

# Interface Augmentée par Effet Miroir

*Frédéric Vernier, Christophe Lachenal, Laurence Nigay, Joëlle Coutaz*

*Laboratoire CLIPS-IMAG*

*Equipe IHM*

*BP 53, 38041 Grenoble, France*

*{frederic.vernier, christophe.lachenal, laurence.nigay, joelle.coutaz}@imag.fr*

## RESUME

Dans cet article, nous définissons le terme d'interface augmentée pour décrire notre approche visant à fusionner, sur un même écran, une interface traditionnelle avec le flux vidéo d'une caméra pointée sur l'utilisateur. Après avoir positionné notre approche par rapport aux deux grands courants de réalité augmentée, nous développerons les bases de programmation que nous avons mises en œuvre dans notre système. Nous présenterons ensuite deux types d'utilisation de ce système. La première classe de domaine applicatif regroupe des systèmes qui intègrent le geste explicatif dans une communication médiatisée en fusionnant le document, source de l'explication, et l'image de l'utilisateur. Le second type d'utilisation que nous proposons dans cet article combine la zone de dessin d'un logiciel de retouche d'images avec l'image de la main de l'utilisateur. Nous terminerons cet article par notre point de vue sur la conception de ces systèmes en mettant particulièrement en avant les contraintes de conceptions que nous avons rencontrées et comment nous les avons résolues.

**MOTS CLES :** Interface augmentée, Réalité augmentée, Vidéo, Effet Miroir.

## REALITE ET INTERFACE AUGMENTEE

Le concept de Réalité Augmentée (RA), qui vise l'association harmonieuse du réel et du virtuel, est issu de courants de pensée disjoints menés en parallèle en Synthèse Graphique comme en Interaction Homme-Machine :

- En Synthèse Graphique, la RA est une dérivation de la Réalité Virtuelle : typiquement, de "vrais" pixels d'origine vidéo sont combinés à des pixels de synthèse [1]. Par exemple, des meubles de synthèse sont agencés par manipulation directe dans le film vidéo d'une cuisine réelle. Ainsi, l'utilisateur manipule-t-il des objets purement électroniques dont certains représentent un monde qui existe quelque part sur Terre (la cuisine en question), mais qui n'appartiennent pas au lieu physique actuel de l'interaction (l'agencement de la cuisine se faisant, par exemple, dans un cabinet d'étude).
- En IHM, la RA, en réaction précisément à l'immersion virtuelle, s'appuie résolument sur la conservation et l'amplification du réel. Par exemple, une feuille de papier physique, objet

familier s'il en est, est amplifiée de capacité de traitements de l'information [9, 17]. Par contraste avec l'approche précédente, l'utilisateur manipule ici des objets réels qui font corps avec le lieu physique actuel de l'interaction.

Dans cet article, nous proposons une nouvelle perspective, intermédiaire entre les deux extrêmes que nous venons de présenter : la notion d'interface augmentée. Une interface augmentée est une interface graphique usuelle amplifiée de pixels vidéo montrant une scène, lieu actuel physique de l'interaction. Techniquement, l'amplification se traduit par la fusion de l'image graphique produite par l'interface homme-machine du système avec une image vidéo temps réel capturée par une caméra pointée sur l'environnement interactionnel pertinent pour la tâche en cours. Comme dans l'approche graphique, une interface augmentée s'appuie sur la superposition de "vrais" et de "faux" pixels mais contrairement à l'approche graphique, les pixels vidéo forment un miroir électronique de la réalité. Comme dans l'approche IHM (cf. Mackay), une interface augmentée exploite l'apport du réel ambiant mais contrairement à Mackay, les objets centraux de la tâche sont électroniques. Les figures 2, 3 et 6 sur lesquelles nous reviendrons, illustrent le concept d'interface augmentée.

Cet article est structuré ainsi : dans la section qui suit, nous décrivons le mécanisme technique de fusion des pixels. Nous illustrons ensuite le concept d'interface augmentée par deux classes d'application représentative. Nous terminons la discussion en précisant les contraintes de conception en sorte que l'effet miroir soit respecté.

## SYSTEME DE FUSION (IHM + VIDEO)

Une interface augmentée s'appuie, on le rappelle, sur la fusion de deux sources de pixels : les pixels vidéo acquis en temps réel par une caméra et les pixels qui constituent l'image graphique de l'IHM. Le matériel nécessaire à la mise en œuvre de cette technique est simple et peu onéreux. Il s'agit d'un écran, d'une carte d'acquisition et d'une caméra.

