Le Tableau Magique

RAPPORT DE PROJET ENSIMAG 1998-99

RÉALISÉ PAR
Sébastien Annedouche,
Benoît Loup,
Michel Prodhomme

SOUS LA RESPONSABILITÉ DE
François Bérard

MAÎTRE D'OEUVRE
Joëlle Coutaz

TUTEUR
Jim Crowley

Ce document est disponible sous forme électronique aux adresses suivantes :

http://iihm.imag.fr/publs/1999/TR199902_MagicBoard.ps.gz

http://iihm.imag.fr/publs/1999/TR199902_MagicBoard.pdf

TABLE DES MATIÈRES

Table des matières ii

Remerciements 1

Avant-Propos 3
- But du document 3
- Organisation des participants 3

Introduction : Le sujet 5

Réalité Augmentée : état de l’art 7
- Réalité Augmentée 7
- Le canal gestuel 8
- Vision par ordinateur et suivi de geste 8
- Contexte concurrentiel : autres tableaux 9

Les objectifs prévus 11
- Scénarios 11
- Les services attendus 11

Réalisation 15
- Environnement matériel et logiciel 15
- Conception globale 15
- Suivi de doigt 16
  - Principe 16
  - Mise en oeuvre 17
  - Problèmes rencontrés 21
  - Évaluation des performances 21
    - Étude de la fréquence de fonctionnement 22
    - Étude de la vitesse maximale de suivi 23
- Contrôle de la caméra 24
  - Principe 25
  - Mise en oeuvre 25
    - Le calibrage de la caméra 26
    - Le calibrage des positions 26
    - Le calibrage du zoom 27
    - Utilisation des tables créées lors du calibrage 28
  - Problèmes rencontrés 30
  - Étude des performances 31
- Nettoyage d’images 31
  - Principe 31
  - Mise en oeuvre 32
    - Technique du «seuillage adaptatif» de P. Wellner 32

            Performances..33
                Temps de calcul 36
        Calibrage caméra 37
            Principe 37
            Mise en oeuvre..38
            Etude des performances 42
        Le mosaïcing...43
            Principe 43
            Mise en oeuvre..43
            Etude des performances 43
                Taille de l'image haute résolution... 43

Bilan 47
        Objectifs atteints 47
        Contribution au projet..47
        Planning suivi 48

Conclusion : leçons et perspectives 51
        Leçons de l'expérience 51
        Perspectives...52

Annexes

La différence d'images 55

Analyse fonctionnelle 57

Le noyau fonctionnel 61
        Gestion du contenu du tableau virtuel 61
        Réseau...62
            Principe 62
            Mise en oeuvre..63
            Performances..63

L'interface Homme-Machine 65
        Les differents affichages sur le tableau magique 65
            Les boutons 65
            Les curseurs 66
        L'interface de démarrage 66

Le modele de couleurs 69
        Principe 69
        Mise en oeuvre..70
            Construction d'un «modèle de couleurs» 70
            Application d'un modèle de couleur à une image. 71
            Performances..72

Définitions, abréviations et terminologie 73

Bibliographie 75

REMERCIEMENTS

Nous tenons à remercier,

Joëlle COUTAZ, pour nous avoir acceuillis dans son équipe, pour les moyens qu'elle a mis à notre disposition et pour le temps qu'elle a consacré à l'aboutissement de ce projet;

François BÉRARD, pour son aide précieuse de tous les instants sur tous les aspects du projet;

Enfin, tous les membres de l'équipe IIHM pour la bonne ambiance qu'ils ont contribuée à créer.

AVANT-PROPOS

BUT DU DOCUMENT

Ce document constitue le rapport de notre projet de fin d'études. Il en présente le sujet et le contexte, il précise ensuite les objectifs fixés et leur réalisation, puis résume les enseignements de cette expérience.

ORGANISATION DES PARTICIPANTS

Groupe de développement :

- Sébastien ANNEDOUCHE
- Benoît LOUP
- Michel PRODHOMME

Maître d'œuvre :

- Joëlle COUTAZ (CLIPS-IMAG - Equipe IIHM, Ingénierie de l'interaction Homme-Machine) encadre le projet et apporte une aide technique.

Consultant :

- François BERARD (CLIPS-IMAG - Equipe IIHM) apporte son aide concernant le domaine de la vision par ordinateur appliquée à l'Interaction Homme-Machine.

Tuteur :

- James L. CROWLEY (GRAVIR-IMAG - Equipe PRIMA) se charge des procédures ENSIMAG.

1

INTRODUCTION : LE SUJET

L’objet de ce projet est le développement de techniques de vision par ordinateur capables de répondre aux requis ergonomiques d’une classe de systèmes de Réalité Augmentée : le Tableau Magique. La Réalité Augmentée (RA) propose, au contraire de la Réalité Virtuelle, de conserver les objets qui nous sont familiers mais ces objets sont améliorés de capacités de traitement de l’information. Les systèmes de RA sont peu nombreux, mal cernés et fonctionnent dans des conditions expérimentales très contraintes. Ce projet a pour but de réaliser une maquette de tableau magique offrant davantage de robustesse et de souplesse d’utilisation que les systèmes de RA similaires.

Techniquement, le tableau magique est constitué d’un tableau blanc réel manipulable de manière ordinaire : des inscriptions peuvent se faire et se défaire avec les stylos feutre et l’effaceur usuels. Mais le tableau est amélioré de deux dispositifs :

- un projecteur relié à la sortie vidéo d’un ordinateur affiche sur le tableau les retours d’information du système;
- une caméra observe et interprète les gestes de l’utilisateur.

Dès lors, le tableau est amplifié de services informatiques tels que couper, coller, déformer, imprimer, qui peuvent s’exprimer avec les mains. Dans ce projet, le tableau magique est envisagé pour une situation particulière d’usage : le “brainstorming”, activité de réflexion collective conduisant à la production d’idées et à leur organisation. Un tableau magique, tel que nous l’entendons, doit offrir les services suivants :

- L’utilisateur peut dessiner au tableau à l’aide de feutres usuels.
- L’utilisateur peut sélectionner une zone de tableau, déplacer le contenu d’une zone sélectionnée, en modifier la taille, l’iconifier et l’effacer. Toutes ces actions doivent se faire au moyen des doigts et sont interprétées par vision par ordinateur.
- Le contenu complet ou partiel du tableau peut être enregistré dans un format numérique propice au stockage sous forme de fichier, à l’envoi vers une imprimante ou dans une page Web.
- Le tableau peut être utilisé à distance par plusieurs utilisateurs reliés par vidéoconférence.

Dans ce qui suit, nous présentons le contexte du projet, puis nous abordons en détail les objectifs que nous nous sommes fixés ainsi que leur réalisation, avant de tirer les leçons de cette expérience.

RÉALITÉ AUGMENTÉE : ÉTAT DE L'ART

2

Ce projet a pour ambition d'aboutir à la réalisation d'une maquette de système de réalité augmentée. Ce projet a également pour vocation d'explorer le canal gestuel en tant que moyen d'interaction entre l'utilisateur et le système informatique. Cette interaction s'appuie sur la vision par ordinateur. Nous reprenons ces trois points (Réalité Augmentée, canal gestuel et vision) aux paragraphes suivants puis nous justifions notre offre au regard de la concurrence.

