

# SUIVI DU DOIGT EN VISION PAR ORDINATEUR : APPLICATION AU BUREAU NUMÉRIQUE

François Bérard\*, Joëlle Coutaz\*, James L. Crowley\*\*

\* Laboratoire de Génie Informatique-Institut IMAG  
BP 53, 38041 Grenoble Cedex 9, tel. 76 51 48 54  
E-mail : Francois.Berard@imag.fr, Joelle.Coutaz@imag.fr

\*\* Laboratoire d'Informatique Fondamentale et d'Intelligence  
Artificielle-Institut IMAG  
46 avenue Félix Viallet, 38031 Grenoble Cedex, tel. 76 57 46 55  
E-mail : jlc@imag.fr

## THEME DES JOURNEES : NOUVELLES INTERFACES

**Résumé :** Cet article présente les résultats de notre réflexion sur l'usage de la vision par ordinateur appliquée au cas du bureau numérique. Nous restreignons l'étude au suivi d'objet de pointage tel le doigt et nous identifions les problèmes posés et les facteurs requis. Nous comparons deux techniques de vision par ordinateur pertinentes : les contours actifs et la corrélation d'image que nous avons appliquées à la mise en œuvre de FingerPaint, une application pilote réalisée sur Macintosh Quadra. FingerPaint permet de dessiner sur une feuille de papier réelle au moyen d'un doigt ou d'un objet physique que l'on aurait plongé dans de la peinture électronique.

**Mots-clés :** suivi du doigt, vision par ordinateur, bureau numérique.

### 1. Introduction

Jusqu'ici, la vision par ordinateur a considéré la robotique comme son domaine de prédilection. Elle trouve maintenant un nouveau champ d'application en communication homme-machine ouvrant, nous en sommes convaincus, des perspectives inattendues. Krueger et son videoplace [KRUEGER 91] est un pionnier en la matière. Aujourd'hui, la tendance se confirme : à CMU, le système DigitEyes offre un modèle 3D temps réel de la main [REHG 93] ; par la technique des "eigenfaces" [TURK 91], il est possible de reconnaître un visage avec une fiabilité raisonnable [BERARD 93] ; de même, la vision par ordinateur est utilisée pour la reconnaissance de l'expression faciale [ESSA 94] ou pour reconstituer le contact visuel dans les activités de téléconférence [OTT 93].

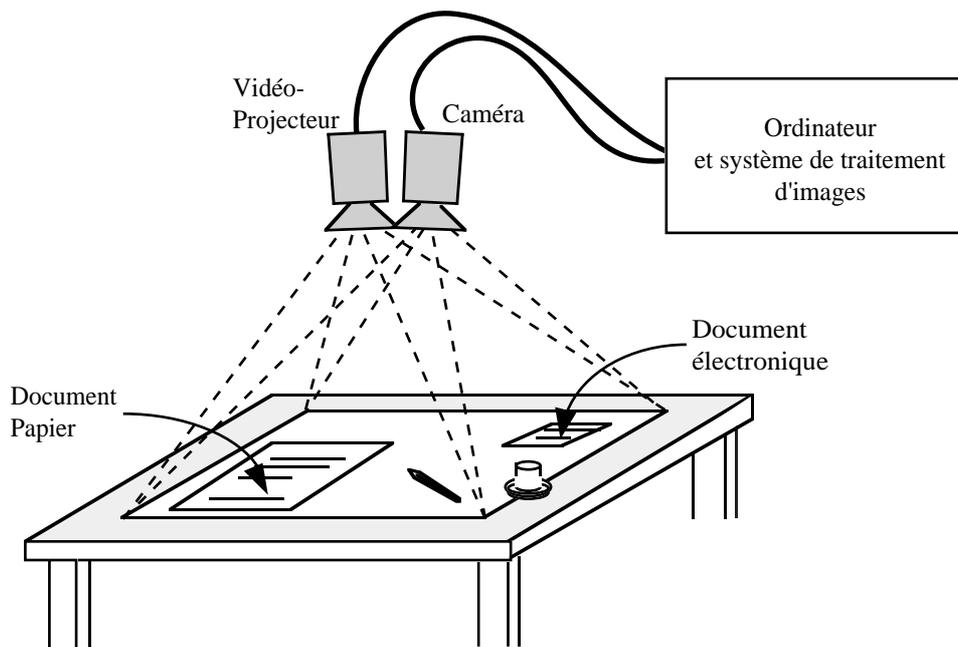
Cet article présente les résultats de notre réflexion sur l'usage de la vision par ordinateur appliquée au cas du bureau numérique. Après avoir présenté ce concept au paragraphe 2, nous orientons en 3 notre analyse vers les problèmes de mise en œuvre. En particulier, nous retiendrons comme sujet d'étude le suivi du doigt et présentons en 4 les requis que les techniques de vision doivent satisfaire. En 5, nous présentons les résultats que nous avons obtenus avec la technique des contours actifs et la corrélation d'image. Nous avons évalué ces deux approches au moyen de FingerPaint, une application pilote réalisée sur Macintosh Quadra 840 équipé d'une caméra. FingerPaint permet de dessiner sur une feuille de papier réelle au moyen d'un doigt ou d'un objet physique que l'on aurait plongé dans de la peinture électronique. Le détail de ces travaux peut être consulté dans [BERARD 94].

