

Interface 3D appliquée au travail coopératif

Cédric Dumas
LIFL
Télévirtualité - INA
E-mail : dumas@{lifl.fr|ina.fr}

Gregory Saugis
LIFL
E-mail : saugis@lifl.fr

Christophe Chaillou
LIFL
E-mail : chaillou@lifl.fr

LIFL - Laboratoire d'Informatique
Fondamentale de Lille
bâtiment M3, Cité Scientifique
F-59 655 Villeneuve d'Ascq cedex

INA - Institut National de l'Audiovisuel
4, avenue de l'Europe
F-94 366- Bry s/ Marne cedex

1. résumé

Nous présentons un nouveau modèle d'interface graphique en trois dimensions. Notre champ d'application est le travail coopératif synchrone, domaine où les participants sont amenés à communiquer de manière simultanée tout en partageant des documents (réunion, enseignement, etc.).

Nous nous plaçons dans un cadre 'bureautique' avec une interface qui recrée virtuellement les conditions d'une réunion, en augmentant (par rapport au monde réel) les possibilités des utilisateurs. A cette fin, nous utilisons le plus souvent possible l'animation et l'interaction 3D temps réel pour renforcer la sensation de collaboration au sein de l'espace de travail virtuel. Ce dernier est créé à partir d'une géométrie artificielle (par opposition à une vraie géométrie 3D) utilisant une distorsion spatiale qui permet de garder toutes les informations visibles, centrant le point d'intérêt de l'utilisateur. Ce dernier peut collaborer avec ses interlocuteurs à travers leurs représentations virtuelles et un système de manipulation adapté.

Mots clés : TCAO synchrone, interface 3D, interaction 3D.

2. Introduction

Les progrès technologiques nous permettent d'accéder à de nouveaux champs, laissés jusqu'alors à notre seule imagination. Ainsi, l'évolution des réseaux et des ordinateurs profite au développement du travail coopératif assisté par ordinateur, domaine qui pour être valorisé nécessite de résoudre beaucoup de contraintes techniques et humaines. En particulier, nous devons tenir compte des contraintes du travail coopératif et des capacités de perception en trois dimensions de l'utilisateur.

Les limites des interfaces 2D

Les interfaces de bureau actuelles, de type WIMP¹ montrent vite leurs limites d'un point de vue ergonomique. L'espace 2D est très contraignant pour l'affichage de grosses quantités d'information : recouvrement des fenêtres, nécessité de réduire l'information à l'état d'icône. Quant au système d'affichage des fenêtres, que ce soit *X11* ou *Windows*, il banalise les applications dans des fenêtres toutes identiques, sans rapport avec la tâche de l'utilisateur.

D'un point de vue technique, les capacités des réseaux imposaient jusqu'ici des sessions asynchrones. Quant à la puissance des machines, elle limitait les interfaces à des représentations 2D de l'espace de travail : mode de déplacement limité, faible degré de réalisme des métaphores, problème de matérialisation des interlocuteurs et de leur relations dans l'interface, niveau d'interaction graphique très faible entre les utilisateurs, problème du rapport de l'utilisateur dans l'espace de travail (système peu proprioceptif).

¹ Window Interface Mouse Pointer

Les nouveaux enjeux des interfaces

Nous proposons un nouveau concept d'interface, fondé sur la synthèse d'images en temps réel, techniquement possible grâce à la vulgarisation des cartes graphiques 3D pour les ordinateurs personnels. En introduisant une nouvelle dimension (la profondeur), on modifie la place de l'utilisateur dans l'interface, on lui donne de nouveaux moyens de navigation, d'action et d'organisation de son espace de travail.

Nous allons exposer maintenant les différents concepts inhérents au travail simultané de groupe (TCAO² synchrone), à la représentation et à l'interaction dans une interface 3D. Puis, nous décrivons notre modèle d'interface et comment l'idée générale a pris forme. Pour conclure, nous décrivons les différents développements suivis et à suivre, concernant l'évolution du prototype et de son évaluation.

3. Concepts

Concevoir un modèle d'interface tridimensionnelle met en jeu plusieurs domaines. Nous avons déjà cité la synthèse d'images temps réel et le travail coopératif assisté par ordinateur, propres à la définition même du cadre des travaux. Mais certaines disciplines des sciences humaines interviennent directement dans l'élaboration, comme l'ergonomie ou la sociologie, pour ce qui concerne le comportement de l'utilisateur dans l'interface ou parmi un groupe d'utilisateurs. La synthèse de ces domaines permet d'avancer des concepts généraux sur la réalisation d'une interface 3D pour les activités coopératives.

3.1 Le travail coopératif synchrone

L'interface doit supporter le travail coopératif synchrone, donc des applications où les participants doivent communiquer pour prendre des décisions, comparer leurs opinions ou trouver des solutions (téléconférence, enseignement). Le réalisme des situations est crucial. L'utilisateur doit sentir immédiatement qu'il est avec d'autres personnes. Des expériences comme celle des *Hydra Units* [Buxton 1992] ou de *Majic* [Okada et al 1994] nous permettent de tirer quelques points fondamentaux des réunions médiatisées :

le contact visuel : un interlocuteur perçoit qu'on le regarde, il peut échanger des regards.

la conscience du regard : on trouve le centre d'intérêt d'une personne en suivant son regard.

les expressions faciales : elles donnent des informations sur les impressions des interlocuteurs, leur acquiescement ou leur mécontentement, et plus encore.

les gestes : ils sont importants pour la désignation (manipulation, interface gestuelle, etc.), mais aussi comme moyen d'expression dans le dialogue.