La fusion s'effectue pixel par pixel en appliquant la formule suivante à chacune des composantes (Rouge, Vert, Bleue) du flux vidéo :

$$RVBRes = \frac{CoeffIhm * RVBIhm + CoeffVid * RVBVid}{CoeffIhm + CoeffVid} \text{ où :}$$

RVBRes désigne la valeur du pixel résultat, RVBIhm, la valeur du pixel source de l'IHM, RVBVid, la valeur du pixel source vidéo. Les coefficients de composition CoeffIhm et CoeffVid définissent les proportions relatives « d'IHM et de vidéo » dans l'image résultante. Afin de simplifier le raisonnement nous ne considérons que les coefficients dont la somme est égale à 1. Cela ne restreint en rien les possibilités de fusion puisque c'est la proportion entre les deux coefficients qui entre en compte dans l'image finale.

La fusion étant appliquée au niveau du pixel, le calcul exprimé par la formule ci-dessus est exécuté près de 30 millions de fois par seconde pour des images sources de 640 par 480 pixels à la fréquence de 30Hz (standard NTSC). Si la formule précédente pouvait être calculée en une instruction de processeur, un processeur à 30MHz serait suffisant mais une telle opération nécessite malheureusement plusieurs dizaines de cycles d'horloges d'un processeur normal où les instructions sur les réels sont plus lentes que les instructions sur les entiers.

Afin de ne pas restreindre notre système aux configurations munis de processeurs scientifiques, il convient donc de proposer une amélioration du coût algorithmique de ce calcul. Une optimisation simple consiste à utiliser des coefficients de la forme  $m/2^n$ , par exemple (1/2, 1/2), (1/4, 3/4) ou (3/8, 5/8). Ce faisant, nous exploitons les capacités d'optimisation intrinsèque des processeurs : les divisions par puissance de 2 se font par instructions de décalage bit à bit. De plus, comme le montre l'expression C ci-dessous, l'utilisation d'un masque spécifique permet d'appliquer le décalage sur les trois composantes R, V, B à la fois.

Cette expression C traduit notre version optimisée de la formule initiale de fusion appliquée à la fusion (1/2, 1/2):

$$RVBRes = ((RVBVid >> 1) \& 0x007F7F7F) + ((RVBIhm >> 1) \& 0x007F7F7F);$$

La valeur du masque 0x007F7F7F se justifie ainsi : un pixel est codé sur 32 bits sous la forme XRGB où R, V et B désignent respectivement un octet par composante Rouge, Verte, Bleue (X n'est pas pertinent et ne sert

qu'à terminer le mot de 32 bits).

Le passage aux autres couples de coefficient s'opère par addition de coefficients de la forme  $1/2^n$ . Pour l'exemple de la fusion (1/4, 3/4) 3/4 se décompose en  $1/2 + 1/4$  :

$$RVBRes = ((RVBVid >> 2) \& 0x003F3F3F) + ((RVBIhm >> 1) \& 0x007F7F7F) + ((RVBIhm >> 2) \& 0x003F3F3F);$$

Si notre formule permet d'obtenir une fusion en temps réel sur les calculateurs actuels, la restriction qu'elle impose sur les coefficients de composition limite la diversité des images résultantes. Cette limitation peut être compensée par l'exploitation de la transparence. La transparence d'un pixel qui appartient à un objet de l'interface est calculée à partir de la valeur du pixel de l'objet (RVBObj), la valeur du pixel de fond (RVBFond) et des coefficients CoeffObj et CoeffFond qui déterminent la puissance de l'effet de transparence.

On détermine la composante RBVIhm préalablement, à la construction ou à la modification de l'interface, en utilisant la même formule que précédemment mais adaptée au traitement du fond et de l'objet de l'interface :

$$RVBIhm = \frac{CoeffObj * RVBObj + CoeffFond * RVBFond}{CoeffObj + CoeffFond}$$

L'introduction de la transparence n'obéit pas aux mêmes contraintes car cette opération n'est pas temps réel. On peut donc introduire ici des coefficients réels quelconques. Notre préoccupation est aussi de prouver que notre effet de miroir n'est pas incompatible avec la transparence.