2.1 RÉALITÉ AUGMENTÉE

La réalité augmentée est une approche centrée sur l'interaction Homme-Environnement qui vise l'intégration des objets du monde réel dans l'acte interactionnel. Elle cherche à étendre le spectre des utilisations des objets usuels en fournissant à l'utilisateur des fonctions issues d'un système de traitement de l'information. Ces fonctions n'ont pas pour but de remplacer la réalité; elles visent à l'amplifier par l'observation de l'environnement et notamment par l'analyse des actions de l'utilisateur sur cet environnement. Un système de réalité augmentée doit donc receuillir des observations sur l'environnnement réel, les traiter et les restituer dans le mode réel.

L'un des précurseurs des systèmes de réalité augmentée est le Digital Desk (Bureau Virtuel) développé par le centre de recherche de Xerox (ex EuroPARC). Le bureau digital est un système de réalité augmentée appliqué au bureau de travail. Celui-ci est amplifié par un système d'information qui receuille l'information depuis l'environnement par les canaux audio et vidéo, effectue des calculs et projette des informations de retour vers le monde réel par le biais du canal vidéo. Ce système permet la manipulation courante des documents papier; grâce à la réalité augmentée, il permet aussi de travailler sur des documents ou des parties de documents numériques qui conservent les mêmes règles d'utilisation que les documents papier traditionnels (déplacement du document, correction ou effacement de certaines zones ...). Ces manipulations se pratiquent avec les doigts, d'où le nom de bureau "digital". Les parties du document numérique et physique sont intimement liées en synergie sur la surface de travail du bureau physique. De nouvelles fonctions sont rendues possibles par l'intervention du numérique, comme le copier-coller d'une zone sélectionnée.

2.2 LE CANAL GESTUEL

Dans le cadre de l'Interaction Homme-Machine, le canal gestuel est encore sous-exploité. Il inclut les gestes de la main, des doigts, des membres ou du corps entier pour commander une machine.

On distingue trois types de geste suivant leur finalité principale [Cadoz 94]:

- Le geste sémiotique : il sert à faire connaître par le moyen de signes (ex : les gestes accompagnant la parole).
- Le geste épistémique : c'est un moyen de connaître (ex : la combinaison de la perception cutanée avec le mouvement des mains et des doigts qui permet de percevoir la forme ou la taille d'un objet).
- Le geste ergotique : il permet d'agir physiquement sur les objets pour les saisir, les transporter, les modeler ...

Dans le cadre de ce projet, les gestes envisagés sont de nature sémiotique : ils servent à exprimer des ordres au tableau magique.

2.3 VISION PAR ORDINATEUR ET SUIVI DE GESTE

La recherche en suivi des mains et reconnaissance de geste est un phénomène récent en vision par ordinateur. Elle se divise en deux écoles de pensée et ceci quelle que soit la nature du geste observé : l'une s'appuie sur les modèles géométriques et cinématiques, l'autre sur les représentations de l'apparence.

Les modèles géométriques trouvent leurs origines dans la synthèse d'images [Thalman 90]. Les efforts d'adaptation aux besoins de la vision s'appuient sur des techniques de suivi et d'estimation au moyen du filtre de Kalman [Koch 93, Crowley 97a]. Les modèles géométriques de la main sont fondés sur des primitives de volume [O'Rourke 80] ou des splines [Dowton 91, Gravila 95]. Les besoins lourds en calcul et les problèmes de complexité rendent cette approche difficile à adapter au suivi en temps réel comme l'exige le tableau magique.

La vision par apparence, à laquelle nous sommes associés, offre une nouvelle voie de réflexion. Elle s'appuie sur l'apparence, c'est-à-dire sur les manifestations visuelles possibles de quelque chose. Par exemple, l'apparence visuelle d'un objet physique est la composition de toutes les images, prises de tous les points de vue possibles, selon tous les éclairages possibles.

Les méthodes d'interprétation des gestes par apparence mesurent les paramètres du geste en utilisant des opérations définies sur le signal 2D [Maggioni 95, Cipolla 96]. Certaines approches emploient les contours déformables tels les “snakes”. D'autres déterminent un ensemble de déformations types au moyen d'une analyse en composantes principales [Cootes 93, Cootes 95, Heap 96]. L'alternative est l'application d'une analyse en composantes principales directement sur l'image de la main [Black 96, Martin 94]. D'autres encore [Wren 96] utilisent la couleur de la peau modélisée par une mixture de Gaussiennes. Des solutions efficaces mais simplistes comme la technique de différence d'images employée pour le bureau digital n'est pas envisageable en situtation réelle : sensibilité au changement d'éclairage, et surtout orientation figée du doigt.

Dans ce projet, nous utilisons des techniques relevant de la vision par apparence capables de répondre, au vu de l'expérience, aux requis imposés par l'usage interactif en situation. Pour ce faire, nous nous appuyons sur une plate-forme d'outils de vision pour l'interaction homme-machine développés au laboratoire CLIPS-IMAG en collaboration étroite avec l'équipe PRIMA du laboratoire GRAVIR-IMAG.

2.4 CONTEXTE CONCURRENTIEL : AUTRES TABLEAUX

Le LiveBoard de Xerox et ses assimilés, peuvent être vus comme une alternative au tableau magique : si ces dispositifs offrent des services d'édition usuels, l'interface personne-système est conventionnelle, l'épaisseur de l'écran et son parallaxe entravent parfois la bonne coordination sensori-motrice, l'usage de marqeurs spéciaux est imposé et le prix d'achat est prohibitif. L'expérience montre qu'ils sont en fait, peu utilisés. Le succès du ZombieBoard de XeroxPARC [Saund 97], de même que la sortie du Digital Flipchart [PD 97] et du SoftBoard sont révélateurs du bien fondé de notre choix.

Le ZombieBoard, dans sa version actuelle, est un simple photocopieur semblable aux copieurs analogiques de tableau. Par des marques de forme conventionnelles dessinées au feutre, l'utilisateur balise une zone d'intérêt que le système imprimera sur support papier. Le système parcourt la surface au moyen d'une caméra montée sur une tourelle mobile qui prend des images successives de la surface puis construit, par la technique du "mosaïcing", une image unique de haute résolution et nettoyée des impuretés d'un tableau mal effacé. Mais, contrairement au Digital Flipchart, ce système n'inclut pas de manipulation interactive des inscriptions. Le Digital Flipchart projette sur le tableau des boutons numériques sélectionnables au moyen du doigt. Il s'agit là d'une forme d'interactivité primitive. Le Holowall est plus riche dans son offre interactive mais utilise un tableau semi-transparent avec projection des images et observation des main par l'arrière [Rekimoto 97]. Ce montage simplifie le traitement des occultations des mains de même que la détection du contact des mains sur le tableau. Mais il ne s'agit plus du tableau familier.

