge symétrique, spécifiant un piqué vers le document. Quant au défilement (*panning, scrolling*), il donne naissance à un flux laminaire homogène spécifiant au contraire que l'observateur se déplace parallèlement à la surface du document.

Ces deux derniers types de champs de flux optique, les seuls que l'on puisse générer dans nos interfaces courantes, sont l'un et l'autre fort problématiques, mais pour des raisons bien différentes. Le *flux radial* fourni par le zoom délivre à l'utilisateur la meilleure information possible pour contrôler son déplacement en cours, mais le déplacement en question, une variation d'altitude, ne rapproche aucunement l'utilisateur de sa cible — ce n'est qu'un détour stratégique servant à faire varier l'échelle de visualisation.

Quant au *flux laminaire* que génèrent les défilements, il délivre de l'information sur un mouvement qui, celui-là, rapproche utilement l'utilisateur de sa cible, mais l'information délivrée est de fort mauvaise qualité : le problème en l'occurrence est que l'interface impose au pilote du véhicule virtuel un point de vue particulièrement inapproprié, celui d'un passager regardant l'environnement à travers une fenêtre latérale du véhicule.

Le flux radial asymétrique de la Figure 9 et le flux laminaire de la Figure 10 ont en commun de spécifier une progression de la position de l'observateur par rapport au sol (ou au document) : ces flux renseignent donc sur des déplacements directement utiles. Mais de ces deux champs de flux, seul le premier est porteur de l'information *prospective* dont le pilote a besoin. Le flux laminaire, certes, autorise la perception du mouvement (une appréciation de la

vitesse notamment), mais il ne permet en aucun cas de guider un véhicule vers sa cible, celle-ci ne figurant pas dans le champ.

Igarashi et Hinckley (2000) ont mis au point une technique visant précisément à atténuer cet inconvénient de l'interface pan-zoom. La technique consiste à faire dépendre à chaque instant l'altitude de la caméra virtuelle de sa vitesse courante de progression horizontale, le zoom devenant de la sorte une fonction automatique. L'efficacité de cette technique, récemment confirmée (REF), tient probablement au fait qu'elle améliore la qualité de l'information prospective, celle relative à la réduction de l'écart à la cible. Mais, le zoom arrière accroissant la sélection dans toutes les directions, il faut noter que ce procédé accroît dans les mêmes proportions l'information prospective et l'information rétrospective, cette dernière probablement en pure perte.

4.5. Implosion de l'échelle et sélection de cible

Les sections précédentes, où il a été question de navigation dans les documents, nous conduisent vers une évaluation critique des interfaces graphiques à l'aune de leur capacité à assurer le bouclage dynamique de la perception et de l'action. Comme ces activités de l'utilisateur ne sont pas toutes modélisables selon la métaphore du déplacement locomoteur, il semble important d'identifier également les propriétés de la vue perspective dans des conditions statiques, en considérant spécialement le cas de la sélection des objets graphiques au moyen du pointage de curseur.

L'instrument de pointage en quoi consiste le curseur d'écran appartient à l'espace de projection, comme son nom l'indique. Confiné à l'intérieur de la vue, il ne peut désigner que des points appartenant à la sélection courante (Figure 3). Dans la technologie courante, la désignation d'un objet du document situé loin de la sélection courante requiert une séquence de navigation assez complexe, initiée par un zoom arrière ; or l'effet immédiat de cette première opération est d'interdire toute désignation d'objet, du simple fait que la réduction d'échelle affecte uniformément toutes les cibles potentielles.² A la suite d'une inclinaison de caméra, en revanche, le pointage reste partiellement possible puisque le rescalement des objets graphiques dans la vue est graduel.

Le problème de la difficulté du pointage en vision perspective recèle un double enjeu, puisque le pointage sert non seulement à la sélection des objets graphiques, mais encore à la sélection du *point de saisie* à partir duquel le document peut être dragué.³ Dans les deux cas, la difficulté du pointage va évidemment dépendre de la distance de l'objet choisi pour cible, et donc de la position de sa représentation dans le gradient d'échelle. Pour évaluer jusqu'où il est possible de pointer dans une vue en perspective, il nous faut préciser la loi de variation de l'échelle dans les vues obliques (pour une analyse mathématique plus précise de cette question, voir l'Annexe 1).

² En pratique, quand le nombre de pixel disponible pour la représentation de l'objet graphique passe sous un certain minimum, ou bien l'objet disparaît, ou bien il est remplacé par un marqueur symbolique de taille conventionnelle, selon le principe du zoom sémantique (REF).

³ A condition de décider d'ignorer certaines connotations parasites, reconnaissons que la métaphore du dragage des fonds portuaires fonctionne aussi bien dans notre langue que dans la langue anglaise.

