HAL
open science

Etude du contrôle du point de vue 3D à partir de la tête et des épaules

Thibaut Jacob, Gilles Bailly, Eric Lecolinet

To cite this version:

Thibaut Jacob, Gilles Bailly, Eric Lecolinet. Etude du contrôle du point de vue 3D à partir de la tête et des épaules. 27ème conférence francophone sur l'Interaction Homme-Machine., Oct 2015, Toulouse, France. pp.w4, 10.1145/2820619.2825005 . hal-01219069

HAL Id: hal-01219069

https://hal.science/hal-01219069

Submitted on 23 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude du contrôle du point de vue 3D à partir de la tête et des épaules

Thibaut Jacob
Télécom Paristech, CNRS LTCI
46, rue Barrault
75013 Paris
thibaut.jacob@telecom-paristech.fr

Gilles Bailly
Télécom Paristech, CNRS LTCI
46, rue Barrault
75013 Paris
gilles.bailly@telecom-paristech.fr

Eric Lecolinet
Télécom Paristech, CNRS LTCI
46, rue Barrault
75013 Paris
eric.lecolinet@telecom-paristech.fr

© ACM, 2015. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Actes de la 27ème conférence francophone sur l'Interaction Homme-Machine, 2015.
http://dx.doi.org/10.1145/2820619.2825005

Résumé
La conception de modèles en trois dimensions nécessite un changement fréquent du point de vue lors de la manipulation d'objets 3D. Afin de réduire le coût de l'alternance entre tâches primaires (création d'objets) et tâches secondaires (contrôle du point de vue), nous proposons l'utilisation des mouvements de la tête et des épaules pour le contrôle de caméra. Nous présentons les résultats d'une étude par élicitation visant à comprendre comment les utilisateurs assignent de manière spontanée des mouvements du torse (tête et épaules) à quatre types de contrôles de caméra dans un contexte bureautique.

Mots-clés
contrôle du point de vue 3D ; étude par élicitation ; mouvements de la tête et des épaules

ACM Classification Keywords
H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

Introduction
La manipulation d'objets 3D nécessite de fréquemment changer le point de vue afin d'éviter l'occultation et de pouvoir apprécier les détails, percevoir la profondeur ou avoir un aperçu global de la scène [2]. Plusieurs techniques d'interaction reposant sur l'utilisation du

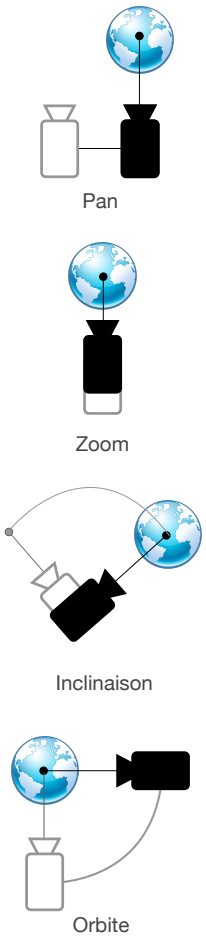

Figure 1 – Les quatre types de contrôles de caméra de cette étude

clavier ou de la souris ont été proposées à cette fin [2](chapitre 4.2). Cependant, ces techniques imposent à l'utilisateur d'alterner entre création ou édition d'objets et contrôle de la caméra. Ceci a pour inconvénient d'interrompre le déroulement du travail et n'est sans doute pas sans impact sur l'attention de l'utilisateur ainsi que le temps d'exécution de la tâche [2, 8].

Utiliser la tête pour contrôler la caméra présente plusieurs avantages. D'une part, ce canal interactionnel supplémentaire permet de réduire les interruptions et d'augmenter le parallélisme : la tête contrôlant la caméra, les mains sont libres pour effectuer des tâches primaires (en l'occurence, manipuler des objets 3D). D'autre part, le couplage tête/caméra peut apporter une perception plus écologique de la scène 3D [4]. Enfin, cette solution est bon marché et facile à mettre en place car ne nécessitant qu'une webcam.