Avec le tableau magique, nous visons une situation d'usage écologique : pas de fil aux mains (comme l'impose un gant numérique) mais un tableau et des marqueurs ordinaires qui restent utilisables même en cas de panne du système. Le SoftBoard répond presque à nos requis, mais fondé sur la technologie laser, il comprend un tableau en porcelaine et des marqueurs habillés d'une surface réfléchissante et d'un code barre identifiant la couleur. Le SmartBoard de Smart Technologies (http://www.smarttech.com) utilise une surface tactile et permet l'usage de feutres ordinaires. Le Mimio, qui sera disponible en Europe en juillet 99, est une approche originale : il comprend un recepteur d'ultrason, sorte de barette de 40X10 cm que l'on fixe sur un tableau ordinaire au moyen de deux ventouses. Les feutres sont insérés dans un gousset qui émet des ultrasons. La vision par ordinateur est une autre approche possible à ces conditions d'usage. C'est celle que nous retenons.

LES OBJECTIFS PRÉVUS

3

Les objectifs prévus comprennent la description des scénarios de recette et la liste des services attendus pour réaliser ces scénarios. En accord avec le maître d'oeuvre, les objectifs ont été classés suivant trois niveaux de priorité: obligation, souhait et éventualité.

3.1 SCÉNARIOS

- Dans le cadre du réaménagement des locaux d'une entreprise, les différents protagonistes souhaitent se mettre d'accord sur les emplacements des nouveaux bureaux au cours d'une réunion à laquelle toutes les parties concernées sont présentes. Au cours de cette réunion, les documents de base incluent les plans papiers décrivant l'architecture des bureaux. Les participants peuvent visualiser les plans sur le tableau blanc, y écrire des annotations au stylo effaçable, mais aussi des annotations projetées qu'ils déplacent, suppriment ou copient à leur guise. Ils peuvent conserver leur travail sous la forme d'un ou de plusieurs fichiers qui contiennent le contenu complet ou partiel du tableau magique au moment de la sauvegarde.
- Le concepteur d'un logiciel a prévu de tenir une réunion avec un responsable client pour se mettre d'accord sur les spécifications du logiciel qu'il doit livrer. Les deux personnes sont sur deux sites distants et utilisent chacun un tableau magique. Ils peuvent tour à tour écrire sur le tableau, ajouter des annotations à ce que l'autre personne a écrit ou dessiné. Un contact de type vidéoconférence est assuré pour qu'un utilisateur puisse désigner les formes dessinées au tableau, ainsi qu'un contact sonore (de type téléphone).

3.2 LES SERVICES ATTENDUS

Nous classons dans le tableau ci-dessous le niveau de priorité de chacun des services que doit fournir la maquette du tableau magique :

Priorité	Libellé du service
Obligation	• Calibrage du système : Mise en correspondance des repères cartésiens de l'acquisition et de la restitution vidéo. • Capture/Numérisation : Nettoyage d'une image (une image est le résultat de la capture vidéo d'une partie du tableau) - Contrôle de l'orientation et du zoom de la caméra - Mosaïcing. • Suivi de doigt : Détection de main dans l'image - Suivi d'entité. • Reconnaissance de gestes : Reconnaissances de trajectoires temporelles (en particulier la pause). • Sélection de zones rectangulaires du tableau. • Copie, déplacement et collage d'une zone sélectionnée (ces fonctions sont réalisées à l'aide du doigt). • Impression, sauvegarde du contenu total ou partiel du tableau, chargement d'un document (image issue de la sauvegarde d'une session précédente ou données externes). • Ouverture et fermeture de connexion entre deux systèmes distants. • Transmission d'informations entre deux systèmes (images, contenu du tableau, matrices de mise en correspondance des repères). • Gestion des tours de parole (mode lecteur/rédacteur).
Souhait	• Acquisition vidéo sur PC • Sélection d'une zone non rectangulaire du tableau (la forme sélectionnée est celle indiquée par la trajectoire du doigt). • Zoom : agrandissement ou réduction d'une zone sélectionnée (par l'une des deux méthodes citées précédemment). • Gestion d'un espace virtuel plus grand que la taille du tableau physique.

Priorité	Libellé du service
Eventualité	• Transmission de données en mode Multicast (envoi d'images à destination de plusieurs systèmes distants). • Réalisation d'une boîte à outils graphiques (tracé de figures géométriques simples à l'aide de gestes de l'utilisateur). • Importation de formats d'information externes (ex : Word, PowerPoint, PostScript). • Reconnaissance de formes simples, reconnaissance d'écriture.

RÉALISATION

4

Dans ce chapitre, nous développons l'aspect réalisation du projet. Nous commencons par préciser l'environnement logiciel et matériel dans lequel nous avons évolué. Nous aborderons ensuite la conception globale et la conception détaillée des composants essentiels de la maquette du Tableau Magique.

4.1 ENVIRONNEMENT MATÉRIEL ET LOGICIEL

Le tableau ci-dessous présente de manière synthétique les outils et matériels utilisés pour la réalisation du projet.

Développement du noyau fonctionnel	C++ et TCl/Tk version 8.0.4
Développement de l'IHM	Tcl/Tk version 8.0.4
Plate-forme matérielle	Silicon Graphics O2, processeur R10000 à 150 Mhz, système IRIX 6.5 PowerMacintosh, processeur PPC 604e à 350 Mhz, système MacOS 8.5
Caméra	Sony EVI-D30
Vidéoprojecteur	Liesegang DDV 820

Tous les tests de performances dont les résultats seront présentés dans la suite de ce chapitre ont été réalisés sur les deux stations de travail.

4.2 CONCEPTION GLOBALE

Le Tableau Magique a été décomposé en plusieurs unités de traitement (ou tâches). Le rôle de chacune de ces unités ainsi que leur fonctionnement global est présenté en annexe (cf. Annexe B : “Analyse fonctionnelle”), de même que leurs relations (cf. figure 27). Cette décomposition fonctionnelle illustre l'architecture logicielle retenue pour la conception globale de la maquette du Tableau Magique. Nous abordons maintenant la conception détaillée des trois unités de traitement les plus importantes, et qui présentent des aspects techniques intéressants : le suivi de doigt, le contrôle de la caméra, et la capture haute résolution.

4.3 SUIVI DE DOIGT

Dans cette section, nous abordons la technique de détection par corrélation qui a été retenue pour l'implémentation du suivi de doigt, puis nous complétons notre description en évoquant la gestion du suivi de doigt au niveau d'abstraction supérieur.

4.3.1 PRINCIPE

La détection par corrélation est sans doute la solution la plus naturelle au problème de la recherche d'un objet dans une image. Son principe de fonctionnement est le suivant : on effectue une "mesure de similarité" d'un morceau de l'image et d'un motif de référence (image représentant l'objet recherché) dans toutes les positions que peut occuper ce motif au sein de l'image de départ. La mesure de similarité est calculée en comparant chaque pixel du motif à son correspondant dans l'image.