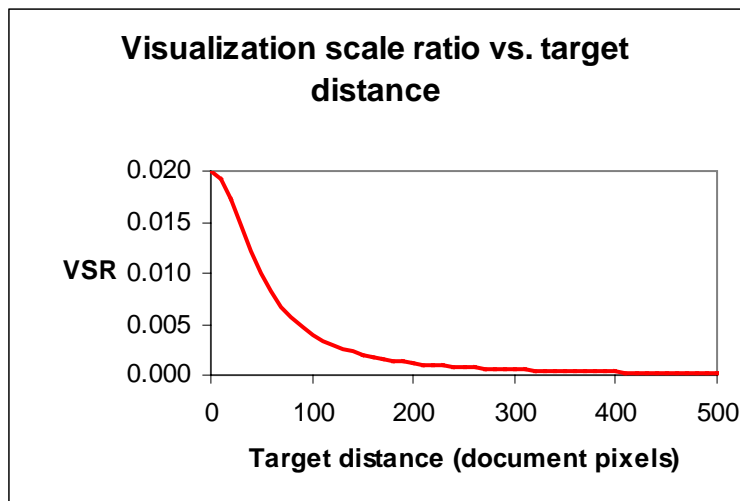


Figure 11. L'échelle, définie comme le nombre de pixels d'écran par pixel de document, en fonction de la distance de l'objet dans le document.

Comme le montre la Figure 11, l'échelle de visualisation de l'objet graphique varie comme une fonction non-linéaire de sa distance horizontale au point d'observation. Quand l'objet s'éloigne, l'échelle commence par se réduire de manière accélérée, puis la décroissance ralentit pour se rapprocher asymptotiquement de zéro.

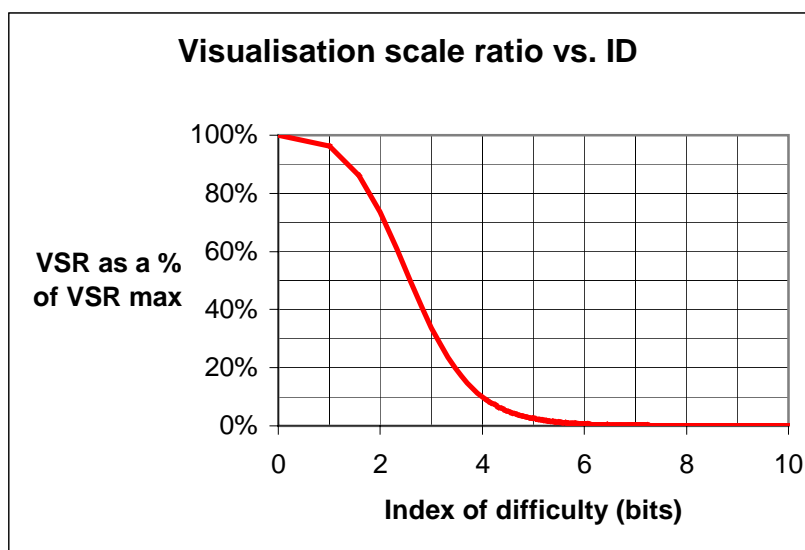


Figure 12. L'échelle exprimée en pourcentage du REV maximal, en fonction de l'*ID* de Fitts (1954).

La Figure 12 exprime l'échelle en pourcentage du maximum possible pour une focale et une altitude données de caméra, ce maximum correspondant au point situé à une distance horizontale nulle et donc observable sous un angle de projection de 90° . En abscisse est représenté l'indice de difficulté du pointage de Fitts (1954), avec $ID = \log_2(D/W)$. Pour un *ID* modéré de 5 bits, l'échelle a déjà chuté jusqu'à ne plus représenter que 2,5% du maximum. Pour un *ID* de 10 bits, correspondant à la région critique du continuum où la navigation multi-échelle commence à devenir vraiment nécessaire (Guiard et al., 2001, 2004), l'échelle ne vaut plus que 0.2% de son maximum. Un *ID* de 10 bits définit par exemple la difficulté du pointage pour un objet graphique mesurant à l'écran, en vision perpendiculaire, 10 pixels de côté et situé à peu près à 10.000 pixels de distance, soit à 10 hauteurs d'écran (ou 5 pixels de côté à 5 hauteurs d'écran, etc.). Il est clair que dans ce cas le pointage direct sur l'objet vu en

perspective serait extrêmement difficile, la cible du curseur étant réduite dans la vue oblique à un unique pixel. Au delà de 10 bits, l'objet ne peut plus être représenté autrement que sous forme d'un marqueur symbolique.