3.2 L'activité de groupe

La parole est en effet loin d'être le seul vecteur de communication pendant un dialogue (voir [Cassel]). Les gestes (volontaires ou réflexes), le regard et l'expression du visage apportent tout autant d'informations et aident à communiquer. De plus, le travail à plusieurs exige de connaître le point de vue et les actions de ses interlocuteurs (voir [Shu] et [Kuzuoka]). Il faut donc définir des métaphores permettant aux différents participants d'une activité coopérative de comprendre les actions des autres et de pouvoir interagir. Tout d'abord, ces derniers peuvent être vus dans l'interface sous forme de vignette fixe, animée (vidéo), ou sous forme 3D. Chacune de ces représentations ayant un contenu sémantique croissant (le plus pauvre étant l'image fixe qui ne fait qu'identifier visuellement une personne). A ce propos, [Benford et al 95] définissent, quand il s'agit de communication, les différents critères à envisager pour représenter une personne dans un environnement virtuel. Ils établissent des règles pour chaque caractéristique, et les appliquent pour leur propre système [DIVE]. Leur travaux mettent en avant les qualités d'une représentation sous forme de clone³, qui permet donc de garder un contact visuel (on maîtrise le mouvement des yeux d'un clone), les gestes et les expressions faciales.

² Travail Coopératif Assisté par Ordinateur

³ représentation tridimensionnelle réaliste d'une personne

Avec sa représentation, chaque intervenant doit posséder un moyen de désignation à distance, visible dans l'interface de ses interlocuteurs, un télépointeur. En le faisant évoluer vers un modèle plus complexe, en dépassant le stade de la simple désignation et en permettant aux différents participants d'interagir ensemble, on améliore les possibilités de coopération.

3.3 L'activité centrée autour de la tâche

Les participants d'une réunion doivent se concentrer autour d'un ou de plusieurs documents partagés (on parle de 'documents' au sens large : données ou applications). Il est donc préférable de renforcer l'impression de collaboration autour de ces derniers, en les plaçant de manière centrale. Ce concept consistant à regrouper spatialement les documents pour recentrer l'attention est née du projet *Rooms* (voir [Henderson et al 86]) de *Xerox*. L'idée originale était de diminuer la concentration et les recouvrements de fenêtres en les regroupant en fonction d'une tâche particulière dans des bureaux virtuels, un seul étant affiché à la fois.

3.4 La métaphore de la table de réunion

Il semble donc logique de vouloir matérialiser cette séparation des tâches. Un espace ouvert et homogène ne conviendrait donc pas. On rencontre souvent dans la littérature la notion de pièce (au sens visuel et sémantique). Elle est caractérisée par un espace clos et associée à une seule tâche. Cela ne permet cependant pas de garder une vue, même subjective, des autres activités se déroulant en parallèle. Nous allons étendre ce concept par la suite à la métaphore de table de réunion, où un utilisateur peut rejoindre un groupe de personnes en gardant un contact (visuel ou sonore) sur les autres tables.

Ce concept est proche de la réalité. Pour que l'utilisateur se retrouve dans cette situation qu'il a l'habitude de rencontrer, un environnement tridimensionnel est une solution idéale - ceci pour obtenir une interface naturellement conviviale et intuitive.

3.5 Perception et soutien de la conscience mutuelle

Certaines tâches requièrent de se concentrer sur un point particulier, tandis que d'autres nécessitent une vue globale du milieu. En général, l'attention est partagée temporellement entre ces deux cas. Cependant, le TCAO exige de chaque participant qu'il sache ce qui se produit, ce qui change, où et par qui : l'interface doit pouvoir soutenir cette conscience mutuelle. L'idéal serait de voir tout ce qui se passe à tout moment dans la pièce (le 'tout visible'). Il y a cependant une limite à la quantité d'information qu'il est possible d'afficher sur un écran. On peut néanmoins tirer parti (tout en s'adaptant) des spécificités de la perception humaine. En particulier, il existe une zone centrale dans le champ visuel où l'image est très précise, alors que la vision périphérique, manquant de netteté, réagit surtout aux mouvements et aux changements. Nous simulons cette notion en plaçant le document de travail au centre de l'écran et en conservant tous les autres documents dans la scène : c'est le 'tout perceptible', que l'on obtient en diminuant l'espace occupé par les objets d'importance mineure, en les dégradant tout en les laissant visible (de manière à conserver toutes les informations à l'écran).