Les mouvements de la tête ont été utilisés pour contrôler la caméra dans le domaine de la réalité virtuelle, via l'utilisation de visiocasques, d'environnements de type CAVE, etc [2]. Cependant, bien que la plupart des ordinateurs possèdent désormais une caméra intégrée, cette approche a été peu étudiée dans le contexte de la bureautique [5]. De fait, sa mise en application dans ce contexte nécessite de déterminer des couplages tête/caméra efficaces, confortables et adaptés au fait que la position du corps et des mains est partiellement contrainte dans ce type de situation.

Nous étudions comment assigner les mouvements de la tête et des épaules pour le contrôle du point de vue dans une scène 3D. Plus précisément, nous nous intéressons aux mouvements de *pan*, *zoom*, *inclinaison* et *orbite* (Figure 1) qui sont les quatre contrôles de caméras les plus utilisés dans les logiciels de 3D (e.g., Google Earth, Blender, Sketchup). Nous pensons que le mouvement des épaules est une modalité pertinente, car, contrairement à la tête, les épaules peuvent être déplacées sans influer sur la vision de l'écran par l'utilisateur. De plus, le mouvement de la tête est relativement indépendant du mouvement des épaules (les utilisateurs peuvent déplacer la tête sans déplacer les épaules et vice-versa). Prendre en compte cette modalité permettrait de définir des langages d'interactions plus expressifs.

Nous employons une méthodologie de "guessability" [14] afin de comprendre comment les utilisateurs contrôleraient la caméra de manière spontanée au moyen de mouvements de la tête et des épaules. Cette approche consiste à présenter les effets (contrôles de caméra) aux participants et éliciter les causes (mouvements de la tête et des épaules) ayant mené à ceux-ci. Nous observons que 1) l'utilisation des épaules est utile pour des mouvements d'orbite ; 2) au moins 10 contrôles de caméra (en considérant les mouvements gauche/droite/haut/bas) peuvent coexister dans une application de modélisation 3D classique ; 3) mis à part le cas de l'*inclinaison* verticale, les contrôles de caméra verticaux sont difficiles à opérer et génèrent peu de consensus.

Nous pensons ces observations utiles pour les concepteurs désirant créer de nouvelles techniques d'interaction basées sur les mouvements du torse, ainsi que dans le domaine de la vision par ordinateur. A cette fin, nous donnerons librement accès aux données expérimentales.

Travaux existants

Contrôle de caméra. Le contrôle du point de vue est un vaste sujet concernant divers domaines comme la réalité virtuelle ou les jeux vidéo. Dans ce projet, nous nous intéressons aux contrôles de caméra courants de type *pan*,

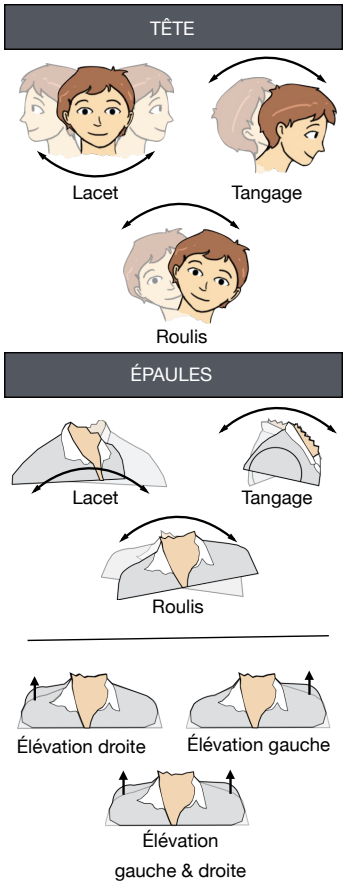

Figure 2 – Espace de conception des mouvements de tête et d'épaules

zoom, *inclinaison* et *orbite* (Figure 1) ainsi qu'au couplage tête/épaules. De nombreux travaux ont été menés sur le contrôle du point de vue au travers des mouvements du corps pour ces différents contrôles de caméra. On trouve par exemple [9] pour un contrôle de type *pan*, [7] pour le contrôle de l'*inclinaison*, [6] pour le contrôle de l'*orbite*, [5] pour contrôler le niveau de *zoom*, ou encore pour les mouvements de type fish tank [13].