Mais il n'y a aucune raison d'imposer le pointage direct dans les vues en perspective. En autorisant la navigation par défilement en vision perspective, par exemple en permettant à l'utilisateur de draguer le document vers lui — ou de rapproche la caméra de la cible, ou de s'en rapprocher lui-même, ce qui revient au même —, on le place dans des conditions analogues à celles du pilote dont le regard, comme il convient, est orienté vers l'avant (Figure 9). Notons toutefois que le caractère hautement non-linéaire de la projection perspective complique le contrôle du défilement. Dans ce type de vue, en effet, le dragage du document va produire un déplacement d'autant plus grand que le *point de saisie* aura été plus lointain dans la sélection. Pour un dragage d'amplitude donnée sur l'écran, la variation du point de saisie peut entraîner des changements considérables de l'amplitude du déplacement par rapport au document, jusqu'à des changements d'*ordre de grandeur*. Ainsi par exemple, pour une amplitude fixe de dragage de 10 pixels-écran, si on s'avise de déplacer le point de saisie de 100 pixels-écran, l'effet du même mouvement de dragage sur la translation de la caméra par rapport au document est susceptible de varier dans un rapport de 1 à 1000 (voir Annexe 2).

5. Expérimentation

Cette expérience a mis à profit un prototype d'interface permettant à l'utilisateur de naviguer dans un vaste document en disposant soit de la vue zoomable classique soit d'une vue en perspective, la tâche étant d'atteindre une cible située à des distances variables du point d'observation offert au départ.

5.1. Méthodes

Choix de la tâche. Nous avons eu recours dans cette étude à la tâche de pointage de Fitts (1954). En effet, l'acquisition d'une cible, qui présente l'avantage de constituer un paradigme expérimental simple et bien contrôlé, nous semble capturer un aspect central de tout acte de navigation, le fait qu'il soit orienté vers une cible et guidé visuellement. La consigne de la tâche de pointage est bien définie, puisqu'il s'agit de s'efforcer d'atteindre la cible désignée en un temps T minimal. Or on connaît une loi quantitative, de validité très générale, permettant de modéliser par une fonction affine la relation entre cette mesure de temps et le cahier des charges du pointage, qui se ramène au rapport entre la distance D à couvrir et la largeur W de la cible. Spécifiquement, ce qu'on appelle la *loi de Fitts* (Fitts, 1954; Fitts & Peterson, 1964) stipule que


$$T = k_1 + k_2 \cdot \log_2(D/W + 1) \quad (1)$$

où T désigne le temps moyen requis pour atteindre la cible, où k_1 et k_2 sont des coefficients ajustables ($k_2 > 0$). On désigne conventionnellement l'expression $\log_2(D/W + 1)$ comme l'*indice de difficulté (ID)* du pointage. Dans une expérience manipulant l'*ID* comme variable indépendante et mesurant le *TM*, la loi de Fitts va nous permettre de caractériser les performances d'un système humain-ordinateur donné par seulement deux nombres, les coefficients k_1 et k_2 .

Il est à noter que la logique du paradigme de pointage de Fitts n'est en aucun cas altérée quand l'échelle intervient comme une variable supplémentaire, comme on peut le faire dans les monde électroniques. La manipulation du facteur d'échelle, quelle que soit la technique mise en œuvre, est un procédé mis au service du participant pour lui permettre de

progresser vers sa cible, un procédé qui n'affecte en rien la définition et la mesure des variables indépendantes du paradigme, D et W . Nous savons que la loi de Fitts conserve de fait son entière validité dans la tâche de pointage assistée par un zoom (Guiard et al., 2001).

Matériel. L'expérience, programmée en C++, a été conduite sur un ordinateur type PC sous Linux, équipé d'un écran $XX \times XX$ pixels. Dans un souci de simplification, nous avons utilisé une souris standard à roulette comme unique organe d'entrée. La roulette de souris, inutilisée dans le cas du pointage facile uni-échelle, avait servi une fonction différente dans les deux conditions multi-échelle : en condition pan-zoom elle permettait de faire varier l'altitude de la caméra, celle-ci restant toujours perpendiculaire au document ; en condition perspective elle permettait de faire varier l'inclinaison de la caméra vers le haut ou vers le bas du document (rotation de type panoramique, sans translation). Dans toutes les cas, le déplacement du corps de la souris déplaçait un curseur d'écran, et le document pouvait être déplacé au moyen du dragage classique avec le bouton gauche de la souris maintenu enfoncé (aucun dragage n'était toutefois nécessaire dans la condition de pointage facile uni-échelle, où la cible était toujours visible dans la vue).

Choix du document électronique à parcourir. Nous avons utilisé pour l'expérience un document calculé conçu sur mesure pour une tâche de pointage. Il s'agissait d'une surface rectangulaire de grande dimension (Y pixels verticalement sur X pixels horizontalement) sur lequel figuraient deux rectangles colorés, situés l'un vers le haut et l'autre vers le bas du document. Ces rectangles constituaient les cibles que le participant devait, en naviguant, atteindre alternativement et cliquer. La cible à atteindre était entourée par un champ de cercles concentriques couvrant la totalité du document et suffisamment dense pour que toute vue intercepte au moins quelques arcs. Ce système de repérage supprimait pour le participant toute incertitude, quand la cible était hors de vue, sur la direction à suivre pour l'atteindre ainsi que s' distance restant à couvrir.