Cette combinaison de miroir et de transparence est illustrée par l'image de la Figure 1 qui, en son centre, est autant réfléchissante (effet miroir) que transparente. L'effet miroir est maximal (en bas à gauche de l'image) lorsque seule l'image source vidéo est visible (CoeffObj = 0, CoeffVid = 1 et CoeffFond = 0). L'effet transparent est maximal (en haut à droite de l'image) lorsque seul le fond est visible (CoeffObj = 0, CoeffVid = 0 et CoeffFond = 1). Le carreau bleu est inchangé (en haut à gauche) lorsque ni le fond ni la vidéo ne sont visibles (CoeffObj = 1, CoeffVid = 0 et CoeffFond = 0).

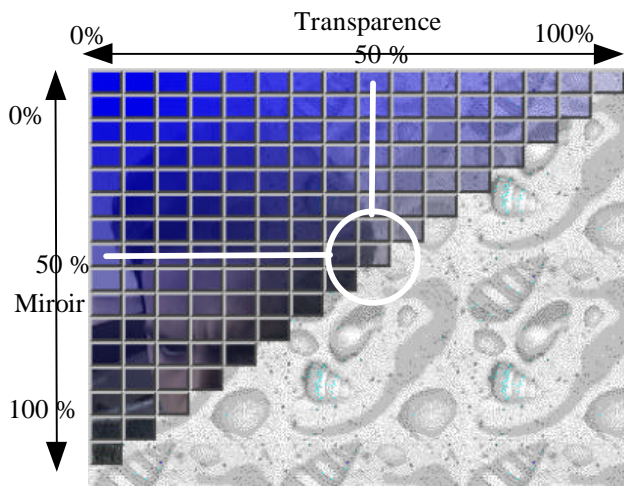


Figure 1 : Carte des effets miroir et transparence.

L'interface de la Figure 1 est composée d'une mosaïque de carreaux bleus qui reflètent différentes valeurs de CoeffObj, CoeffFond et CoeffVid. L'image de fond de l'interface est une image répétitive composée de motifs marins. L'incrément entre deux valeurs consécutives de CoeffVid et CoeffFond est de 1/16.

#### APPLICATIONS DU SYSTEME DE FUSION

La fusion de l'IHM graphique et de l'image vidéo d'une scène dans laquelle s'insère l'utilisateur, revient à superposer deux repères spatiaux : celui de l'IHM graphique et celui de l'environnement réel d'interaction rendu visible par effet miroir. Nous présentons ci-dessous deux exemples complémentaires d'exploitation de cette superposition réalisée avec notre système de fusion. Il s'agit de télé-explication et de retouche d'images. Dans le premier cas, l'effet miroir est une modalité de sortie alors que dans la seconde application l'effet miroir sert de modalité d'entrée.

##### Application à la télé-explication

Une explication fait souvent usage de gestes déictiques à l'adresse d'objets, sujets du discours. En coprésence, l'orateur et l'observateur partagent un même univers physique. A distance, une interface augmentée fondée sur la métaphore du miroir, permet de pallier la discontinuité des espaces.

L'effet miroir, on l'a vu, consiste à intégrer dans l'interface graphique électronique la scène réelle pertinente pour la tâche. Dans le cas de la télé-explication, l'orateur et ses gestes sont les entités pertinentes du monde réel : ils seront donc intégrés à l'interface augmentée. L'orateur se voit dans l'interface : comme devant un miroir, il peut désigner du doigt les concepts électroniques présentés dans l'interface. Il lui suffit de placer la main de façon à ce que l'image de son doigt se superpose à l'objet en question. A distance, l'observateur reçoit l'image

fusionnée. L'espace électronique, objet de la discussion, et l'orateur ne font qu'un.

La Figure 2 montre un exemple de télé-explication inspirée de la présentation de la météo et réalisée avec notre système. Dans cette situation interactionnelle, le présentateur voit sur un moniteur (ou projeté en face de lui) l'objet du discours (la carte) et sa propre image inscrite dans la scène. Cette interface augmentée offre deux avantages sur le système en vigueur à la télévision :

- Le présentateur ne masque pas la carte puisqu'il est transparent. En conséquence, le téléspectateur peut à tout instant inspecter la région qui l'intéresse,
- Le présentateur, qui voit la carte devant lui, n'est pas contraint de se tourner sur le côté pour en consulter le contenu : son espace interactionnel (carte et téléspectateurs) est devant lui et il en fait partie intégrante.