Mouvement de la tête. Les mouvements de la tête ont été utilisés dans différents types d'applications comme la téléopération (e.g. [15]), les jeux (e.g. [10]) ou en réalité virtuelle (e.g. [13]). Dans [11], Terziman et al. ont étudiés les mouvements de la tête lors de la marche et mis au point un algorithme exploitant trois degrés de liberté (translation horizontale, verticale et roulis) permettant à l'utilisateur de simuler le déplacement dans une scène virtuelle tout en étant assis. D'autres travaux s'appuient sur la métaphore du Fish Tank [13] afin d'augmenter la perception de profondeur via la parallaxe de mouvement : la position de la caméra dans le monde virtuel est directement couplée à la position de la tête de l'utilisateur. Bien que ce couplage permette d'augmenter la sensation d'immersion dans les applications de réalité virtuelle et améliorer les performance de l'utilisateur dans les jeux vidéos, ce souci de réalisme n'est pas forcément utile pour toutes les applications. A contrario, l'efficacité et la qualité de l'ergonomie sont des critères dominants pour la plupart des applications [2] (chapitre 6.4) et en particulier celles qui relèvent de la bureautique.

Dans un contexte de bureautique, les mouvements de la tête ont été utilisés pour modifier l'échelle de documents (e.g. pages web) dans des interfaces 2D où ils permettent de contrôler plusieurs niveaux discrets de zoom [5]. Les mouvements de la tête ont également été utilisés pour contrôler un curseur afin d'aider les personnes en situation de handicap (e.g. [12]) ou pour contrôler la navigation dans un document en deux dimensions (défilement)[1].

Certaines de ces implementations couplent la position de la tête avec la position de la taille ou celle des pieds, cependant, à notre connaissance, aucun travail n'a été effectué sur le couplage tête/épaules pour le contrôle du point de vue dans une scène 3D.

Espace de conception. Bien que des études aient été conduites en psychologie expérimentale sur l'utilisation des mouvements de la tête (e.g.,[3]), celles-ci n'ont pas pris en compte les spécificités du contexte (bureautique) ni la tâche (contrôle de caméra) qui nous intéressent dans cette étude. À partir d'une session de brainstorming nous avons défini un espace de conception pour les mouvements de la tête et des épaules dans un contexte de bureautique (Figure 2). Bien que d'autres mouvements d'épaules soient théoriquement possibles, comme la rétractation ou la prolongation, ceux-ci sont difficiles à accomplir avec les mains sur le clavier ou la souris.

Les travaux sur les mouvements de la tête ont principalement cherché à augmenter l'immersion et à identifier de nouveaux usages. Notre objectif est d'explorer plus systématiquement l'espace de conception du couplage tête/épaules : Nous observons comment les utilisateurs réalisent spontanément des mouvements de tête et des épaules pour contrôler une caméra dans une application 3D. Nous pensons que ces résultats peuvent constituer une base utile pour faciliter la création de nouvelles techniques d'interaction.

Élicitation

Le but de notre étude était de : (1) déterminer comment les utilisateurs utilisent les mouvements de la tête et des

$$A_c = \sum_{P_g \subseteq P_c} \left(\frac{\mid P_g \mid}{\mid P_c \mid} \right)^2 \quad (1)$$

Equation 1 - Formule de calcul du consensus [Wobbrock, 2009]

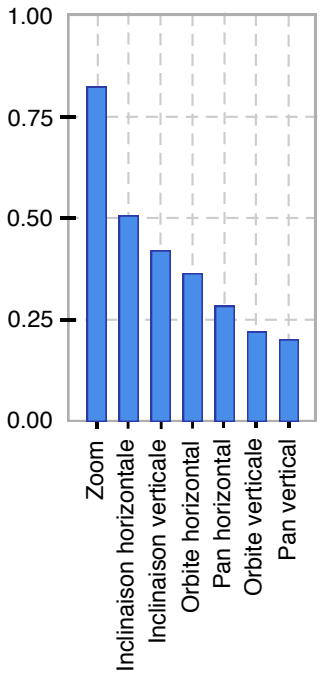

Figure 3 – Consensus pour chaque contrôle de caméra trié par ordre décroissant. Les gestes symétriques (haut/bas, gauche/droite, avant/arrière) ont été rassemblés.