Quand la cible ne pouvait être affichée en raison d'une échelle de visualisation insuffisante, son emplacement sur le document était signalé par un marqueur symbolique (un rectangle bleu ?? de taille constante), selon le principe du zoom sémantique (REF).

Conditions expérimentales. Le but principal de l'expérience était d'évaluer les performances de pointage avec les deux techniques concurrentes de navigation implémentées dans notre prototype d'interface, le zoom perpendiculaire de la technologie courante et la vue en perspective, comme schématisé dans la Figure 13.

Nous avons confronté les sujets à sept niveaux de difficulté, $ID = 3 / 5,5 / 8 / 9 / 11 / 13 / 15$ bits. Tandis que les trois premiers ID nous ont permis de tester nos participants dans le cas du pointage traditionnel (direct, sans navigation), les quatre niveaux de difficulté suivants, impraticables sur un écran d'ordinateur en l'absence de quelque procédé de manipulation de l'échelle, ont été l'occasion de mettre en œuvre les deux techniques multi-échelle que nous souhaitions comparer.

Cette première expérience sur la navigation en vue perspective étant exploratoire, nous avons développé une interface de conception minimaliste, délibérément dépourvue de tout artifice logiciel de facilitation, dans le but de produire, outre les mesures de temps nécessaires pour le calcul des performances de pointage, des enregistrements continus de nature à révéler les stratégies de navigation spontanément utilisées par les participants, sans interférence du système.

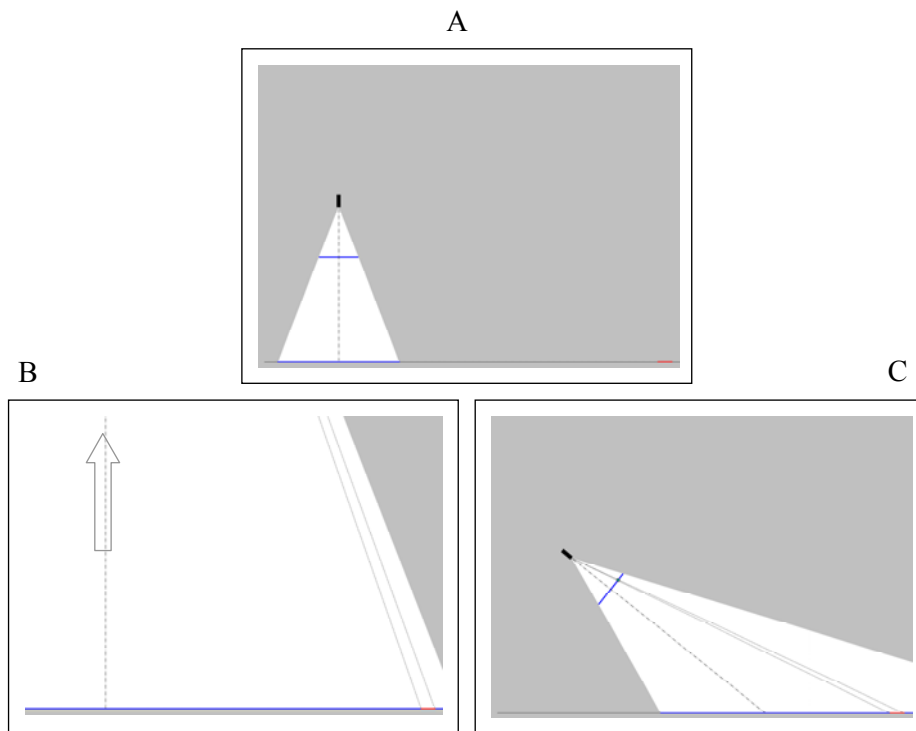


Figure 13. Les deux techniques alternatives de variation de l'échelle de visualisation implémentées dans le prototype d'interface expérimental. Dans le schéma A, la situation de départ : la cible (trait rouge) se trouve loin de la sélection (trait bleu). En B, un zoom arrière étend la sélection dans toutes les directions jusqu'à inclure la cible. En C, une inclinaison de caméra déplace et étend la sélection vers la cible, jusqu'à l'inclure.