### 2. Le bureau numérique

Le concept de bureau numérique (ou "digitaldesk") nous vient de RXRC (ex EuroPARC) [WELLNER 93]. Il se veut une réponse aux ambitions irréalistes de la bureautique qui clamait dès le départ, que l'objectif "zéro-papier" serait atteint en quelques années. La solution du tout électronique était un choix extrême que la pratique a rapidement mis en

cause : de source sérieuse, on note que le travail sur papier dans les bureaux a été multiplié par 6 depuis 1970 et qu'il progresse chaque année de 20% environ [WELLNER 93]. C'est pourquoi le concept de bureau numérique, qui prône une coopération harmonieuse entre documents papier et numériques, semble prometteur.

Comme le montre la figure 1, les services informatiques sont intégrés au bureau réel via une caméra et un projecteur fixés verticalement à la surface de travail. Les actions de l'utilisateur sont fournies à la machine par l'intermédiaire de la caméra alors que la projection de l'écran sur le bureau procure l'équivalent d'une interface graphique, avec en plus la possibilité de fusionner les objets réels (ceux qui sont posés sur la table) avec les objets virtuels (ceux qui sont projetés sur la table).



**Figure 1** : Configuration matérielle du bureau numérique.

Cette installation satisfait trois caractéristiques essentielles : 1) une image est projetée sur le plan de travail mais aussi sur les documents qui y sont entreposés ; 2) l'utilisateur dispose du clavier et de la souris mais conserve ses outils du monde réel : stylos, gomme, doigts, etc. 3) les documents posés sur le bureau peuvent être lus par l'utilisateur mais aussi par la machine.

La combinaison de ces caractéristiques montre que le bureau numérique est un exemple de réalité augmentée au sens de notre taxonomie présentée dans [BERARD 94] et que la combinaison du réel et du virtuel est présente aussi bien en entrée qu'en sortie. De fait, le bureau numérique ouvre un vaste champ d'applications. Prenons par exemple, une feuille physique sur laquelle seraient inscrites des valeurs numériques. L'utilisateur pourrait désigner du doigt les valeurs utiles et prononcer "additionner". Le résultat serait projeté de manière électronique sur le papier. Le papier réel gagne ici les capacités d'une calculatrice.

Les problèmes que soulève la réalisation du bureau numérique au moyen de la vision par ordinateur sont maintenant présentés

### 3. Les problèmes

Nous avons relevé trois classes de difficultés : le calibrage, la correction d'image et la reconnaissance d'objet. Nous évoquerons rapidement les deux premiers et insisterons au paragraphe suivant sur la reconnaissance d'objet qui sert de fondement à la réalisation du suivi de doigt.

*Calibrage.* Le problème du calibrage tient à deux raisons : les objectifs de la caméra et du projecteur ont deux points de vue différents même si ces deux dispositifs sont rapprochés. Il en résulte deux images de forme distincte. La seconde difficulté tient aux changements de configuration des dispositifs dûs aux vibrations ou à un souffle d'air sur le papier. En conséquence, il convient d'appliquer régulièrement un algorithme d'auto-calibrage qui soit rapide et qui ne nécessite pas l'intervention explicite de l'utilisateur.

La solution adoptée pour le digitaldesk consiste à projeter quatre croix sur le bureau, à les localiser sur l'image captée par la caméra, puis à localiser une cinquième croix pour vérifier les prédictions du système. Ces opérations s'appuient sur un système d'équations que les auteurs ne justifient pas. Dans [BERARD 94], on trouvera une formulation précise de la fonction de calibrage au moyen d'une transformation projective pour laquelle nous proposons une implémentation efficace.