Figure 2 : Interface augmentée : Carte météorologique avec effet miroir.

Nous envisageons d'autres domaines d'application de la télé-explication fondés sur notre système : le dépannage à distance et le télé-tutorat.

Un dépanneur reçoit l'image électronique d'un objet à réparer. Sur l'interface augmentée par effet miroir, le dépanneur montre à distance comment réparer l'objet en désignant sur l'image les manipulations à effectuer. Il peut en particulier reproduire les gestes à effectuer sur l'objet réel. Cette technique, qui s'appuie sur du matériel courant du commerce, est plus simple à réaliser que la GestureCam [7] ou le GestureLaser [8] qui nécessitent l'un et l'autre du matériel dédié (mini-robot manipulateur de la GestureCam et miroirs montés sur un moteur pas à pas dans le cas du GestureLaser). Cependant, nous ne sommes pas en mesure de comparer l'efficacité interactionnelle de ces diverses approches.

Le télé-enseignement constitue un autre domaine applicatif privilégié de notre technique. L'enseignant peut fournir des explications à l'étudiant distant qui rencontre des difficultés dans l'utilisation d'un logiciel. Il montre la façon de faire directement sur l'interface du logiciel. L'étudiant effectue les actions en même temps qu'il reçoit les explications. De fait, l'étudiant est en situation d'apprentissage actif. Avec un système de prise de contrôle à distance comme VNC [14], soit l'étudiant agit et le tuteur ne peut pas montrer en même temps qu'il explique, soit le tuteur agit et l'étudiant est en situation de consommation passive. Ce système d'enseignement à distance, même s'il s'avère coûteux en bande passante réseau, va dans le sens du plein développement des réseaux de visioconférence et de visioconférence. Il propose en outre une utilisation productive de l'image envoyée.

### Application à la retouche d'images

La Figure 3 montre un exemple de retouche d'images au moyen d'une interface augmentée par effet miroir de la main. Dans ce scénario, la caméra observe, comme dans le cas du bureau digital, une feuille de papier sur laquelle l'utilisateur dessine avec les instruments de son métier (pinceau, crayon, gomme, etc.). Face à lui, est présentée sur un écran, l'image de l'interface augmentée. Comme le montre la Figure 3, l'utilisateur y voit l'objet de la tâche (l'image électronique à retoucher), sa main et les inscriptions produites sur la feuille de papier.



Figure 3 : Combinaison d'un dessin en cours d'édition avec la main de l'utilisateur.

Au départ, la feuille de papier est blanche et le portique de l'image n'a pas de titre. Celui-ci va être peint à la main au moyen de l'outil qui convient à la situation. Lorsque l'utilisateur amène son outil (crayon, pinceau, stylo, etc.) sur la feuille, l'image de la main est affichée à l'écran et fusionnée avec l'image à retoucher. Cette juxtaposition permet de déposer de l'encre au bon endroit dans l'image électronique (dans notre cas, le texte « Plage du soleil .. » sur le portique). A la

demande de l'utilisateur, le système met en œuvre un algorithme d'extraction de l'encre déposée sur la feuille [16]. Ce faisant, l'encre physique de la feuille de papier pénètre sans rupture interactionnelle dans l'espace électronique. Chaque extraction d'encre donne lieu à la création d'un calque électronique qui se superpose à l'image électronique. L'utilisateur termine sa retouche en fusionnant les calques.

L'apport majeur de cette interface augmentée est la souplesse offerte dans le choix des modalités d'entrée : pour une tâche donnée, l'utilisateur a le choix entre les instruments du métier et les outils informatiques. De plus, ce choix est décidable à tout instant : l'utilisateur peut dessiner avec son feutre favori et changer ensuite la couleur du trait en utilisant les services du logiciel.

### AUTRES SYSTEMES DE FUSION AVEC EFFET MIROIR

Les systèmes de visioconférence classiques offrent des interfaces à effet miroir dégradé : les sites distants et l'image de la scène locale sont rendus en juxtaposition dans des fenêtres disjointes. Il en résulte une rupture visuelle que des systèmes comme Majic [6] et HyperMirror [12] ont tenté de combler.