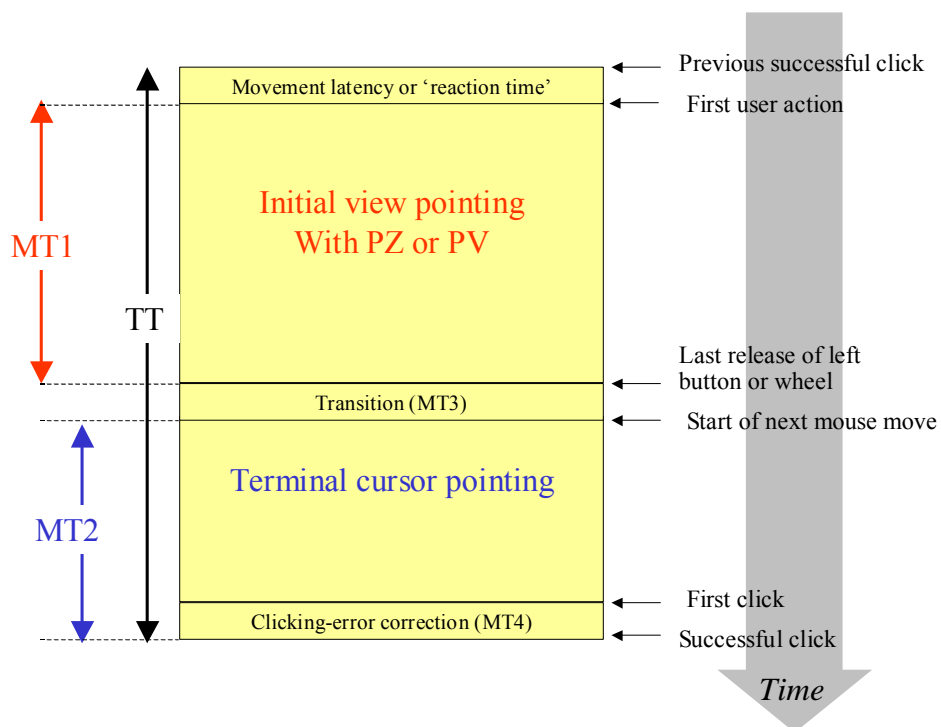


Figure 14. Définition des variables dépendantes temporelles de l'expérience.

Décomposition du temps. Comme illustré dans la Figure 14, ce que le paradigme de pointage désigne comme le temps de mouvement, ici le temps total TT écoulé entre deux clics réussis, se dichotomise dans l'expérience en un premier temps TM_1 consacré à la navigation de vue et un second temps TM_2 , typiquement beaucoup plus court, consacré au pointage terminal avec le curseur. Le premier temps correspond aux mouvements de caméra par rapport au document requis pour visualiser la cible, le second au mouvement plus simple du curseur dans l'espace de la vue pour acquérir la cible. Nous mettrons l'accent d'une part sur la variable TT , qui nous informe globalement sur la performance, et d'autre part sur la variable TM_1 , qui intéresse spécifiquement l'objet central de cette étude, la navigation de vue (pan-zoom classique vs. vue inclinable).

Procédure. Seize volontaires adultes non rémunérés ont participé à l'expérience, en deux sessions d'environ une heure consacré chacune à l'une des deux techniques de navigation multi-échelle, la moitié des sujets commençant par le pan-zoom, l'autre moitié par la vue perspective. Chaque session commençait avec une tâche facile mettant en jeu le pointage uni-échelle classique, sans navigation de vue : avec un ID de 3, 5.5 ou 8 bits, en effet, il est possible de visualiser dans une même vue en plein écran les deux cibles alternatives su pointage. Dans la seconde partie de la session, la gamme d' ID était déplacée vers le haut (9, 11, 13 et 15 bits), imposant l'obligation de faire précéder le pointage de curseur par une navigation multi-échelle de la vue.

Nous avons choisis, pour des raisons de commodité, le protocole du pointage alternatif, dans lequel deux cibles s'offrent au pointage, impliquant un va-et-vient. Le programme exigeait que le participant clique effectivement une cible pour que l'autre cible lui soit proposée, toute imprécision de pointage devant être corrigée immédiatement. Cette disposition simplifie l'analyse puisqu'elle garantit que la mesure du temps d'acquisition de cible soit associée à un pourcentage d'erreurs nul.

5.2. Résultats et discussion

La Figure 15 montre de manière comparative comment le temps total de pointage, mesuré de clic à clic, varie avec la difficulté du pointage dans nos deux conditions de navigation multi-échelle. Aucune des deux courbes ne s'inscrit dans le prolongement de la courbe obtenue dans le pointage facile uni-échelle : ce décalage s'explique aisément par le fait qu'une unique souris est inadaptée **temps de transition à vérifier**.

A l'évidence, à en juger par la proximité des deux courbes, la navigation en vue perspective donne lieu à des performances tout à fait satisfaisantes, du même ordre que celles que permet la technique standard.