Le calcul de la similarité entre deux images n'admet pas une solution unique. C'est pourquoi on trouve une grande variété de détections par corrélation, chacune appliquant sa mesure de similarité. Nous présentons ici deux mesures de similarité (une seule des deux a effectivement été utilisée dans ce projet). Dans la suite du paragraphe, les mesures de similarité sont calculées pour un motif M de taille mxn comparé à la zone de même taille dans l'image I à la position (i,j).

La similarité de deux images peut être vue comme la somme des écarts des valeurs de pixels. Plus cette somme est proche de zéro, plus les images sont similaires. En pratique, ils est plus intéressant de sommer les carrés de différences de pixels car cela fait intervenir une multiplication, moins coûteuse en temps de calcul qu'une valeur absolue. On parle alors de SSD (Sum of Squared-Differences) :

$$SSD(i, j) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} (M(u, v) - I(i+u, j+v))^2 \tag{1}$$

Cette formule est idéale lorsque les conditions lumineuses restent inchangées entre les différentes images. La fonction s'annule lorsque le motif coïncide avec son double dans l'image. Cependant, les conditions lumineuses varient en fonction du temps et de la position de la cible. La deuxième mesure de similarité permet de résoudre ce problème.

En effectuant une normalisation par la luminosité, on obtient une mesure plus robuste aux variations lumineuses. On parle alors de NCC (Normalized Cross-Correlation) :

$$NCC(i, j) = \frac{\sum_{u=0}^{m-1} \sum_{v=0}^{n-1} M(u, v) I(i+u, j+v)}{\sqrt{\sum_{u=0}^{m-1} \sum_{v=0}^{n-1} (M(u, v))^2 \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} (I(i+u, j+v))^2}} \tag{2}$$

A l'évidence, l'expression mathématique ci-dessus indique que la mesure NCC est plus coûteuse en temps de calcul que la mesure SSD (équation 1). Afin de réduire le coût de l'algorithme, la recherche du motif ne s'effectue pas sur toute l'image mais seulement sur une partie de celle-ci. Cette partie de l'image, appelée zone de recherche, est déterminée en fonction de la dernière position connue de la cible dans l'image. L'étude de la taille optimale de la zone de recherche et du motif est présentée dans [Bérard 94].

4.3.2 MISE EN OEUVRE

L'algorithme de suivi commence par une phase d'initialisation pendant laquelle on effectue le choix du motif à suivre. La figure suivante montre un exemple de motif, ici un doigt, dans une image du tableau.

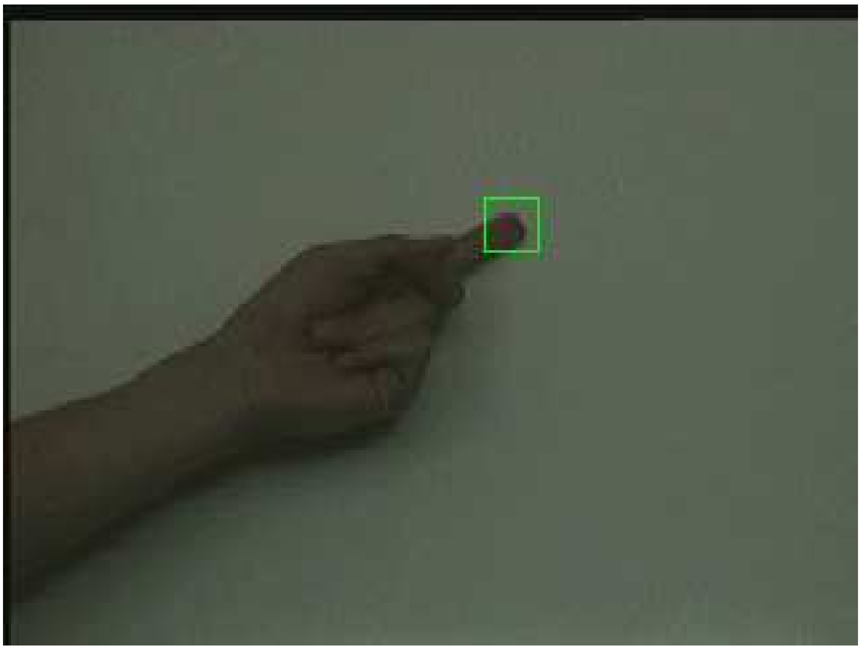


Figure 1 : Initialisation du suivi de doigt.

Une fois le motif choisi, le processus de localisation peut commencer. La zone de recherche est déterminée à partir de la position précédente du motif. Puis le motif est recherché dans cette zone, c'est-à-dire la mesure de similarité est calculée pour chacune des positions que le motif peut prendre dans cette zone. La figure 2 illustre ce processus.

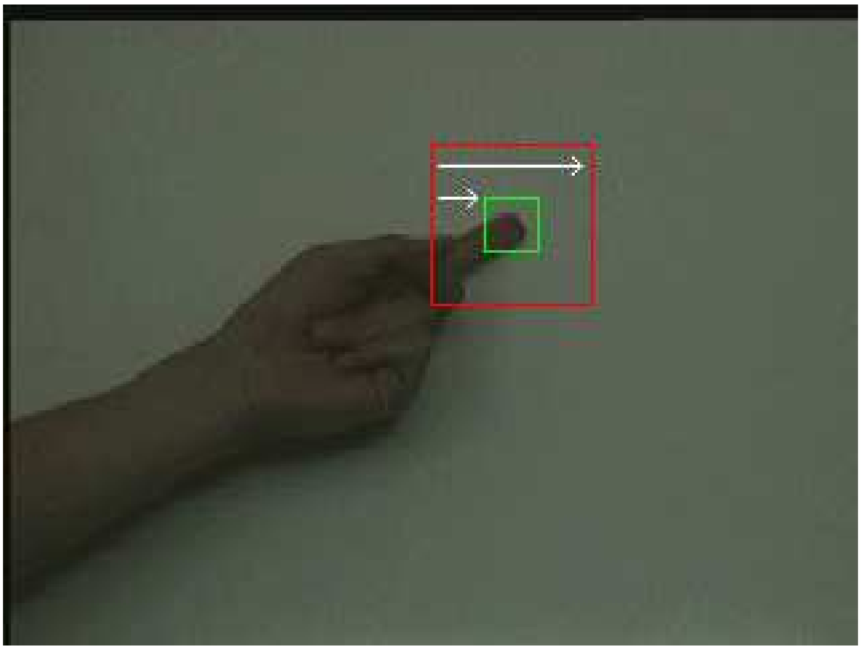


Figure 2 : Recherche par corrélation.

De cette recherche, on extrait une carte de corrélation qui représente les résultats du calcul des mesures de similarité. Sur la figure suivante, qui est la carte de corrélation résultant du processus de localisation, plus le pixel est blanc, plus la valeur de similarité est bonne.