Dans Majic, les images vidéo des sites distants sont assemblées en un collage sans couture. Ainsi l'utilisateur voit-il en face de lui sur grand écran une image « grandeur nature » de ses collègues qui, bien que géographiquement répartis, donnent l'illusion presque parfaite qu'ils sont dans la même salle. Mais, comme dans une réunion en face à face, l'utilisateur ne se voit pas en réflexivité dans le montage vidéo : Majic fusionne mais n'est pas réfléchissant.

Contrairement à Majic, HyperMirror ne vise pas la reproduction du présentiel, mais le WYSIWYS strict : tous les participants, locaux et distants, se voient dans la même scène projetée en grand sur le mur. La nouveauté de la situation interactionnelle se fonde entièrement sur l'effet miroir obtenu par la composition des deux sources vidéo. HyperMirror fusionne et est réfléchissant. On relèvera dans les conclusions de cette étude le « caractère naturel » des gestes parce que l'utilisateur se voit agir.

Comme dans Majic et HyperMirror, Fovea [4, 5] produit une image composite à partir de deux sources vidéo : une image à basse résolution capturée par une première caméra fixe est composée avec une image haute résolution acquise par une seconde caméra mobile et pilotable à distance à la souris. Comme le montre la Figure 4, la composition, qui s'appuie sur un algorithme similaire au nôtre, s'inspire de l'architecture de la vision humaine : l'image à haute résolution fournit un vue détaillée du site distant, tandis qu'à la périphérie,

on perçoit le contexte général de la scène distante. Fovea fusionne avec un effet fisheye mais n'est pas réfléchissant.

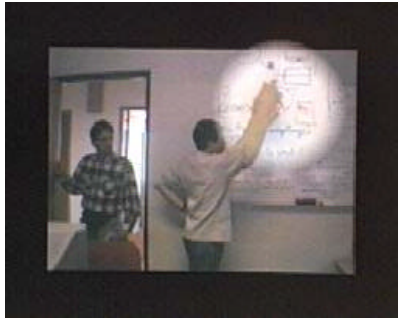


Figure 4 : Fovea.

Les systèmes que nous venons de présenter pratiquent tous trois, la fusion d'images vidéos. Contrairement à notre concept d'interface augmentée par effet miroir, l'espace de travail électronique y est absent. Dans [2], l'IHM graphique est amplifiée d'acteurs vidéo pré-enregistrés. Ceux-ci s'affichent par-dessus l'IHM graphique et apparaissent à la demande pour fournir des explications (en ce sens, ils sont interactifs). Ce système superpose (pas d'effet de transparence) et n'est pas réfléchissant.

La machine Mandala [15] et ALIVE [10] répondent à notre définition d'interface augmentée par effet miroir. Dans la machine Mandala, la vidéo des musiciens est intégrée à une scène de synthèse dont le comportement dépend des gestes du musicien. De manière similaire, ALIVE fusionne l'image de l'utilisateur dans une scène animée de créatures intelligentes (en l'occurrence un chien). Toutefois, ALIVE, démonstrateur d'architecture multi-agent et d'intégration de techniques de vision par ordinateur, n'a pas été envisagé, ni étudié, pour son apport interactionnel. En particulier, les contraintes concernant les conditions de réalisation d'un « bon » rendu de l'effet miroir, n'ont pas été clairement identifiées.

#### EFFET MIROIR : CONTRAINTES DE CONCEPTION

La métaphore du miroir met à profit notre expérience interactionnelle avec les surfaces réfléchissantes de la vie courante. A priori, la mise en œuvre électronique de l'effet miroir doit être réaliste, c'est-à-dire, respecter autant que possible, le vécu interactionnel tiré du monde réel. Nous avons retenu deux éléments de réalisme tirés de notre expérience avec les applications de télé-exploration et de retouche d'images : la conservation du contact du regard et la conservation de la qualité de l'image fusionnée.

#### Conservation du contact du regard

La conservation du contact du regard impose que la caméra soit placée au centre de l'écran. Or, les caméras sont le plus souvent installées au-dessus du moniteur.

Comme le montre la Figure 5, il en résulte un angle  $\alpha$  entre l'image source et l'image réfléchi par le miroir. Lorsque cet angle est trop grand, le contact visuel "regard à regard" est perdu. La Figure 1 montre un bon exemple d'angle  $\alpha$  inadapté : ici le regard de l'image de l'utilisateur dans le miroir est résolument tiré vers le bas.