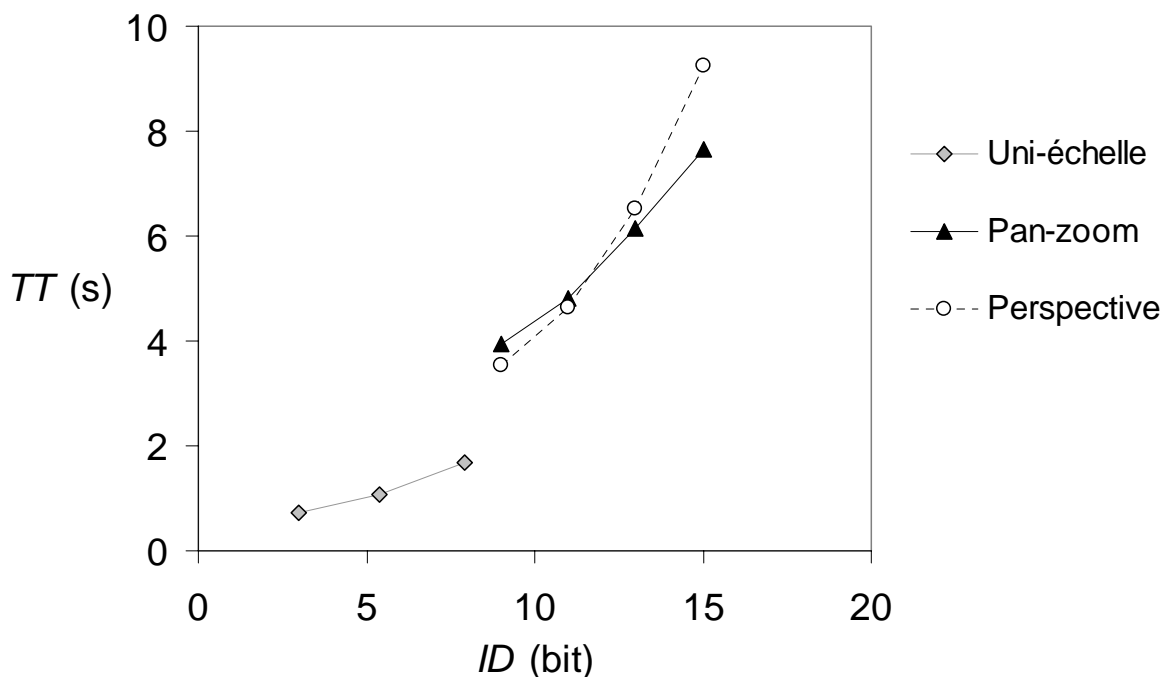


Figure 15. Variation du temps global de pointage avec le niveau d' ID , d'une part pour le pointage multi-échelle réalisé au moyen des deux techniques concurrentes, et d'autre part pour le pointage uni-échelle.

En condition pan-zoom, le TT se laisse décrire conformément à notre prédiction (Guiard & Beaudouin-Lafon, 2004 ; Guiard et al., 2001) comme une fonction affine de l' ID , les 16 estimations individuelles du r^2 de la régression linéaire étant assez bien regroupés (entre .915 et .99995), reproduisant la loi de Fitts. Avec la navigation en perspective, la croissance du TT avec l' ID devient plus raide, et l'ordonnée à l'origine intervient plus bas, deux effets tout à fait significatifs observés chacun 16 fois sur 16 ($p < .01$ par le test du signe bilatéral). Le point d'intersection, assez variable d'un participant à l'autre (dans l'intervalle 9-23 bits) se situe en moyenne à 10.9 bits. En d'autres termes, on peut dire qu'en deçà de ce niveau d' ID , la nouvelle technique de navigation l'emporte sur la technique pan-zoom, et qu'au delà c'est l'inverse.

On observe pour la condition perspective une discrète concavité en U de la fonction liant le TT à l' ID . Digne de considération puisqu'elle s'est avérée systématiquement présente chez tous les participants ($p < .01$), cette concavité est similaire à celle d'une fonction exponentielle. Si l'on extrapole vers des niveaux de difficulté nettement supérieurs à ceux que nous avons utilisé, il est probable que le temps d'acquisition de cible en navigation perspective va rapidement atteindre des valeurs problématiques, alors que la technique pan-zoom, nous le savons, permet à ce temps de continuer de croître linéairement.

Ainsi, une première conclusion de ce travail est que la technique de navigation en perspective testée dans cette expérience est tout à fait efficace pour des niveaux de difficulté modérés, dans la gamme 9-11 bits, mais elle n'est praticable que jusqu'à un certain niveau de difficulté — fait compréhensible si l'on se rappelle combien est sensible, en vision perspective, le choix du point de saisie du document (Section 4.5).

6. Conclusion et perspectives

L'une des options qui caractérisent notre approche dans cette étude est de tenter de comprendre les contraintes qui pèsent sur l'humain dans son interaction avec ses objets documentaires en considérant parallèlement le monde physique réel et les mondes électroniques. Il est à peu près certain que la forme, remarquablement stable, vers lesquelles ont convergé en pratique la plupart de nos documents matériels, celle du livre, ou *codex* (qui est avec nous depuis seize siècles), représente une solution optimale. La question de savoir quelle est la nature du cahier des charges auquel le livre apporte une réponse optimale reste en revanche une question ouverte pour la recherche (REF demander à Fekete) — c'est un cas de figure de la forme « trouver le problème de cette solution ».