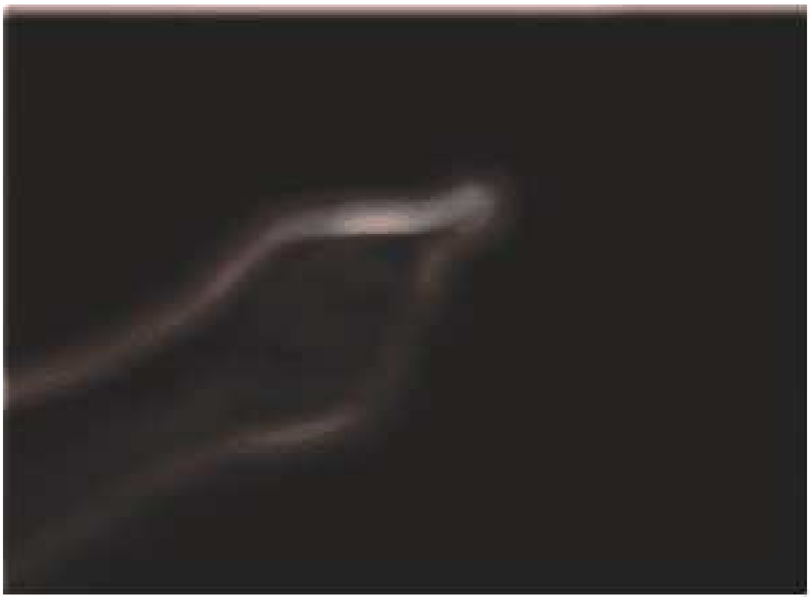


Figure 3 : Carte de corrélation.

La position du motif est alors déterminée en prenant la meilleure valeur de similarité.

L’initialisation d’un suivi de doigt par corrélation pour l’usage interactif du Tableau Magique a été réalisée de la manière suivante :

- une “zone sensible” est fixée, pour laquelle le système calcule en permanence la différence entre deux images successives (voir explication de la technique de différence d’images en annexe G), les images étant issues de la caméra. Le résultat de ces différences est une énergie qui représente le degré de différence entre les deux images. En comparant les variations de cette énergie à un seuil judicieusement choisi, il est possible de détecter les mouvements qui se produisent dans cette zone sensible, comme par exemple l’entrée dans cette zone du doigt de l’utilisateur.
- une fois que le doigt de l’utilisateur est stabilisé dans cette zone, on peut faire l’acquisition de l’image de cette zone du tableau comme motif de référence pour un suivi par corrélation.

Plus précisément, on peut arriver à détecter que le doigt de l’utilisateur s’est stabilisé dans cette zone dès que l’on a observé un retour à des valeurs normales de l’énergie à la suite d’un pic. En effet, l’entrée du doigt dans la zone sensible provoque une élévation rapide de cette énergie. Lorsque le doigt se stabilise, la différence entre deux images successives est suffisamment faible pour que l’énergie de la différence chute significativement. En pratique, on observe les variations illustrées par le graphe suivant.

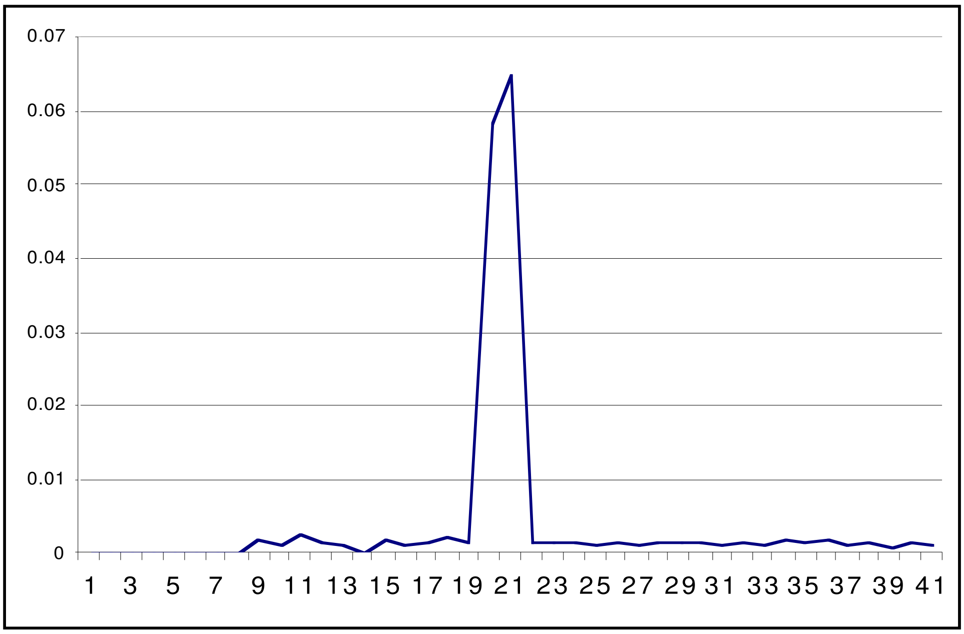


Figure 4 : Variations d’énergie dans la zone sensible (en ordonnées, les valeurs d’énergie, en abscisse les numéros de frames d’image).

En se fixant un seuil d’énergie, on peut donc détecter tour à tour l’entrée du doigt dans la zone sensible et sa stabilisation dans cette zone.

Le problème du choix du seuil est crucial. En effet, la variation des conditions d'éclairage et la taille de la zone observée influent sur les valeurs de l'énergie de différence et sont donc, entre autres, deux facteurs qui empêchent l'utilisation d'un seuil fixé manuellement. Nous avons donc travaillé à l'élaboration d'un procédé de calcul automatique du seuil. La méthode adoptée est la suivante :

- Dans un permier temps, la moyenne et l'écart-type des valeurs de l'énergie de la différence sont calculées (le calcul est effectué sur un nombre de valeurs qui doit être suffisamment grand pour que le calcul soit fiable).
- Le seuil est ensuite determiné en ajoutant à la moyenne un multiple de l'écart-type. Soit m la moyenne des énergies, σ l'écart-type et p le paramètre additionnel, alors le seuil S est calculé selon la formule suivante :

$$S = m + p\sigma \qquad (3)$$

Le paramètre p peut être considéré comme la sensibilité de la zone d'initialisation. Le graphe suivant représente les variations au cours du temps de l'énergie de la différence d'images calculée dans la zone sensible, ainsi que celles de m et de S.

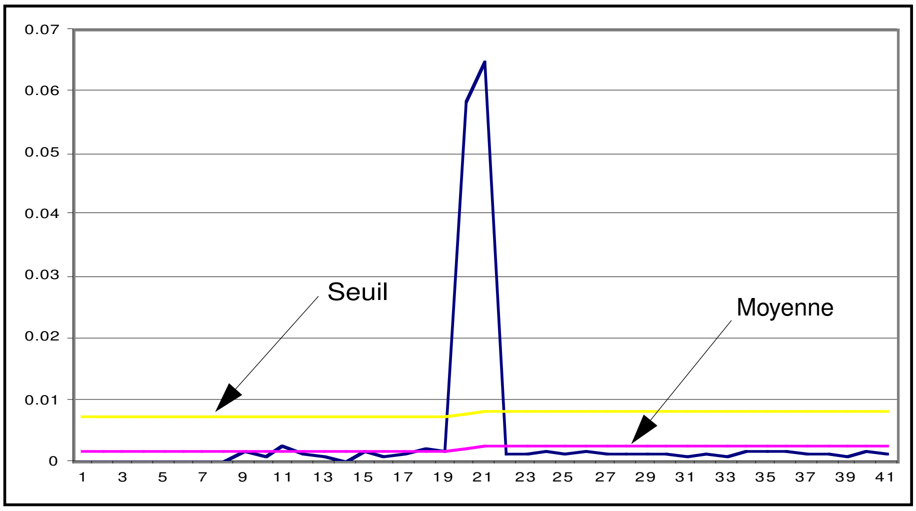


Figure 5 : Variations du seuil et de la moyenne.