*Correction d'image.* Pour exploiter les informations contenues dans les documents papier, il faut pouvoir disposer d'un logiciel capable d'interpréter le contenu des images. L'inclinaison des images due à l'angle que forment le projecteur et la caméra par rapport à la verticale doit être corrigée afin de reconstituer une image complète du document. Cette étude dépasse le cadre de cet article.

*Reconnaissance d'objet.* Le digitaldesk implique que le système de vision soit capable de reconnaître le type d'outil emprunté par l'utilisateur. Le problème de la reconnaissance en vision est un vaste domaine de recherche mais les besoins sont ici cernés par le nombre d'objets susceptibles d'être identifiés : l'utilisateur aura l'occasion de se servir de ses doigts, de stylos ou de gommes. Le processus de reconnaissance peut être encore simplifié si l'on accepte que l'utilisateur se serve toujours des mêmes outils et qu'une phase d'apprentissage soit demandée à l'initialisation du système.

#### **4. Le suivi d'objet**

Le suivi d'objet fait parti des thèmes essentiels de la recherche en vision par ordinateur. Il est le sujet d'une multitude de publications majoritairement centrées sur les problèmes canoniques : suivi en mono ([MEYER 92], [KOLLER 92]) ou stéréo vision ([REHG 93]) ; suivi d'un motif ([MEYER 92]), d'un modèle géométrique ([HARRIS 92], [KOLLER 92], [REHG 93]), ou d'un contour actif ([KASS 87], [TERZOPOULOS 92], [WILLIAMS 90], [UEDEA 92]). Certaines recherches portent aussi sur des applications précises du suivi comme le contrôle du trafic automobile ([KOLLER 92], [KOLLER 93]), la reconnaissance des gestes voire le suivi des mains ([REHG 93]). Cependant, à l'heure actuelle, il n'existe aucune étude sur l'utilisation du suivi pour le bureau numérique.

*Objectif.* Notre objectif est de connaître en temps réel la position sur le bureau d'un objet de pointage, essentiellement un doigt, un stylo ou une gomme. Nous cherchons à conserver l'environnement naturel de l'utilisateur : l'emploi de gants ou de marquages spécifiques sur les objets à suivre est proscrit. Notre seule source d'information est un flot d'images en 256 niveaux de gris provenant d'une caméra.

Nous avons retenu le suivi d'objet dans un espace plan : dans le cas du bureau numérique, l'espace de travail est essentiellement 2D. La détection du contact et de sa levée peut se faire au moyen d'un microphone intégré au bureau. C'est la solution adoptée pour le digitaldesk. Nous avons vérifié que ce procédé est efficace en l'améliorant par l'utilisation d'une caisse de résonance.

Ceci posé, les facteurs directeurs de la solution sont la vitesse, la robustesse, la souplesse, la reprise sur erreur.

*Vitesse.* La loi de Fitts [CARD 83], qui permet de prédire le temps mis par un sujet humain pour placer sa main sur une cible, s'applique directement à notre étude. Dans le cas du bureau numérique, la diagonale capturée par la caméra mesure environ 70cm et les objets à manipuler (des icônes) ont une taille de l'ordre de 2cm. Ces données nous

permettent d'estimer la vitesse de déplacement de la main au moyen de la loi de Fitts, soit : 1,8 m/s.

Nous avons pratiqué une évaluation expérimentale de ce résultat en filmant un va-et-vient rapide de la main sur le bureau. On trouvera dans [BERARD 94] le détail de cette expérience qui nous conduit aux conclusions suivantes : dans le cas du mouvement de la main sur le bureau, la loi de Fitts a tendance à surestimer la vitesse du doigt puisque lors de l'expérience, le geste a été effectué de manière très rapide et sans la contrainte d'atteindre une cible. Or la vitesse maximale atteinte est de 1,39 m/s, là où la loi de Fitts prédit des vitesses de l'ordre de 1,8 m/s pour un individu moyen et de l'ordre de 4,2 m/s pour un individu rapide.

*Robustesse.* Nous traduisons le critère de robustesse d'un algorithme de suivi d'objet par sa tolérance aux rotations, sa résistance aux variations de l'éclairage et sa tolérance à l'encombrement du bureau.

La technique doit tolérer une totale liberté de rotation de l'objet suivi dans le plan du bureau mais aussi hors de ce plan. Nous ne voulons pas imposer d'environnement lumineux spécifique. Par conséquent, les variations lumineuses naturelles doivent être prises en compte. L'objet à suivre peut se déplacer sur un bureau nu ou bien encombré. Le suivi doit donc autoriser l'utilisation d'un fond bruité.