3.6 L'animation interactive

Elle permet deux choses. Tout d'abord, on peut densifier la quantité d'informations affichées sur l'écran, et rendre parfaitement lisible seulement une faible partie de ces dernières (voir [The Perspective Wall] ou [Cone trees]), le reste donnant, sous sa forme dégradée, une idée globale du contenu. Une animation interactive de ces données permet à l'utilisateur de visualiser correctement et rapidement chaque partie du document. L'utilisation de ce concept va nous permettre d'en respecter un autre par la suite, celui du 'tout perceptible'.

Une autre utilisation de l'animation interactive est celle de l'interface elle-même, qui devra être fluide. Nous voulons éviter les ruptures abstraites dans la continuité de la scène, qui augmentent la charge cognitive de l'utilisateur. Une simple observation doit permettre d'éviter un effort (intellectuel) inutile. Par exemple, certains systèmes graphiques décomposent la réduction d'une fenêtre à l'état d'icone en une animation linéaire, au lieu de fermer brutalement la fenêtre et de créer

un nouvel icône : plus besoin de chercher à reconstituer mentalement le lien entre les deux formes du même objet. C'est la perception qui remplace l'effort cognitif.

3.7 la navigation

Nous définissons la *navigation* comme les déplacements de l'utilisateur à l'intérieur de l'espace 3D, ce qui se traduit par des changements de point de vue de la scène. L'*interaction*, quant à elle, correspond aux actions de l'utilisateur dans la scène, qui manipule des objets sans changer de point de vue global sur cette dernière.

Ces deux éléments, navigation et interaction, sont liés car on doit pouvoir se déplacer pour agir dans l'interface. Mais la troisième dimension pose de nouveaux problèmes comme celui du positionnement ou de la direction du point de vue de l'utilisateur. En fait, il faut gérer les limites de sa perception pour ne pas le désorienter (voir [Gomez 94]) - à fortiori dans notre application où le point central n'est pas la navigation à travers l'interface, mais le travail qu'on y effectue. En effet, dans le cadre du travail coopératif, l'utilisateur n'a pas seulement une position physique dans l'interface, mais aussi une position liée au groupe. Au delà de la conscience de groupe, chacun possède un rôle (par exemple le chef de réunion et un intervenant ne feront pas la même chose). Ce rôle peut induire des limites, et l'utilisateur devra disposer d'une instance de l'interface adaptée à son propre rôle, avec sa propre vue des choses. Il devra se trouver d'autant plus impliqué dans son espace de travail, donc ne surtout pas être désorienté à cause de problèmes de déplacements. Cela contraint à définir un cadre géométrique de navigation limité, suffisant et facile à utiliser.

3.8 La manipulation

Si la navigation est limitée, il n'en va pas de même pour les actions. Nous nous plaçons dans le cas de la manipulation directe. L'utilisateur peut interagir, manipuler directement les objets de l'espace. En étendant les possibilités de mouvements dans l'interface, on peut se rapprocher d'une interaction réaliste. Comme nous nous dirigeons vers un modèle où la représentation, la navigation et l'interaction sont tridimensionnelles (cela met en avant le lien étroit que l'on va établir entre le lieu de travail virtuel et réel dans l'interface), on peut compléter ce modèle en développant l'utilisation des deux mains pour interagir. [Kabbash] met en avant les applications pour lesquelles l'utilisation des deux mains est positive, et n'apporte pas une charge cognitive supplémentaire pour l'utilisateur. En effet, il a constaté que l'utilisation des deux mains peut être pire que d'une seule si les applications assignent une tâche indépendante à chaque main. Cependant, cela permet aussi dans certains cas de s'adapter beaucoup plus vite, d'aller plus vite à l'information ou de manipuler plus facilement l'interface. Pour trouver les applications à deux mains adéquates, on veillera à respecter certaines contraintes qui placent la main gauche (pour les droitiers) comme une main qui sera le référentiel pour l'action, l'initiateur de l'action ou encore celle qui aura la partie de l'action la plus facile.

3.9 la désignation

Cependant, la liberté dans les gestes augmente également les problèmes de positionnement des mains, de perception des mouvements entre l'espace réel et l'espace virtuel (voir [Venolia]). Le modèle d'interaction doit donc être en étroite relation avec le moyen (matériel) de navigation que l'on propose : l'étude doit comprendre en parallèle le mode et le moyen d'interaction.

Pour cela, on différencie les périphériques d'entrée 3D (voir [Fuchs 1995]), qui permettent l'acquisition de données en trois dimensions (les classiques souris ne fonctionnant que suivant deux axes). Ils se présentent sous la forme de *trackballs* ou de capteur que l'on fixe au doigt et que l'on déplace en trois dimensions. Ils renvoient à l'ordinateur des indications quant à la position du périphérique : on parle de degrés de liberté. Si le périphérique gère les déplacements suivant 3 axes (les translations suivant X, Y et Z) alors il a trois degrés de liberté; si en plus, il mesure les rotations autour de ces axes, il possède six degrés de libertés (X, Y, Z et l'angle de rotation du périphérique autour de chacun de ces axes). Désormais, on utilisera le mot *périphérique* seul pour désigner les périphériques d'entrée 3D. On peut les classer suivant trois types :