La stratégie qui semble avoir été poursuivie dans l'histoire de l'artisanat du livre consiste à augmenter la quantité d'information de la scène optique pour maximiser l'information accessible en vision stationnaire, c'est à dire en position assise. On s'est employé d'une part à améliorer la qualité de l'impression, de manière à accroître la densité d'information (e.g., le nombre de caractères d'imprimerie lisibles par unité de surface). Une fois atteint la limite supérieure de densité (pour des raisons physiologiques), il reste d'autre part la ressource d'accroître la surface de l'objet, mais on atteint bientôt une limite, celle-là de nature optique (liée aux propriétés de la visualisation oblique). On s'arrête finalement au format A4, qui doit correspondre à peu près à ce que nous avons appelé le maximum d'information visualisable par scène optique (MIVSO, voir la Section 3.4). Si l'on veut un objet qui contienne davantage d'information que le MIVSO, on aura recours à l'empilement des feuilles (le livre) ou à leur pliage (la carte routière ou le journal quotidien).

Aussi éprouvées soit-elles, ces techniques ne sont qu'un compromis partiellement satisfaisant. Certes, elles donnent naissance à un objet dont le contenu informationnel peut dépasser largement le MIVSO, mais elles imposent un parcours strictement *séquentiel* du document : la forme du livre ne donne jamais à voir qu'un sous-ensemble réduit du document — on ne peut voir qu'une page à la fois. Or la page, même si elle nous est extraordinairement familière (au point d'être toujours scrupuleusement reproduite dans nos éditeurs de textes), reste un principe entièrement arbitraire de segmentation du document.

D'où l'intérêt de la stratégie alternative suivante. Abandonnons la contrainte de la vue stationnaire et demandons à l'utilisateur du document de bien vouloir se déplacer : c'est le cas de figure de l'*affiche*. Ici, le contenu du document peut être déployé librement sur un plan de dimensions arbitrairement grandes, sans éprouver la contrainte du MIVSO et celle, associée, de la segmentation en pages. Mais le document sera désormais montré verticalement, et il devra être fixé sur d'un *panneau* de grande taille et fixe dans l'environnement. L'exploration du document va donc, dans cette seconde option stratégique, revêtir la forme d'une locomotion.

Deux types de contraintes semblent fortement contribuer à la marginalisation du panneau, comparé au *codex*, en tant que technique d'exposition de l'information en direction de l'œil humain. On mentionnera en premier lieu, la contrainte évidente d'encombrement que l'on peut résumer à l'écrasante supériorité du livre en termes de rendement informationnel (évalué en bits par gramme de matière). Citons ensuite les contraintes propres à la locomotion humaine, qui se trouve être énergétiquement coûteuse et de performances médiocres: la

locomotion humaine non seulement lente, mais aussi terrestre — d'où la forme généralement oblongue des surfaces d'exposition que l'on crée en juxtaposant plusieurs panneaux dont la hauteur ne doit pas excéder 2 m.

Le fait que l'une et l'autre de ces deux familles de contraintes (encombrement, énergétique de la locomotion) disparaissent entièrement avec le document électronique nous amène à reconsidérer la question de la solution optimale dans le nouveau contexte. On peut douter, par exemple, du bien fondé du livre électronique conçu comme une réplique fidèle du livre matériel compulsable, comme le *eBook* de Card et al. (REF). Puisque l'encombrement n'est pas un souci dans les mondes électroniques et que, d'autre part, le déplacement virtuel de l'observateur peut parcourir, à des vitesses que l'on choisira librement, les trois dimensions de l'espace euclidien, la métaphore locomotrice de la navigation documentaire semble appeler un examen sérieux. Il nous paraît spécialement judicieux d'explorer soigneusement le nouvel espace de conception qui s'ouvre. Les techniques courantes de visualisation interactive nous semblent devoir être remplacées par des techniques mieux adaptées aux spécificités du navigateur humain.

7. Remerciements

Nicolas Roussel, Jean-Daniel Fekete.

8. Références

- Poupyrev, I., Weghorst, S., & Fels, S. (2000). Non-isomorphic 3D rotational techniques. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems, CHI 2000* (pp.). April 2000.
- Duhem, P. (1908/1990). *Sozein ta phainomena: Essai sur la Notion de Théorie physique de Platon à Galilée*. Vrin : Mathesis.