*Souplesse.* La technique de suivi à retenir doit pouvoir suivre la forme de l'extrémité d'un doigt, d'un stylo et d'une gomme. La structure de ces trois formes étant assez proche, ce critère devrait être satisfait assez aisément. Cependant il faut prévoir l'ouverture de l'algorithme de suivi au cas de nouveaux objets (qui peuvent avoir un rôle différent du pointeur). Par exemple, on voudra connaître la position d'une feuille de papier pour pouvoir y projeter un contenu virtuel quand l'utilisateur la déplace. Il convient donc de considérer la capacité d'adaptation d'une technique de suivi à une variété de formes.

*Traitement des échecs.* Aucun des algorithmes de suivi actuels ne peut garantir un suivi parfaitement fiable. On sera donc amené à comparer leurs capacités à détecter la perte de la cible. S'il y a possibilité de détection précoce, on examinera l'aptitude à entrer dans une procédure d'exception qui tente de rétablir une situation fiable avant la perte. Si le suivi n'a pas de moyen de prévenir une perte imminente, il est néanmoins nécessaire de détecter l'échec afin de réagir soit en entrant de nouveau dans une phase d'initialisation si celle-ci est automatisée, soit en avertissant l'utilisateur (par exemple disparition du curseur, ou réapparition de la boîte d'initialisation du suivi) afin que celui-ci réinitialise lui-même le suivi. On trouvera dans [BERARD 94] l'analyse des problèmes posés par l'initialisation

## **5. Etude comparative de deux techniques de suivi d'objet**

Nous avons choisi de réaliser deux techniques différentes de suivi d'objet en vision par ordinateur afin de les comparer dans le cadre de leur utilisation pour le bureau numérique : le suivi par corrélation et le suivi à l'aide d'un contour déformable.

Le *suivi par corrélation*, déjà présenté dans [MORAVEC 80], s'effectue en recherchant la cible du suivi autour de sa dernière position connue. Une image de la cible (le "motif") est mémorisée, puis comparée à des zones de même taille dans l'image de la caméra. La zone qui "ressemble" le plus au motif est choisie comme nouvelle position de l'objet suivi. Il existe plusieurs formules de calcul de la ressemblance entre deux images ; dans tous les cas la mesure est basée sur les comparaisons des valeurs de pixels de même position dans les deux images.

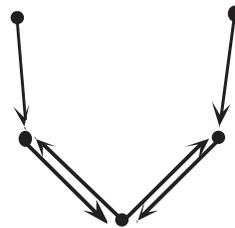
Nous avons identifié trois caractéristiques qui font du suivi par corrélation un bon candidat pour une utilisation dans la réalisation du bureau numérique. Tout d'abord la simplicité de ses algorithmes est à l'origine de sa rapidité : notre réalisation sur système Macintosh Quadra 840 autorise une vitesse de déplacement de la cible de 0,6 m/s. Ensuite l'adaptation du suivi à une cible nouvelle est immédiate : il suffit pour cela de "prendre une photo" de l'objet à suivre. Enfin il est possible d'avoir une information sur l'état de

l'algorithme grâce à la mesure de ressemblance entre motif et image : lorsqu'une chute de cette valeur est observée, il est fort probable que la cible ait été perdue. Il est alors possible de réinitialiser l'algorithme.

Cependant, le suivi par corrélation présente deux limitations directement issues de son principe : en premier lieu la mesure de ressemblance est fortement influencée par les variations d'intensité lumineuse puisqu'elle est calculée sur la valeur des pixels qui représentent leur luminosité. Un autre problème apparaît lorsque l'objet suivi effectue une rotation : comme celle-ci n'est pas prise en compte dans le calcul de la mesure de ressemblance, les pixels comparés dans le motif et l'image ne se correspondent plus, ce qui entraîne une perte de la cible.

C'est pourquoi nous nous sommes intéressés à une technique d'un niveau d'abstraction plus élevé puisqu'elle se base sur les contours plutôt que sur les pixels : il s'agit du *suivi à l'aide d'un contour actif (snake)* qui fut présenté pour la première fois dans [KASS 87]. Son principe est de créer un contour (le snake) n'ayant aucune existence physique mais dont le comportement est modélisé de façon à ce qu'il suive sa cible. Ceci est réalisé en format une attirance du snake aux forts gradients (contours) de l'image. Ses déplacements vers le gradient sont toutefois limités par une certaine élasticité et rigidité qui modélisent sa capacité de déformation en élongation et courbure.