épaules pour contrôler la caméra de point de vue dans les applications 3D classiques ; (2) comprendre les modèles mentaux associés à ces contrôles ; (3) évaluer le confort de l'utilisateur qui interagit avec les mouvements du corps ; (4) guider les concepteurs sur les façons d'assigner des mouvements de la tête et des épaules aux contrôles de caméra. Pour ce faire, nous avons adopté une approche de type "user-defined interaction", basée sur la présentation des effets d'une action aux participants et l'élicitation par la suite des causes ayant mené à ceux-ci [14]. Les utilisateurs visionnaient une vidéo montrant un contrôle de caméra et devaient ensuite proposer les mouvements de la tête et des épaules ayant conduit à ce mouvement.

Nous avons suivi une méthodologie similaire à celle utilisée par Wobbrock et al. [14], mais avec pour différence notable que nous nous intéressons à des actions de contrôle continu plutôt qu'à des commandes discrètes et nous observons si les utilisateurs sont capables d'effectuer ces actions de manière confortable et précise. Des mouvements larges peuvent être inconfortables et inappropriés pour des activités quotidiennes sur ordinateur, tandis que des mouvements trop petits peuvent manquer d'amplitude pour faire du contrôle précis. Afin de répondre à ces questions, nous avons demandé à nos participants de faire des mouvements aussi larges que possible mais qui restent confortables, ceci afin d'estimer quelle amplitude maximale peut être utilisée en pratique. Nous leur avons également demandé de réaliser ces mouvements aussi rapidement et précisément que possible.

Nous distinguons d'un côté le confort physique et de l'autre le confort visuel. Le confort physique est lié à la fatigue musculaire lors des mouvements. Le confort visuel est lié à la facilité avec laquelle l'utilisateur peut suivre l'écran pendant l'interaction. En effet, pendant que les utilisateurs effectuent des tâches primaires (par exemple entrer du texte ou manipuler un objet), certains mouvements de la tête peuvent rendre la vision difficile ou même ne permettre qu'une vision périphérique dégradée.

Participants et appareillage. Dix participants (6 hommes, 4 femmes) âgés de 24 à 33 ans (moyenne = 26.4 ; σ = 2.8) ont été recrutés dans notre laboratoire et ont reçu des friandises pour leur participation. Aucun d'eux n'avait d'expérience antérieure en ce qui concerne le contrôle de caméra par les mouvements de la tête.

Les participants étaient assis sur une chaise bureautique confortable. En accord avec les recommandations ergonomiques en ce domaine (ISO 9241), nous avons utilisé un écran de 15 pouces placé à une distance de 1 mètre des participants. Un dispositif ART a été employé pour la capture de mouvement afin d'acquérir des données précises. Trois cibles comprenant plusieurs marqueurs passifs (positionnés sur la tête et les deux épaules des participants) ont été utilisées.

Procédure. Les participants devaient penser à voix haute pendant l'expérience. Chaque contrôle de caméra, réparti par carré latin, devait être répété 5 fois pour chaque variation (gauche/droite/haut/bas/ ou avant/arrière). Les participants étaient autorisés à prendre des pauses entre les essais.

Résultats

Suite aux propositions des participants pour chaque contrôle de caméra, nous avons groupé les propositions similaires. La taille de ces groupes a ensuite été utilisée pour calculer le score de consensus chez les participants, selon la formule décrite dans [14] (Figure 4). Le niveau de consensus est représenté à la Figure 3. Les mouvements opposés sur un même axe (gauche/droite, haut/bas) ont

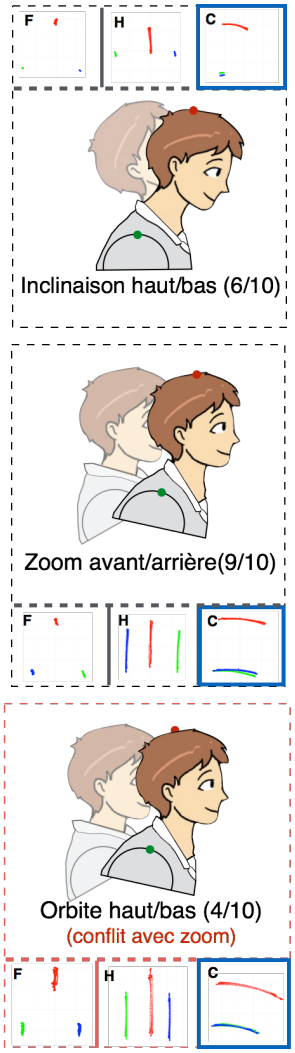

Figure 4 – Propositions pour l'inclinaison/orbite vertical et le zoom de caméra vue de face (F), de haut (H) et de coté (C)

été rassemblés, car les participants ont proposé des mouvements symétriques cohérents.