Deux solutions sont envisageables :

1. Utiliser deux caméras comme dans [13]. Par fusion des deux images vidéo sources, la direction du regard est redressée. L'image produite est floue et consomme de précieuses ressources de calcul.
2. Placer la caméra au-dessus du moniteur et trouver le compromis avec l'angle  $\alpha$ . C'est ce que nous avons adopté, moyennant l'application des règles 4, 5 et 6 présentées dans la section suivante.



Figure 5 : Angle entre la caméra, l'utilisateur et l'écran.

#### Amélioration de la qualité de l'image finale

L'image résultante, doit, comme pour un miroir réel, rester lisible. Nous avons choisi d'exploiter les notions de contraste et de luminosité par le truchement de la technique des histogrammes d'intensité lumineuse.

Un histogramme d'intensité lumineuse est un graphique qui fait correspondre à chaque valeur d'intensité lumineuse possible (depuis le blanc jusqu'au noir), le nombre de pixel de l'image qui ont cette intensité.

L'histogramme de la Figure 7 caractérise l'image de la Figure 6 résultant d'une fusion. On constate que la plupart des pixels d'une telle image sont centrés sur les valeurs moyennes. Ceci se justifie ainsi : un pixel très sombre (resp. très clair) de l'image finale s'obtient par fusion de pixels sources très sombres (resp. très clairs). Ce cas se présente rarement dans la fusion de deux images indépendantes. Aussi, l'histogramme d'intensité lumineuse de l'image finale se resserre-t-il autour des

valeurs moyennes ce qui se traduit par le voile terne qui recouvre l'image de la Figure 6.

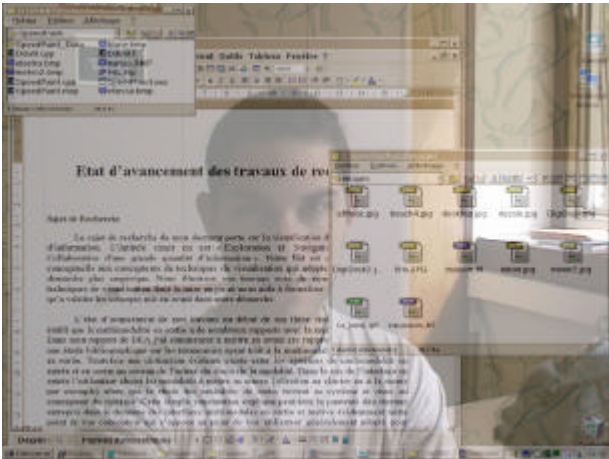


Figure 6 : Image résultante d'une fusion de deux images (interface + vidéo).

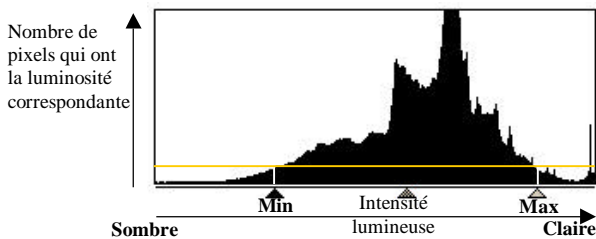


Figure 7 : Histogramme d'intensité lumineuse correspondant à la Figure 6.

Pour obtenir une image plus lumineuse et plus contrastée, nous désirons utiliser une technique appelée en infographie « technique du point noir et du point blanc ». Cette technique consiste à calculer une valeur minimale (resp. maximale) qui cerne au mieux les bords des pics de l'histogramme. Pour cela, nous calculons une nouvelle valeur relative du blanc et du noir. Une fois ces deux valeurs calculées, l'intervalle [min ; max] regroupe la majorité des valeurs des pixels. Nous définissons enfin la fonction affine adéquate par laquelle nous étendons l'intervalle : un pixel ayant la valeur min (resp. max) aura, après application de la fonction, la valeur 0 (resp. 255). Cette technique, qui accentue les composantes RVB d'un pixel, permet d'obtenir une image comme celle de la Figure 8. Dans cette image, les objets de l'IHM graphique se distinguent davantage. L'effet de voile terne de la Figure 6 est nettement atténué car, comme le montre l'histogramme de la Figure 9, les pixels sont moins regroupés au centre que sur l'histogramme de la Figure 7.