Mathématiquement, le choix de cette technique se justifie par le fait qu'il existe une valeur du paramètre p telle que, en supposant que la distribution des valeurs de l'énergie suit une loi normale, la probabilité qu'une variation normale d'énergie (c'est-à-dire due seulement au bruit de caméra ou aux variations des conditions lumineuses) entraîne un dépassement du seuil ainsi calculé est inférieure à 2%. Cette valeur se situe aux alentours de 3. Nous avons pris une valeur par défaut de 5, ce qui assure plus de sécurité. Ainsi, une telle technique limite les risques de dépassement intempestif du seuil.

Ce procédé de détection de mouvement a fait l'objet du développement d'un composant dédié (que nous avons appelé détecteur de mouvement) qui a été réutilisé plus tard, notamment pour le développement de l'interface Homme-Machine (voir Annexe D : "L'interface Homme-Machine").

La méthode de corrélation nous permet donc, comme nous venons de le voir, de connaître la position d'un doigt dans l'image issue de la caméra. A partir de la suite des positions successives du doigt, nous évaluons la trajectoire saptio-temporelle de ce doigt et nous générons les événements "pause" ou "mouvement" selon le cas. Les composants utilisant le suivi de doigt se servent de ces événements. Techniquement, on peut détecter le mouvement ou la pause du doigt en intégrant les données issues de l'algorithme de suivi sur une durée plus ou moins longue. En effet, le mouvement du doigt se traduit par une variation de la position instantanée, tandis que la pause est caractérisée par une absence prolongée de mouvement.

En pratique, nous détectons la pause lorsque la position du doigt ne quitte pas une certaine zone durant une durée déterminée (la durée de la pause). Cette tolérance vis-à-vis de la position du doigt est indispensable vu les imprécisions inhérentes à la technique de suivi par corrélation. En effet, la position exacte du doigt sur le tableau peut ne correspondre à aucun pixel particulier de l'image (la position exacte se situe "entre" deux pixels) : dans ce cas, l'algorithme de corrélation peut fournir deux positions valables pour le doigt, différant d'un pixel seulement, entraînant ainsi une oscillation de la position retournée. De même, le bruit de caméra provoque des oscillations de ce type. La mise en place d'une zone de tolérance permet de nous affranchir de ce problème.

Les clients du suivi de doigt s'abonnent à ces événements et peuvent leur associer des traitements. C'est le cas notamment du noyau fonctionnel et de l'interface Homme-Machine (voir Annexe C : "Le noyau fonctionnel" et Annexe D : "L'interface Homme-Machine").

4.3.3 PROBLÈMES RENCONTRÉS

Le principal problème posé par la mise en œuvre du suivi de doigt est l'obtention d'un bon niveau de robustesse. La principale difficulté vient du fait qu'il est difficile de trouver un compromis entre la taille du motif à rechercher et la fréquence de fonctionnement (un motif de petite taille améliore la fréquence de fonctionnement et la vitesse de suivi mais ne permet pas de mémoriser un objet de forme significative, ce qui rend le suivi plus aléatoire).

Pour gagner en robustesse, nous avons réduit la taille des images issues de la caméra par quatre (réduction par deux sur chaque dimension). La fréquence de fonctionnement ne dépendant que de la taille du motif et de la zone de recherche, il est ainsi possible pour une fréquence de fonctionnement équivalente de suivre un motif représentant une partie du tableau quatre fois plus grande (et donc de forme plus significative) dans une zone de recherche quatre fois plus grande, ce qui autorise une vitesse de déplacement du doigt plus importante (voir l'évaluation des performances au paragraphe suivant).

4.3.4 ÉVALUATION DES PERFORMANCES

L'usage interactif en situation du Tableau Magique nécessite un suivi de doigt très performant. Dans l'idéal, nous devons avoir un suivi de doigt en temps-réel, c'est-à-dire un suivi qui réponde à la Loi de Fitts [Card 83]. En pratique, les limitations matérielles telles que la puissance des machines utilisées nous imposent de nous limiter, lors du calcul de corrélation mis en oeuvre pour le suivi de doigt, à une zone de recherche plus petite que l'image elle-même. Cela se traduit par une limitation des performances du suivi de doigt.

Les performances de notre suivi de doigt peuvent être mesurées par deux paramètres :

- La fréquence de fonctionnement
- La vitesse maximale du doigt autorisée

ÉTUDE DE LA FRÉQUENCE DE FONCTIONNEMENT

Supposons fixe la taille du motif recherché. Pour une position de ce motif dans la zone de recherche, le temps de calcul de la valeur de corrélation (voir 4.3.1) est fixe. Suivant la taille de la zone de recherche, le temps de localisation du motif dans cette zone, c'est-à-dire le temps nécessaire au calcul de l'ensemble des valeurs de corrélation pour toutes les positions possibles du motif, va être différent. Intuitivement, plus la taille de la zone de recherche est grande, plus le temps de localisation du motif est important, et par conséquent, plus la fréquence de fonctionnement est faible.

Notre expérience de l'utilisation de la technique de corrélation pour le suivi d'un doigt nous suggère l'utilisation d'un motif de taille (32x32). La figure suivante présente les variations de la fréquence de fonctionnement du suivi de doigt[1] en fonction de la taille de la zone de recherche.

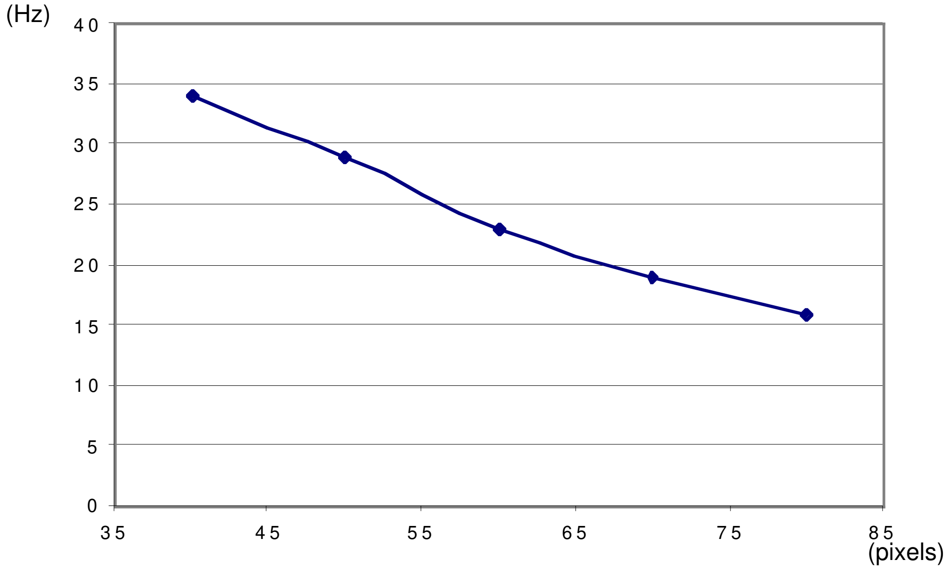


Figure 6 : Fréquence de fonctionnement du suivi de doigt en fonction de la taille de recherche.