Zoom avant/arrière Nous observons que le *zoom* (9/10) a obtenu le plus haut niveau de consensus, les utilisateurs se penchant en avant ou en arrière (tangage tête + épaules) pour respectivement, agrandir ou réduire l'échelle.

Contrôle de caméra horizontal. L'*inclinaison* horizontale de caméra a également obtenu un niveau de consensus important : tous les participants ont effectué un mouvement de tangage avec la tête, la moitié d'entre eux réalisant en plus une torsion des épaules sur l'axe vertical (lacet). On en déduit que la reconnaissance de ce mouvement pourrait être basée sur le tangage de la tête seule. Pour le mouvement d'*orbite* horizontal, le consensus est d'utiliser seulement le mouvement de torsion des épaules (lacet). Le modèle mental diffère donc des autres couplages, la caméra n'étant *pas* couplée avec le mouvement de la tête mais avec celui des épaules. Enfin, le *pan* horizontal est basé principalement sur la conjonction du roulis de la tête et des épaules (5/10).

Contrôle de caméra vertical. Mis à part l'*inclinaison* haut/bas (6/10), les autres contrôles de caméra verticaux ont un niveau de consensus faible. Les interviews des participants offrent plusieurs réponses à ces observations. Tout d'abord, ceux-ci mentionnent que le *pan* ainsi que l'*orbite* verticaux étaient "fatiguant car ceux-ci nécessitaient de compenser des mouvements de la tête et des épaules de faible amplitude par des mouvements du torse (redressement)". D'autre part, "ces mouvements nécessitaient de se soulever de la chaise, ce qui est inconfortable". Enfin, il semble plus difficile d'effectuer un mouvement d'amplitude similaire sur l'axe haut/bas que sur l'axe gauche/droite, l'utilisateur étant assis.

Ces gestes sont représentés sur les figures 4 et 5.

Discussion et travaux futurs

Ensemble de gestes. Cette étude montre que certains contrôles de caméra (pan haut/bas, orbite haut/bas) sont difficiles à effectuer, ont un faible niveau de consensus entre utilisateurs et/ou entrent en conflit avec d'autres contrôles de caméra. Cependant, 10 contrôles de caméra (zoom avant/arrière, inclinaison haut/bas/gauche/droite, orbite gauche/droite, pan gauche/droite) peuvent être facilement effectués par les participants avec un niveau de consensus relativement élevé. De plus, ces mouvements peuvent coexister au sein d'une même application.

Le mouvement des épaules : un canal d'entrée prometteur. Nos résultats suggèrent d'exploiter le mouvement des épaules est important pour le contrôle orbite droite/gauche car celui-ci permet aux utilisateurs d'interagir sans bouger la tête. Une conséquence pratique de cette observation est qu'il serait utile de développer des algorithmes de vision par ordinateur à la fois capables de détecter le mouvement de la tête et des épaules. Ceci nécessite de plus d'orienter les caméras de manière appropriée, ce qui n'est toujours le cas actuellement.

Mouvements du corps non utilisés. Certains mouvements d'épaules présentés en Figure 2 n'ont pas été utilisés par les participants pour contrôler la caméra. Ces mouvements sont autant de gestes disponibles pour d'autres interactions.

Gestes de la main. Bien que ce papier traite principalement de situations dans lesquelles les utilisateurs effectuent des tâches primaires via l'utilisation du clavier ou de la souris, des dispositifs d'entrée tels que le Leap motion ou le Kinect permettent de détecter les gestes de la main. Des travaux supplémentaires sont