Les périphériques isométriques

Ils possèdent une résistance infinie, ils sont donc immobiles. Ils mesurent le mouvement à partir de la force et du couple qu'on leur applique. Ils ont pour défaut le manque de rapport direct avec le mouvement; en effet c'est la force appliquée qui détermine le mouvement et non pas le déplacement, quasi nul, de la main. Le manque de retour (donc de proprioception) est un autre défaut, d'où un certain temps d'adaptation pour les tâches complexes. Les *trackballs* 3D en sont un exemple (figure 1a).



figure 1a : *trackball* (isométrique)



figure 1b : capteur à 3° de liberté

(isotonique)

Les périphériques isotoniques

Ils bougent librement avec le mouvement de l'utilisateur, ils possèdent donc une résistance nulle. C'est le cas des souris en 2D, ou des gants de données. Ils ont pour défauts la fatigue éventuelle due à leur utilisation prolongée, et pour certains types d'applications les limites de leur champ d'action - suffisant en ce qui nous concerne, l'utilisateur restant devant son ordinateur. Cependant ils permettent souvent d'avoir six degrés de liberté. Les capteurs à ultrasons en sont un exemple (figure 1b).

Les périphériques élastiques

Ils sont à mi-chemin entre les deux derniers types. On doit appliquer une force qui varie en fonction du déplacement, et ils sont autocentrés dans leur position stable. Ils sont réputés plus proprioceptifs pour l'utilisateur et donc plus faciles à utiliser.

Il existe des controverses sur le périphérique qui permet d'obtenir les meilleures performances (voir [Zhai]). Les isométriques fonctionnent mieux dans les modes de contrôle de taux (robotique), alors que les isotoniques sont plus adaptés aux modes de contrôle de position, lorsqu'il y a un rapport direct entre le mouvement de la main et le mouvement du pointeur.

4. Notre modèle

Nous allons maintenant exposer notre modèle d'interface, en donnant une forme à chacun des concepts énoncés ci-dessus, en définissant l'organisation de l'espace, puis les moyens d'y travailler et de coopérer.

4.1 L'organisation spatiale

Le Bureau

Bien qu'ayant des éléments communs avec la réalité virtuelle, nous nous plaçons plus dans un cadre bureautique (pas de périphérique encombrant du type casque ou gants), par opposition aux environnements immersifs. Ces derniers posent souvent des problèmes de lenteur, de précision ou de perception. Nous cherchons plutôt à éviter à l'utilisateur tous les mouvements dus aux contraintes naturelles et qui ne sont pas nécessaires à une vraie réunion. Autrement dit, la conception même de

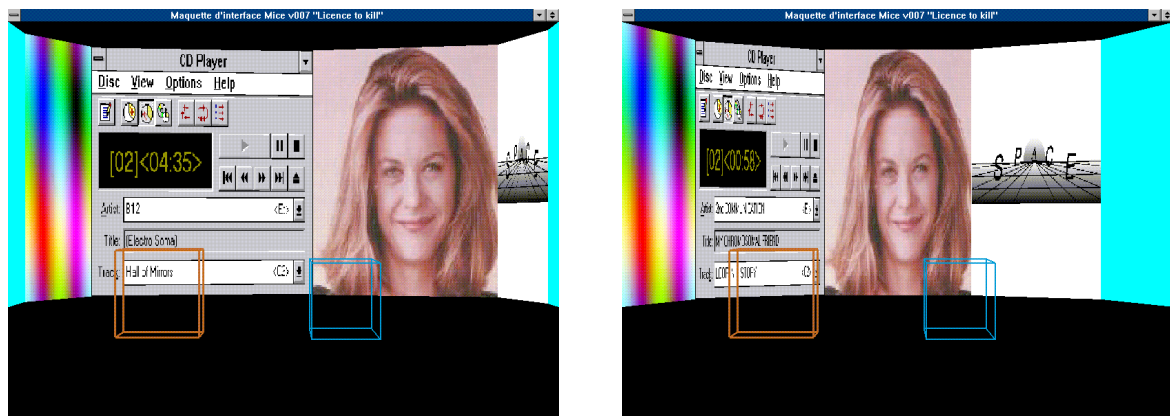
l'espace de travail devrait résoudre les problèmes de navigation, en réduisant le nombre de mouvements inutiles et perturbants pour l'utilisateur. Par exemple, des personnes assises autour d'une table ne peuvent pas lire facilement un même document en même temps. On analyse et on n'utilise pas les mouvements qui ne sont pas indispensables afin de rendre un espace simplifié et pratique.

Cet espace comprend deux types d'objets de base que sont les acteurs et les artefacts. Les acteurs sont des représentations des interlocuteurs à distance ou des assistants artificiels. Les artefacts sont des documents ou des outils.

La table de réunion

La métaphore de l'interface est la table de réunion (*rooms*), qui correspond à une seule et même activité. Nous y opérons un découpage spatial et sémantique en deux parties. La première partie est un bandeau où l'on place les acteurs et les documents partagés (voir ci-dessous). La seconde correspond à l'espace intérieur de ce bandeau - c'est un espace de travail particulier (pour un document occupant une position centrale), et c'est aussi l'espace privé de l'utilisateur (pour ce qui n'est pas partagé).