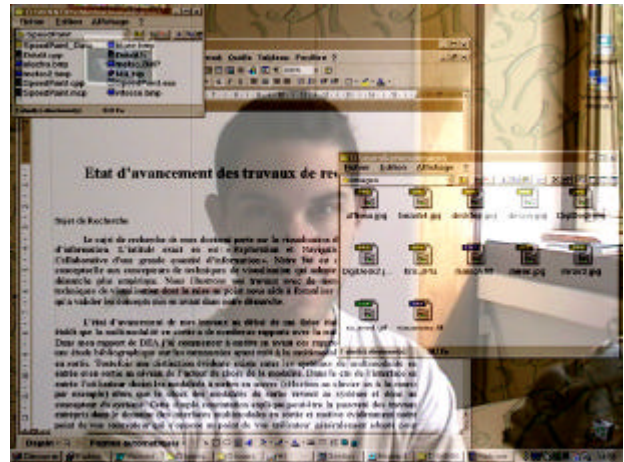


Figure 8 : Image de la Figure 6 après le traitement.

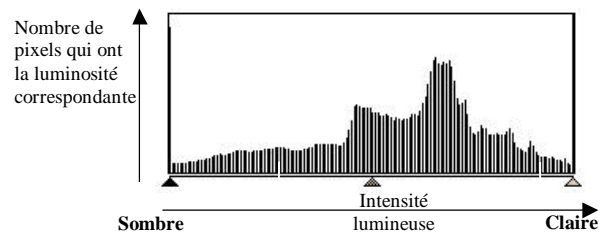


Figure 9 : Histogramme d'intensité lumineuse correspondant à la Figure 8.

### Règles de conception

Pour garantir l'effet miroir, les règles suivantes de conception doivent être satisfaites :

1. La fréquence d'affichage doit être supérieure à 20 Hz.
2. La latence entre l'acquisition et la restitution doit être imperceptible.
3. La luminosité des images sources doivent être similaires.
4. L'effet miroir doit s'appliquer au plein écran.
5. La caméra doit rester pointée sur la tête de l'utilisateur.
6. L'utilisateur doit se tenir éloigné de l'écran.

La règle 1 est liée au système perceptif humain : en-dessous de 20 Hz, l'homme perçoit une série d'images au lieu d'une animation continue.

La règle 2 tient au fait qu'un vrai miroir renvoie l'image instantanément alors que notre système introduit un délai qui comprend le temps d'acquisition, le traitement et l'affichage. Ce délai est gênant puisqu'il fixe la perception de causalité entre le geste et le retour d'information du miroir. La règle 2 impose au concepteur logiciel de traiter l'information la plus récente et de minimiser le stockage dans des espaces de mémoire-tampon dont l'accès ralentit le traitement.

La règle 3 est liée à la fusion des deux images. L'image la plus lumineuse sera davantage prégnante. En particulier, la fusion d'une image très sombre avec une image très lumineuse rend l'effet miroir inopérant.

Les règles 4, 5 et 6 visent à réduire l'angle  $\alpha$ . La règle 6 (éloignement de l'utilisateur) peut avoir une incidence sur la conception de l'IHM (par exemple, grossir la taille des icônes et celle de la police de caractères).

## CONCLUSION

L'introduction dans l'interface de l'image de l'utilisateur laisse entrevoir de nouvelles solutions d'interaction dans de nombreux domaines. Au-delà de nos deux applications qui soulignent l'intérêt de fusionner l'interface avec l'image de l'utilisateur, nous énonçons des règles de conception liées au réalisme nécessaire de l'effet miroir. Ces règles ont des conséquences sur la conception de l'interface mais aussi sur sa mise en œuvre logicielle pour réaliser les performances nécessaires (liées au nombre de registres, au jeu d'instructions et à la mémoire tampon du processeur).

Il est intéressant de noter que l'effet miroir dans le cas de l'application de télé-explication est une modalité de sortie uniquement alors que dans l'application de retouche d'images l'effet miroir devient une modalité d'entrée. Dans cette dernière application, nous avons cependant noté la difficulté d'utilisation due aux changements de modalités d'entrée entre d'une part le clavier et la souris et d'autre part le feutre et la feuille de papier (temps de prise en main des dispositifs à chaque changement [3]). Pour réduire ces changements, les tâches les plus fréquentes effectuées actuellement avec le clavier et la souris pourraient être déclenchées par des gestes reconnus par un système de vision par ordinateur [11].