On trouvera dans [BERARD 94] les détails de la réalisation du comportement du snake ainsi que le raisonnement qui nous a amené à définir un modèle spécifique du snake pour le suivi du doigt. Ce modèle est basé sur l'utilisation d'une "forme préférée" (voire figure 2) que prend le snake lorsqu'il n'est soumis à aucune force externe.



**Figure 2 :** Forme préférée du snake : les points définissent la forme du snake, les flèches représentent la modélisation de son élongation.

La programmation du suivi à l'aide d'un snake est plus complexe que celle du suivi par corrélation. La vitesse de déplacement maximale de la cible est donc réduite (sur le même système nous obtenons 0.08 m/s). De plus l'adaptation du suivi à de nouvelles formes est rendue difficile par l'utilisation d'un modèle spécifique du snake. Cependant on constate un réel gain en robustesse par rapport au suivi par corrélation, notamment au niveau des tolérances aux variations lumineuses et aux rotations de la cible. C'est pourquoi nous préconisons l'utilisation d'un snake comme technique de suivi du pointeur pour le bureau numérique.

## 6. Conclusion

Dans cet article, la présentation du concept du bureau numérique nous a permis de mettre en évidence l'importance du rôle de la vision par ordinateur dans sa réalisation. Parmi les problèmes soulevés dans ce domaine nous retenons le plus crucial : le suivi d'objet. Ce sujet est analysé dans le contexte précis de la réalisation du bureau numérique. Nous donnons alors les résultats de notre étude comparative menée sur deux techniques de suivi d'objet en vision par ordinateur : nous mettons en évidence que si le suivi par corrélation est rapide, son manque de robustesse nous pousse à préconiser l'utilisation de la technique des snakes, améliorée grâce à l'utilisation d'un modèle spécifique. D'un point de vue plus général, cette étude nous a permis de constater que la réalisation du bureau numérique n'est pas utopique, et nous pousse à poursuivre nos recherches, en particulier en étudiant la possibilité d'un suivi multiple pour une interaction à plusieurs doigts.

## 7. Références bibliographiques

- [BERARD 93] F. Bérard. "Vision et IHM : Application à l'analyse des visages". Rapport de Magistère, Université Joseph Fourier, Grenoble, 1993.
- [BERARD 94] F. Bérard. "Vision par Ordinateur pour la Réalité Augmentée : Application au Bureau Numérique". Rapport de DEA, E.N.S.I.M.A.G., Grenoble, 1994.
- [CARD 83] S. K. Card, T. P. Moran, A. Newell. "The Psychology of Human-Computer Interaction". Lawrence Erlbaum Associates, 1983.
- [ESSA 94] I. A. Essa, A. Pentland. "A vision system for observing and extracting facial action parameters". M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 247, Juin 1994.
- [KASS 87] M. Kass, A. Witkin, D. Terzopoulos. "Snakes : Active Contour Models". Proc. 1st International Conf. on Computer Vision, pp. 259-268, 1987.
- [KOLLER 92] D. Koller, K. Daniilidis, T. Thórhallson, H. H. Nagel. "Model-Based Object Tracking in Traffic Scenes". Proc. 2nd European Conf. on Computer Vision, pp. 437-452, 1992.
- [KRUEGER 91] M. W. Krueger. "Artificial Reality II". Addison Wesley, 1991.
- [MEYER 92] F. Meyer, P. Bouthemy. "Region-Based Tracking in an Image Sequence". Proc. 2nd European Conf. on Computer Vision, pp 476-484, 1992.
- [MORAVEC 80] H. P. Moravec. "Obstacle Avoidance and Navigation in the Real World by a Seing Robot Rover". Phd Thesis, Stanford University, 1980.
- [OTT 93] M. Ott, J. Lewis, I. Cox. "Teleconferencing
- [REHG 93] J. M. Rehg, T. Kanade. "DigitEyes : Vision-Based Human Hand Tracking". Carnegie Mellon University Technical Report CMU-CS-93-220, December 1993.
- [TERZOPOULOS 92] D. Terzopoulos, R. Szeliski. "Tracking with Kalman Snakes". in "Active Vision", The MIT Press, 1992.
- [TURK 91] M. A. Turk. "Interactive Time-Vision : Face Recognition as a Visual Behavior". Phd thesis, Massachusetts Institute of Technology, 1991.
- [WELLNER 93] P. Wellner. "Interacting with paper on the DigitalDesk". Communications of the ACM, Vol.36 No.7, July 1993.
- [WILLIAMS 90] D. J. Williams, M. Shah. "A Fast Algorithm for Active Contours". Proc. 1st European Conf. on Computer Vision, pp 592-595, 1990.