Nous plaçons donc les acteurs et les documents partagés (2D ou 3D) autour d'une table, sur des 'murs'. Par commodité, on place un document par mur. Pour en représenter autant que l'on veut à la fois, on les place suivant un demi-cercle afin de garder l'ensemble des documents visibles. L'utilisateur peut recentrer son point d'intérêt en tournant ce bandeau, action assimilable à celle de tourner la tête dans la réalité. Cette répartition est naturellement uniforme et simule une situation réelle de réunion où plusieurs personnes sont assises autour d'une table. Cela permet d'insérer tout nouvel arrivant (ou un nouveau document) sans qu'il y ait d'a priori sur l'espace qu'il occupe par rapport aux autres interlocuteurs présents dans la scène.



l'interface avec deux niveaux de déformation différents

Distorsion

Si le nombre d'objets autour de la table augmente, ils deviendront tous trop fins pour que l'on puisse les utiliser. On définit donc une notion de focus, qui met en valeur le document dans la zone principale de vision, au centre de l'écran. Cela impose une déformation des autres documents sur les côtés.

Notre déformation a lieu dans un sous-espace (parallélépipède rectangle), uniquement en abscisse. On considère qu'un point d'abscisse nulle sera au milieu de la boîte. A un point d'abscisse $x \geq 0$ (la déformation est symétrique par rapport au plan $x=0$), on applique la formule suivante :

$$x' = \left(1 - \left(\frac{S-x}{S} \right)^\alpha \right) \times S$$

où S est la demi-largeur de la boîte et α le facteur de déformation. Avec $\alpha=1$, la fonction est l'identité. Pour $\alpha > 1$, les points se rapprocheront du bord, avec pour résultat d'étirer les objets centraux, tandis que ceux situés près du bord seront tassés vers l'extérieur. Nous multiplions α plutôt que de lui ajouter une valeur fixe pour avoir des changements plus linéaires de la déformation.

Le tout Visible

Grâce à cette déformation, les parties importantes sont donc parfaitement lisibles, le reste conserve une place dans l'environnement. Et lorsque l'on tourne le bandeau, ce qui disparaît d'un côté réapparaît immédiatement de l'autre (les autres éléments à l'intérieur de la scène ne bougent pas).

Cela permet à l'utilisateur de toujours tout percevoir dans l'interface. Il est en effet essentiel en TCAO que chaque artefact et chaque acteur impliqué dans une tâche coopérative soit présent à l'écran.

Un point de focus

Cette distorsion amène l'utilisateur à placer naturellement au centre la partie du bandeau avec laquelle il travaille. Cette information sur le centre d'intérêt permet de reproduire, sans dispositif spécial, la sensation de contact visuel et de direction du regard d'un intervenant dans les interfaces de ses interlocuteurs.

En dehors du bandeau, l'espace intérieur est un espace de travail, où l'utilisateur peut apporter (agrandir) le document sur lequel il travaille, y placer des documents (2D ou 3D) pour les éditer ou les manipuler. La navigation de l'utilisateur est donc réduite virtuellement à une rotation (bandeau) et à un déplacement en profondeur (zoom du point de focus).

4.2 L'interaction

Le pointeur

Pour interagir, nous avons préconisé dans la première partie l'utilisation des deux mains de l'utilisateur afin de profiter au maximum des possibilités de la 3D et de son rapport à la réalité. L'utilisateur possède tout d'abord un pointeur qu'il déplace au moyen d'un périphérique d'entrée. Dans l'interface, le pointeur symbolise les déplacements de votre point d'action.

Le périphérique associé au déplacement du pointeur sera isotonique et placé dans la main dominante de l'utilisateur. On utilise alors trois degrés de liberté (les trois axes X, Y et Z). On complète ce système d'interaction par un périphérique dans l'autre main pour aider lors des rotations, déplacements et faciliter la manipulation instantanée du bandeau (rotation et zoom). Pour minimiser les déplacements de la main et éviter ainsi les problèmes de synchronisation des bras, on pourrait préférer un périphérique isométrique dans la seconde main (un *trackball*). Cependant, pour retranscrire plus fidèlement les gestes, on peut utiliser un second périphérique isotonique (soit deux *polhemus* par exemple). Ces deux possibilités (isométrique et isotonique contre deux isotoniques) feront l'objet d'une étude.

En effet, pour capter les mouvements, on utilise les 6 degrés de libertés de la main dominante et si possible également de l'autre main, sous peine de devoir déterminer arbitrairement les mouvements de cette dernière. Il faut alors deux capteurs isotoniques à six degrés de liberté, système minimum pour pouvoir calculer la cinématique inverse des bras à partir du mouvement des mains (voir [Sadeghin]).

La transcription du mouvement de la main dominante (pour le pointeur) dans l'interface est instantanée, et agit dans le cadre de la pièce. En atteignant les murs, le pointeur 3D se transforme en pointeur 2D agissant uniquement sur le mur, jusqu'à ce que l'utilisateur l'en éloigne. Il s'agit d'une action implicite, qui se déclenche suivant la position du pointeur dans l'interface. On utilise cette méthode des actions implicites pour ouvrir des menus, des aides ou des panneaux de commande. Par exemple, l'ouverture d'un panneau de commande du bandeau pourra être accessible dans une fenêtre en transparence sur le bas du mur lorsque l'utilisateur se situe à proximité du bandeau, et proposera des opérations de rotation, de zoom, etc.