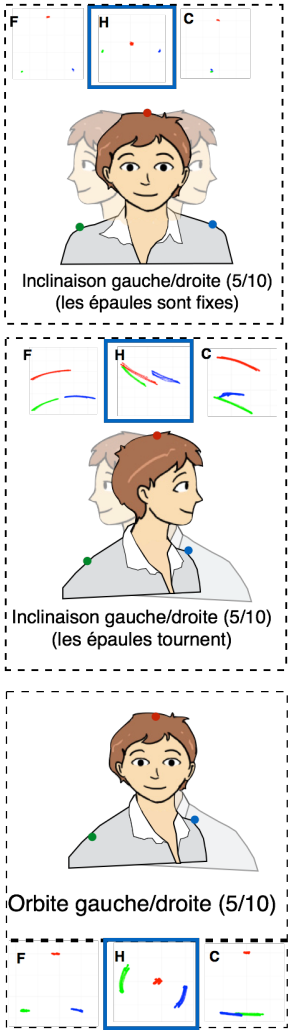

Figure 5 – Propositions pour l'inclinaison/orbite horizontale de caméra vue de face (F), de haut (H) et de coté (C)

vraisemblablement nécessaires pour adapter les algorithmes de vision par ordinateur afin d'assurer la compatibilité entre ces modalités d'interaction et l'utilisation de la tête et des épaules.

Notre objectif était de valider une approche rendant de nouvelles techniques d'interaction possibles afin d'augmenter la bande passante interactionelle. Nous avons étudié comment les mouvements du torse peuvent être utilisés de manière confortable pour contrôler différents types de contrôles de caméra. Les résultats de cette étude peuvent également aider les chercheurs en vision par ordinateur à développer des algorithmes plus robustes prenant en compte le mouvement des épaules et compatibles avec les webcams actuelles. A cette fin, les données acquises sont disponibles en ligne.

De nombreux aspects restent à étudier tels que l'impact du gain entre espaces moteur et visuel sur les performances et le confort, ainsi que le choix du délimiteur pour l'activation du contrôle du point de vue.

Remerciements

Le travail présenté dans cet article a été financé par l'agence nationale de la recherche (ANR 13 CORD 008) et réalisé au sein du consortium EDISON 3D.

Références

[1] Bérard, F. The Perceptual Window : Head Motion as a new Input Stream. INTERACT'99 (1999).

[2] Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. *3D User Interfaces : Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., 2004.

[3] Descarreaux, M., Pssmore, S., and Cantin, V. Head movement kinematics during rapid aiming task performance in healthy and neck-pain participants : the importance of optimal task difficulty. *Manual Therapy 15* (2010), 445–450.

[4] Gibson, E. J., Gibson, J. J., Smith, O. W., and Flock, H. Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology* (1959), 40–51.

[5] Harrison, C., and Dey, A. K. Lean and zoom : Proximity-aware user interface and content magnification. Dans *ACM CHI'08* (2008), 507–510.

[6] Koller, D. R., Mine, M. R., and Hudson, S. E. Head-tracked orbital viewing : an interaction technique for immersive virtual environments. Dans *UIST '96*, ACM (Nov. 1996), 81–82.

[7] LaViola, J. J., Jr, Feliz, D. A., Keefe, D. F., and Zeleznik, R. C. Hands-free multi-scale navigation in virtual environments. Dans *I3D '01*, ACM (Mar. 2001).

[8] Ortega, M., and Vincent, T. Direct drawing on 3d shapes with automated camera control. Dans *ACM CHI'14* (2014), 2047–2050.

[9] Song, D., and Norman, M. Nonlinear interactive motion control techniques for virtual space navigation. Dans *VRAIS '93*, CS (Sep. 1993), 111–117.

[10] Teather, R. J., and Stuerzlinger, W. Exaggerated head motions for game viewpoint control. Dans *ACM Future Play'08* (2008), 240–243.

[11] Terziman, L., Marchal, M., Emily, M., Multon, F., Arnaldi, B., and Lécuyer, A. Shake-your-head : Revisiting walking-in-place for desktop virtual reality. VRST '10, ACM (2010), 27–34.

[12] Varona, J., Manresa-Yee, C., and Perales, F. J. Hands-free vision-based interface for computer accessibility. *J. Netw. Comput. Appl. 31*, 4 (Nov. 2008), 357–374.

[13] Ware, C., Arthur, K., and Booth, K. S. Fish tank virtual reality. Dans *INTERACT '93 and ACM CHI '93* (1993), 37–42.

[14] Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. Dans *ACM CHI '09* (2009), 1083–1092.

[15] Zhu, D., Gedeon, T., and Taylor, K. Exploring camera viewpoint control models for a multi-tasking setting in teleoperation. Dans *ACM CHI'11*, ACM (2011), 53–62.