Nous avons pris une zone de recherche de taille (50x50). Ce choix se justifie par le fait que cette valeur est la plus grande taille telle que la fréquence de fonctionnement est au moins égale à 25Hz. En pratique, il est inutile de prendre une taille inférieure car la fréquence de l'acquisition vidéo est constante et de 25Hz (de plus, une zone de recherche trop petite fait chuter la vitesse de suivi - voir paragraphe suivant).

---

[1] La fréquence de fonctionnement a été mesurée en situation d'usage interactif du Tableau Magique, dans des conditions "idéales" (pas d'autre traitement en tâche de fond - suivi de doigt seul)

ÉTUDE DE LA VITESSE MAXIMALE DE SUIVI

La vitesse maximale de suivi est la vitesse maximale de déplacement du doigt autorisée sans qu'il y ait “décrochage”, c'est-à-dire sans que le suivi perde le motif à suivre. En effet, comme l'algorithme effectue la recherche du motif dans une zone réduite autour de la dernière position connue du motif, si celui-ci sort de la zone de recherche entre deux instants de calcul, le suivi perd le motif. De fait, le doigt ne doit pas quitter l'espace restreint situé entre la position du motif et la frontière de la zone de recherche. La figure 7 montre le déplacement maximal autorisé entre deux instants de calcul pour que le suivi ne perde pas le motif.

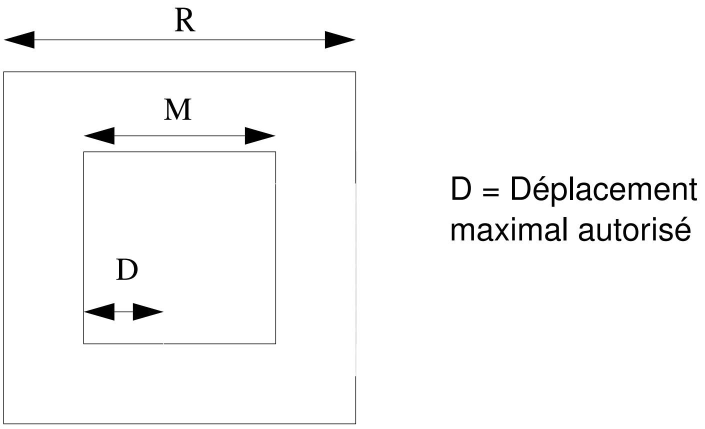


Figure 7 : Vitesse maximale de suivi

Ce déplacement peut facilement être chiffré en pixels. En effet si on note D ce déplacement, M la taille du motif, et R la taille de la zone de recherche, on a :

$$D = \frac{M}{2} + \left(\frac{R}{2} - M\right) = \frac{R}{2} - \frac{M}{2} \tag{4}$$

La vitesse maximale de suivi, en pixels/seconde en fonction de la taille de la zone de recherche est donc donnée par la formule suivante :

$$V = \frac{R - M}{2T} = \frac{F(R - M)}{2} = \frac{F(R)(R - M)}{2} \tag{5}$$

où V représente la vitesse de suivi en pixels/seconde, T le temps séparant deux instants de calcul de la corrélation et F=1/T la fréquence de fonctionnement (qui dépend elle aussi de la taille de la zone de recherche, voir 4.3.4).

A cette vitesse en pixels/seconde correspond une vitesse en centimètres/seconde. Cette vitesse dépend de la position de la caméra par rapport au tableau (plus la caméra est éloignée du tableau, plus un pixel correspond à une distance élevée). Dans cette étude de la vitesse maximale de suivi, la position de la caméra par rapport au tableau est fixe. Si on fait l'approximation que la relation entre la distance en centimètres sur le tableau et la distance en pixels sur l'image caméra est constante sur toute la surface du tableau, la vitesse du suivi exprimée en centimètres/seconde est directement proportionnelle à la vitesse en pixels/seconde. La figure 8 illustre les variations de cette vitesse en fonction de la taille de la zone de recherche.

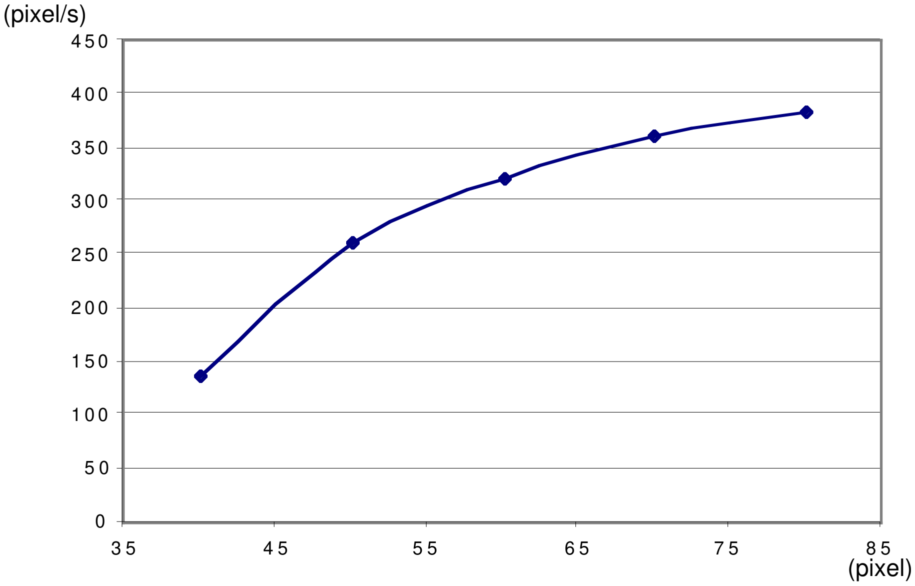


Figure 8 : Vitesse maximale du suivi en fonction de la taille de la zone de recherche.

Il est intéressant de comparer les résultats obtenus aux performances requises par l'usage interactif en situation. En effet, le principal défi que représente le développement d'un suivi de doigt est l'adaptation de la vitesse de suivi à la vitesse des mouvements naturels de l'utilisateur. Nous avons mesuré expérimentalement la vitesse des mouvements naturels d'un utlisateur en situation d'usage. La valeur mesurée est de 290 pixels / s. (pour une caméra située à environ 1m40 du tableau). Notre système est capable de suivre des cibles à cette vitesse pour une taille de motif de 32*32 pixels et une taille de zone de recherche de 60*60 pixels. Avec ces paramètres, la fréquence de fonctionnement du suivi est de 23 Hz.

Cette étude comparative met en évidence le fait qu'il est possible d'adapter le suivi de doigt aux mouvements naturels de l'utilisateur.