Les tests préliminaires ont déjà montré l'intérêt que les utilisateurs portent à notre système ainsi que quelques limitations. Plusieurs d'entre eux ont tendu le doigt vers l'écran pour pointer un objet ce qui a eu pour conséquence de faire disparaître leur doigt du champ de la caméra. Les utilisateurs ont ainsi montré à la fois leur compréhension intuitive de la métaphore du miroir mais ont également mis en évidence une restriction du système. Enfin, si aucun d'entre eux n'a réitéré son erreur une seconde fois, ce comportement appelle à une expérimentation plus poussée de nos deux systèmes.

Notre effort à venir portera également sur une meilleure intégration entre notre système de dessin augmenté avec un système professionnel de retouche d'image. Nous escomptons ainsi pouvoir observer le comportement d'utilisateurs experts en retouche d'image.

## BIBLIOGRAPHIE

1. Azuma R. T., Tracking Requirements for Augmented Reality, Communication of the ACM n°7, Juillet 1993, Publication ACM. pp. 50-51.
2. Breiteneder C., Gibbs S., Interactive Video Actors, Actes de la conférence CHI94, Conference Companion. pp. 447-48.
3. Card S, Moran T., Newel., The psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, 1983.
4. Coutaz J., Bérard F., Carraux E. et Crowley J. *Early Experience with the Mediaspace CoMedi*, Actes de la conférence EHCI'98, Héraklion, Sept. 1998. Disponible à l'adresse [http://iihm.imag.fr/publs/1998/EHCI98\\_CoMedi.pdf](http://iihm.imag.fr/publs/1998/EHCI98_CoMedi.pdf).
5. Coutaz J., Crowley J., Bérard F., Carraux E., Astier W., Using Computer Vision to Support Awareness and Privacy in Mediaspaces, Video demo, CHI99, ACM, à paraître.
6. Ken-Ichi okada, Fumihiko Maeda, Yusuke Ichikawaa, Yutaka Matsuushita Multiparty Videoconferenceing at virtual Social Distance : Majic Design, . Actes de la conférence CSCW'94, Octobre 1994, Chapel Hill, NC, USA. pp. 385-394.
7. Kuzuoka H., Kosuge T. et Tanaka M., GestureCam : A Video Communication System for Sympathetic Remote Collaboration. Actes de la conférence CSCW'94, Octobre 1994, Chapel Hill, NC, USA. pp. 35-43.
8. Kuzuoka et al, GestureLaser : Supporting Hand Gestures in Remote Instruction. Actes de la conférence CSCW'98, Seattle, Washington USA. Vidéo pp. 424.
9. Mackay W., Velay G., Carter K., Ma C. et Pagani D., 1993, "Augmenting Reality : Computational Dimensions to paper", Communication of the ACM n° 7, Juillet 1993, Publication ACM. pp. 96-97.
10. Maes, P., Darrell, T., Blumberg, B. and Pentland, A. *The ALIVE System: Wireless, Full-body Interaction with Autonomous*. M.I.T. Media Laboratory Perceptual Computing Technical Report No 257. . Disponible à l'adresse <http://www-white.media.mit.edu/vismod/publications/publications.html>.
11. Martin J. et Crowley J. An Apperance-Based Approach to Gesture-Recognition", 9th International Conference on Image Analysis and Processing, Florence, Italie, Sept. 1997.
12. Morikawa O. et Maesako T., HyperMirror : Toward Pleasant-to-use Video Mediated Communication System. Actes de la conférence CSCW'98, Seattle, Washington USA. pp.149-158.

13. Ott M., Lewis J. et Cox I., Teleconferencing eye contact using a virtual camera. Actes de la conférence CHI'93, Amsterdam. Netherlands, 24-29 avril 1993. pp. 109-110.
14. Richardson T., Stafford-Fraser Q., Weatherall J., Wood K., Harter A., McLachlan C. et Webster P., VNC Virtual Network Computing. Disponible à l'adresse <http://www.uk.research.att.com/vnc/>.
15. Shibata M., Proposal or Desk Top Program Production. Proc. 134<sup>th</sup> SMPTE Technical Conf. 1992, pp. 10-13.
16. Thevenin D., Bérard F., Coutaz J., Capture des inscriptions d'une surface plane par vision active. Soumis à la conférence IHM'99.
17. Wellner, P. *Interacting with paper on the DigitalDesk*, Communication of the ACM n°7, Juillet 1993, Publication ACM, pp. 87-96.