Malgré l'utilisation de périphériques appropriés, il est essentiel que l'utilisateur ne perde pas son pointeur lors de ses déplacements. Plusieurs techniques existent pour cela. Tout d'abord, le changement d'orientation en fonction de la position du curseur dans l'espace renforce l'effet de mouvement. Ensuite, pour améliorer sa visualisation en trois dimensions, nous jalonnons régulièrement le bureau (dans l'interface, l'équivalent du sol) de repères qui permettent avec la perspective un repérage facile et direct (un quadrillage par exemple). Enfin, nous utilisons les effets de lumière en donnant au curseur une ombre portée sur le sol, dont l'intensité diminue avec sa hauteur, afin de conserver un rapport taille/intensité constant. Cela permet de bien situer sa profondeur.

La forme de ce pointeur a une importance non négligeable. Si on est tenté d'utiliser une flèche, digne translation du pointeur 2D vers la 3D, nous avons plutôt opter pour une solution adaptée aux possibilités de l'interface. Autrement dit, l'animation du pointeur nous permet de le transformer

en une forme correspondant à chaque type d'action possible. L'action la plus fréquente, la préhension, nous amène au modèle de la main. Nous en utilisons une forme simple (graphiquement), masquant peu les objets, et rendant ces derniers transparents quant le pointeur passe derrière afin de toujours avoir celui-ci visible.

Interaction 3D

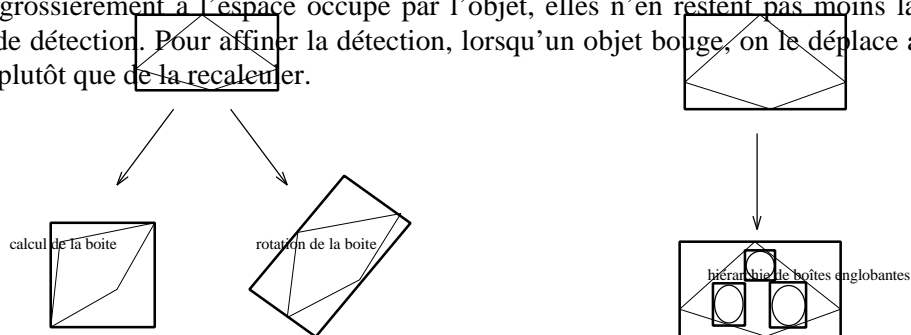
A l'aide du pointeur, l'utilisateur doit pouvoir agir dans l'interface. On lui permet donc de sélectionner tous les objets graphiques. Le mode de sélection doit être simple. On utilise des signaux visuels (apparition d'entourage graphique autour de l'objet) pour montrer qu'un objet est sélectionné ou sélectionnable. Une fois celui-ci sélectionné, les actions à opérer sur l'objet peuvent ensuite apparaître autour de ce dernier sous forme de symboles explicites (ouverture, déplacement, etc.). Le reste dépend principalement de l'application, qui pourra posséder ses propres primitives d'interaction.

Après avoir sélectionné, l'utilisateur peut vouloir manipuler l'objet. Si on dispose d'un seul périphérique dans une seule main, l'opération s'avère peu pratique. Une première solution consiste à faire appel aux *widgets*, pour des raisons de facilité d'utilisation. Il s'agit alors de faire apparaître autour de l'objet des icônes 3D (comme dans les visualiseurs *VRML* par exemple) qui pourront appliquer chacun, quand on clique dessus, une rotation ou un déplacement à l'objet. Il s'agit de 'périphérique virtuels'.

L'utilisation des deux mains est une autre solution (pas forcément incompatible avec la première) qui s'avère dans ce cas un moyen efficace de manipuler les objets, à raison d'un périphérique dans chaque main. En dehors de la manipulation, cela permet également un repérage et une orientation en trois dimensions plus aisés. La position des mains l'une par rapport à l'autre permettant de mieux évaluer l'espace, la deuxième main sert alors de référentiel, ou vient comme un appui et une aide pour travailler de façon plus naturelle. Cette réalité nous engage à utiliser les possibilités d'interaction à deux mains quand cela est possible, en fonction de l'application et de la situation, afin de faciliter le travail dans l'interface.

Boîtes englobantes hiérarchiques

Pour que l'on puisse interagir sur des objets, il faut détecter les interactions lors des déplacements. Il existe plusieurs manières de détecter des collisions en trois dimensions. Nous travaillons dans une interface, avec peu ou pas d'objets de très petites tailles, et notre but est la détection globale de collision, nous ne cherchons pas à mesurer la zone de choc. Les contraintes étant relativement faibles, on se contente donc de boîtes englobantes. Si elles ne correspondent parfois que grossièrement à l'espace occupé par l'objet, elles n'en restent pas moins la méthode la plus rapide de détection. Pour affiner la détection, lorsqu'un objet bouge, on le déplace avec sa boîte englobante, plutôt que de la recalculer.