Annexe 1. Variation du rapport d'échelle de visualisation dans l'espace du document, en vue perspective

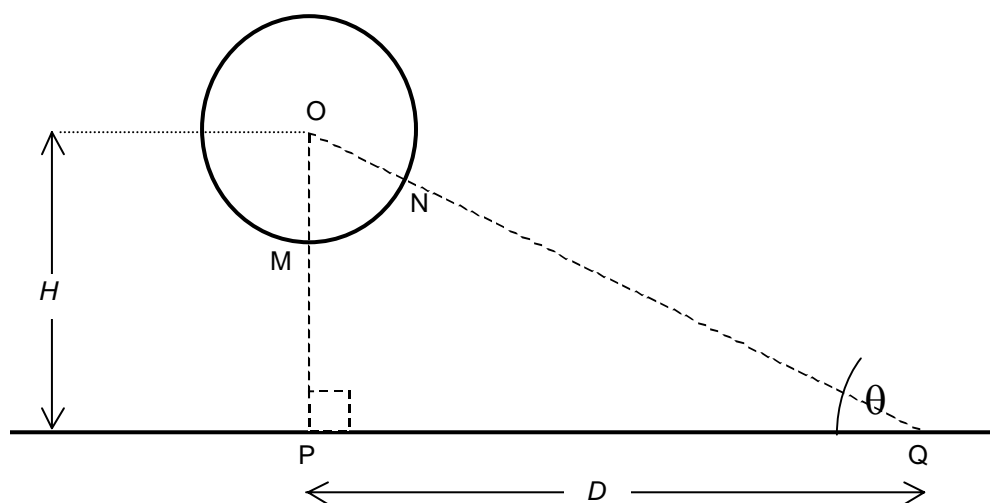


Figure A1.

Dans l'espace du document, soit P la position de la caméra, Q la position de la cible, et D la distance horizontale de la caméra à la cible. H est l'altitude de la caméra. Le rayon $OM = ON$ représente la focale f de la caméra. θ est l'angle de projection du point Q.

Le point d'observation O est le centre d'un cercle dont le rayon est la focale f de la caméra. Le plan de projection est une tangente à ce cercle. Pivoter la caméra autour de O (rotation panoramique) revient à choisir d'autres tangentes. Pour des valeurs élevées du rapport ON/OQ qui seules nous intéressent dans le contexte de la visualisation multi-échelle, on commet une erreur négligeable en identifiant par commodité tous les plans de projection au cercle lui-même.

Nous nous conformons ici à la convention des cartographes, pour qui l'échelle (*visualization-scale ratio*, *VSR*) est le rapport entre les tailles sur la carte et leurs correspondances dans le monde réel. Ainsi l'échelle sera définie comme le rapport entre la grandeur de l'objet à l'écran et la grandeur 'réelle' de cet objet sur le document.

Pour le point P situé à la verticale de la caméra et vu par conséquent en vision perpendiculaire, on a

$$VSR(P) = \frac{OM}{OP}, \text{ c'est à dire}$$

$$VSR(P) = \frac{f}{H}, \tag{A1}$$

ce qui représente l'échelle maximale de visualisation accessible dans un document, étant donné les paramètres f et H .

Pour tout point Q de position quelconque sur le document, on a

$$VSR(Q) = \frac{ON}{OQ} \sin \theta = \frac{f}{\sqrt{H^2 + D^2}} * \frac{H}{\sqrt{H^2 + D^2}}, \text{ c'est à dire}$$

$$VSR(Q) = \frac{fH}{H^2 + D^2}, \quad (A2)$$

ce qui donne l'échelle à laquelle est visualisé tout objet situé autour du point Q vu obliquement.

Quand $D \gg H$, l'Equation A2 se simplifie :

$$VSR(Q) \approx \frac{fH}{D^2}. \quad (A3)$$

Annexe 2. Illustration par un exemple numérique de la non-linéarité du contrôle dans le dragage d'un document vu en perspective

La caméra est placée à une altitude de 1 cm par rapport au plan du document (retenons le cm comme unité pour décrire le système caméra-document). Elle est inclinée à 60° dans le sens du tangage vers le haut, son angle de vue est de 60° . La taille de l'écran est de 600 pixels.

On considère un mouvement de dragage dont l'amplitude dans la vue est constante ; fixons cette amplitude à 10 pixels. Comment varie l'effet de ce mouvement donné de curseur quand varie l'emplacement du point de saisie dans le gradient d'échelle de la vue ?

Dans les conditions que nous venons de fixer, la limite distale de la sélection est à une distance infinie, ce qui signifie que l'horizon apparaît à la limite supérieure de la vue. Si l'on indexe les pixels d'écran de 0 à 599, le pixel #0 correspond à l'infini. Pour un mouvement de dragage dont la saisie intervient au pixel #101 et le relâchement au pixel #111, le document sera rapproché de 0,618 cm. Le même mouvement effectué plus haut dans la vue, du pixel #1 au pixel #11, rapprochera le document de 629.8 cm, soit un effet plus que mille fois supérieur. Cette non-linéarité sera bien sûr moins radicale si le point de saisie intervient dans des régions moins éloignées du document.