Nous utilisons également une hiérarchie d'objets (donc de boîtes englobantes) pour certains objets 3D complexes ou pour des zones d'espace. On détecte donc les collisions avec un premier niveau de boîtes englobantes, à gros grain. Puis à l'intérieur de la boîte détectée, on affine en utilisant les boîtes des objets qu'elle contient, et ainsi de suite, on peut cumuler plusieurs niveaux.

L'utilisateur peut donc se déplacer et toucher certains objets, mais il faut également qu'il puisse agir dessus. En approchant simplement l'objet de près, on affiche un menu 3D autour de l'objet qui contient les actions applicables à ce dernier.

Menus et widgets 3D

Nous utilisons actuellement des menus 3D circulaires, entourant l'objet et composés d'icônes 3D. Ces icônes déclenchent des actions lorsque l'on les sélectionne ou sont directement manipulables en tant que *widjets* 3D (comme pour le déplacement ou la rotation). Ces menus sont facilement manipulables.

Acteurs & Artefacts

Comme nous l'avons précisé plus haut, une personne peut être représentée par plusieurs moyens. Si la vidéo permet de voir réellement une personne, cela nécessite des réseaux très performants, et les données ne sont quasiment pas manipulables. De plus, en fonction du point de vue de la caméra, les gestes et l'attitude de l'interlocuteur prennent des sens différents. Sans poser ce type de problème, les clones possèdent plus de fonctionnalité.

Un clone est une représentation synthétique et tridimensionnelle des intervenants. Le modèle 3D du visage est obtenu à partir de deux photos (face et profil), qui permettent également la création d'une texture complète de la tête, mappée ensuite sur le modèle 3D du visage (voir le projet [télévirtualité]). Le clone est donc très ressemblant et permet d'identifier chacun. Son rôle est de montrer précisément les actions des utilisateurs à distance, de façon intuitive et cohérente. Par exemple, les mouvements de tête et des yeux sont contrôlables et permettent d'indiquer le centre d'intérêt, les bras permettent de manipuler à distance, etc. C'est la première utilisation des clones. La seconde utilisation, plus intuitive, est de l'utiliser comme représentation visuelle et gestuelle d'un interlocuteur. Le clone reproduit alors les expressions faciales, les mouvements de tête et des bras [Viaud et Saulnier 1995]. L'utilisation du clone permet donc de reproduire les gestes au cours d'une conversation, voire de les provoquer.

D'un point de vue technique, il ne s'agit plus de transmettre les images sous leur forme brute (comme la vidéo), mais seulement les informations relatives aux transformations du modèle dans le temps. Le petit volume de données à échanger permet d'utiliser des réseaux à bande étroite. Mais les clones ne suffisent pas pour permettre à chaque utilisateur de connaître la position et les actions des interlocuteurs distants. Nous avons donc développé la notion de télépointeur.

télépointeur

Un télépointeur est un pointeur distant, apparaissant dans votre univers de travail comme l'outil de manipulation d'un de vos interlocuteurs. Son déplacement est couplé à ses gestes grâce au périphérique d'entrée. Il a un domaine d'action restreint, propre à l'application partagée par les intervenants. Sa représentation est liée à celle de l'interlocuteur distant dans votre interface, sous forme de bras dans le cas d'un clone : son premier rôle est la désignation, l'annotation. Ce bras peut alors opérer directement sur les applications environnantes.

Lorsqu'un interlocuteur ne place plus son pointeur dans une zone d'applications partagées, on ne le place plus comme un télépointeur équivalent dans les autres interfaces; mais on interprète ses mouvements comme des gestes dans le dialogue que son clone continue de répéter - les gestes étant un vecteur d'informations essentiel dans la conversation.

5. Implémentation

Les figures illustrant ce papier sont tirées du prototype de notre interface. C'est est un produit fonctionnel qui intègre les concepts cités précédemment. Il a été développé sur PC sous *windows NT*. Nous avons volontairement évité les stations de travail haut de gamme, notre objectif étant d'avoir des résultats sur des machines à faible coût. Nous avons utilisé *OpenGL* comme librairie graphique - standard et présente sous *Windows NT* et *95*, elle permet un bon compromis entre les primitives graphiques disponibles et les performances de l'affichage sur une machine possédant un accélérateur graphique.

6. Conclusion

Nous avons présenté un nouveau modèle d'interface 3D pour le TCAO. L'architecture centrée sur la tâche correspond à l'idée de garder tout perceptible à tout moment. La métaphore utilisée pour

l'interface est la table de réunion. Les acteurs et les documents sont placés autour de la table, sans recouvrement. Une distorsion spatiale tire partie des particularités de la vision humaine, mettant en valeur la partie centrale de l'écran pour que l'utilisateur en fasse son centre d'intérêt. Ce dernier conservant une vue dégradée des autres objets, mais suffisante pour percevoir les changements.

Le premier prototype va maintenant servir de point de départ à l'évaluation de notre interface.

Par la suite, d'autres aspects pourront être envisagés. Nous pourrions faire des études sur la mise en valeur de l'espace de travail 3D grâce à la visualisation informatique (voir le projet [Digital Libraries]).

Au niveau de l'interaction, si le gestuel est important dans le dialogue, il peut également l'être au niveau de l'interface. Dans le cas d'un périphérique à six degrés de liberté, les mouvements de rotation de la main peuvent permettre d'affiner le déclenchement de certaines fonctions, voire de les associer à un geste logique (tourner une poignée pour une porte, tourner la main pour la rotation d'un objet, etc.). On parle alors d'interface gestuelle. Cependant ces gestes peuvent également compliquer l'interface et augmenter le temps de réaction des utilisateurs.

L'exploitation du sens tactile est aussi une voie intéressante, qui peut apporter des informations supplémentaires à l'utilisateur. En particulier, nous pourrions renforcer l'impression visuelle de collision avec un objet grâce à un retour de contact tactile. Cependant les périphériques avec un retour d'effort sont actuellement assez rares et peu pratiques. Un retour léger de toucher pourrait être suffisant.

On veut ainsi tirer parti aux mieux des capacités de l'Homme (utilisation du regard, des deux mains), de ses sens (toucher, ouïe) souvent sous-exploités pour rendre notre interface efficace et intuitive.

Au niveau du rendu, on peut également tirer profit des possibilités d'éclairage de l'interface pour montrer qu'un objet est sélectionné (ou que quelqu'un parle), en l'éclairant davantage, voire en l'animant.

Nous appliquerons, entre autres, ces possibilités afin de dégager un modèle d'interface 3D optimum.

BIBLIOGRAPHIE

[Buxton 1992]

William A. S. Buxton, **Telepresence: Integrating Shared Task and Person Spaces**, p 123-129, Proceedings of Graphics Interface '92

[Okada et al 1994]

Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ickawaa et Yutaka Matsushita, **Multiparty Videoconferencing at Virtual Social Distance: MAJIC design**, p 385-393, ACM 94 Conference on CSCW

[Cassel et al 1994]

Justine Cassel, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost et Mththew Stone, **Animated Conversation : Rule-based Generation of Facial Expression, Gesture and Spoken Intonation, for Multiple Conversational Agents**, p 413-419, proceedings Siggraph'94

[Fuchs 1995]

Philippe Fuchs, **Introduction aux Techniques de la Réalité Virtuelle**, Ecole des Mines de Paris, 1995

[Gomez et al 1994]

J.E. Gomez, Dan Venolia, Andries van Dam, Tony Fields, Rikk Carey, **Why is 3d interaction so hard and what can we really do about it ?**, Proceedings Siggraph'94

[Kabbash et al 1994]

Paul Kabbash, William Buxton et Abigail Sellen, **Two-Handed Input in a Compound Task**, Proceedings of CHI'94

[Venolia 1993]

Dan Venolia, **Facile 3D Direct Manipulation**, Proceedings of InterCHI'93

[Viaud et Saulnier 1995]

Marie-Luce Viaud et Agnès Saulnier, **Real Time Analysis and Synthesis Chain**, proceedings of International Workshop on Automatic Face and Gesture Recognition, 1995

[Zhai et al 1994]

Shumin Zhai and William Buxton and Paul Milgram, **The “ Silk Cursor ”, Investigating Transparency for 3D Acquisition**, Department of Industrial Engineering, University of Toronto, 1994

[Digital Libraries]

le projet au PARC de Xerox : <http://www.xerox.com/PARC/dlbx/library.html>

[DIVE]

le projet dans sa globalite : <http://www.sics.se/dce/dive/dive.html>

[Kuzuoka et al 1994]

Hideaki Kuzuoka, Toshio Kosuge et Masatomo Tanaka, **GestureCam; a Video Communication System for Sympathetic Remote Collaboration**, p. 35-43, ACM 94 Conference on CSCW

[Benford et al 1995]

Steve Benford, Chris Grenhalgh, John Bowers, Dave Snowdon et Lennart E. Fahlén, **User Embodiment in Collaborative Virtual Environments**, p. 242-248, Proceedings of CHI'95

[Shu et Flowers 1994]

Li Shu et Woodie Flowers, **Teledesign: Groupware user experiments in three-dimensional computer-aid-design**, p. 1-14, Collaborative Computing, volume 1, number 1, 1994

[Sadeghin 96]

Aliréza Sadeghin, **Réalisation d'un système de cinématique inverse pour**

l'animation d'un modèle de corps humain synthétique à partir de capteurs magnétiques 3D, rapport interne INA, octobre 1996

[télévirtualité]

projet télévirtualité : <http://www.ina.fr/INA/Recherche/TV/TV.fr.html>

[Henderson et al 86]

D. A. Henderson, Jr. et S. K. Card, **Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface**, p. 211–243, ACM Transactions on Graphics, Vol. 5, No. 3, July 1986

[The Perspective Wall]

J.D. Mackinlay, G.G. Robertson et S.K. Card, **The perspective wall: detail and context smoothly integrated**, p. 173-180, Proceedings of CHI'91

[Cone trees]

G.G. Robertson, S.K. Card et J.D. Mackinlay, **Cone Trees: animated 3D visualizations of hierarchical information**, Proceedings of CHI'91