4.4 CONTRÔLE DE LA CAMÉRA

Dans cette section, nous abordons le contrôle de l'orientation et du zoom de la caméra pour la capture haute résolution.

4.4.1 PRINCIPE

Le but du contrôle de la caméra est de se donner la capacité de la centrer sur n'importe quelle partie du tableau : étant donnée une zone rectangulaire du tableau, nous devons pouvoir en déduire le "**pan**" (positionnement horizontal), le "**tilt**" (positionnement vertical) ainsi que le zoom pour centrer cette zone dans l'image. Ce positionnement devant se faire rapidement afin de ne pas déranger l'utilisateur, la caméra devra subir une phase de calibrage à l'initialisation du système. Cette phase aura pour but de lui faire "découvrir" le tableau.

4.4.2 MISE EN OEUVRE

Comment centrer une zone rectangulaire donnée du tableau ?

Nous allons décrire dans ce paragraphe un algorithme simple qui sera utilisé plus tard comme un élément essentiel de la solution retenue.

Supposons que nous désirons centrer la caméra sur une zone rectangulaire du tableau. La caméra est en plan large et visualise le tableau dans son intégralité. Effectuons les 3 pas suivants :

- Nous allons tout d'abord projeter un rectangle sur cette zone puis récupérer l'image visualisée par la caméra.
- Ensuite nous effaçons ce rectangle et recupérons de nouveau l'image visualisée.
- Nous calculons ensuite la différence de ces deux images. En appliquant un algorithme de connexité à l'image obtenue, nous pouvons trouver le rectangle et ainsi en déduire ses coordonnées dans l'image caméra.

Maintenant, effectuons un mouvement de caméra le plus grand possible sans perdre de vue le rectangle. Si nous recommençons les 3 phases précédentes nous récupérons de nouvelles coordonnées du rectangle dans la nouvelle image caméra.

A ce stade, nous avons donc deux positions différentes du même rectangle. Nous pouvons en déduire les ratios suivant :

$$r_{pan} = \frac{pan_2 - pan_1}{x_2 - x_1} \tag{6}$$

$$r_{pan} = \frac{pan_2 - pan_1}{y_2 - y_1} \tag{7}$$

où, pour i=1 ou 2 :

- pan_i est le pan de la caméra lors de la capture i du rectangle
- $tilt_i$ est le tilt de la caméra lors de la capture i du rectangle
- x_i est l'abscisse du centre du rectangle sur l'image caméra lors de la capture i
- y_i est l'ordonnée du centre du rectangle sur l'image caméra lors de la capture i

Grâce à ces ratios, on peut déduire une approximation du pan et du tilt permettant de centrer les rectangles

$$\mathrm{pan} = \mathrm{pan}_2 + (x_2 - x_c) \cdot r_{\mathrm{pan}} \tag{8}$$

$$\mathrm{tilt} = \mathrm{tilt}_2 + (y_2 - y_c) \cdot r_{\mathrm{tilt}} \tag{9}$$

où :

- x_c est l'abscisse du centre de l'image (sur l'image caméra)
- y_c est l'ordonnée du centre de l'image (sur l'image caméra)

Ainsi, on parvient à centrer la caméra sur la zone désirée. Le principal problème de cette technique est qu'elle nécessite l'affichage puis l'effacement d'un rectangle. Par conséquent, elle ne peut pas être utilisée telle qu'elle sous peine de rendre l'utilisation du Tableau Magique très fastidieuse. Nous allons donc utiliser cette technique uniquement pendant une période de calibrage de la caméra.

LE CALIBRAGE DE LA CAMÉRA

Le but du calibrage est de fournir un contrôle de haut niveau de la caméra : être capable de centrer la caméra sur une zone du tableau de façon immédiate (c'est-à-dire être capable de régler le pan, le tilt et le zoom de la caméra pour n'importe quel point du tableau).

LE CALIBRAGE DES POSITIONS

Le principe est de récupérer le pan et le tilt pour chacun des points d'un quadrillage recouvrant le tableau. Nous constituons ainsi une table associant un point du tableau à un pan et un tilt permettant de centrer la caméra dessus. Nous appellerons cette table "*table des positions*".

Pour constituer la table, nous utilisons, pour chaque point du quadrillage, l'algorithme décrit précédemment avec quelques raffinements supplémentaires :

- nous centrons la caméra sur un rectangle (dont le centre est un des points du quadrillage) en position de zoom grand angle (le tableau est alors entièrement visible par la caméra)
- nous zoomons la caméra au maximum. Le rectangle est toujours visible (sa taille a été choisie suffisamment petite par rapport aux capacités de zoom de la caméra pour être visible en zoom maximum) mais plus aussi bien centré puisque le centrage a été effectué en grand angle. Nous l'affinons donc.
- le pan et le tilt de la caméra sont maintenant modifiés d'une unité à la fois et la position du rectangle réévaluée. Ceci jusqu'à ce qu'il soit centré au mieux. A ce stade, un mouvement de caméra ne peut plus apporter d'amélioration. Mais on peut encore bouger le rectangle lui-même!
- l'abscisse et l'ordonnée du rectangle sont modifiées d'un pixel à la fois et la position du rectangle réévaluée. Ceci jusqu'à ce qu'il soit centré au mieux.
- nous avons maintenant une association très précise (relativement aux capacités de la caméra et du projecteur) entre la position d'un point et la position de la caméra.

A la lecture de l'algorithme ci-dessus, on peut constater que la table des positions obtenue n'est pas un quadrillage parfait puisque l'on peut être amené à modifier la position des rectangles (dernière phase de l'algorithme). Ceci n'est pas un problème, l'important étant d'avoir des correspondances les plus exactes possibles.

LE CALIBRAGE DU ZOOM

Une méthode choisie est de consituer une autre table : la *table des zooms*.

A cette fin, nous allons d'abord projeter un rectangle au centre du tableau et centrer la caméra sur lui. Ensuite, nous faisons varier le zoom du plan large au plan serré avec un pas donné. Pour chaque position de zoom on récupère la surface du rectangle (toujours grâce à une différence d'images) et on conserve l'association surface/zoom dans la table.

Pour compléter nos informations, une *table des surfaces* est aussi créée lors du calibrage des positions : une fois un point du quadrillage centré, nous conservons sa surface en mémoire dans la table en fonction des coordonnées du point. Il est à noter que cette surface est obtenue pour un zoom en plan serré, c'est-à-dire pour le même zoom que la dernière surface contenue dans la table des zooms.

Pour la phase de calibrage, nous utilisons un rectangle au format 4/3 (longueur = largeur x 4/3). Ce format étant celui de l'image caméra, il était logique de l'utiliser. Et par conséquent, pour centrer une zone rectangulaire nous utiliserons toujours le rectangle au format 4/3 englobant.

UTILISATION DES TABLES CRÉÉES LORS DU CALIBRAGE

A partir de la table des positions, nous approximons linéairement le pan et le tilt pour un point donné en utilisant les informations obtenues pour les 4 points qui l'entourent. Voici la description de cette interpolation :

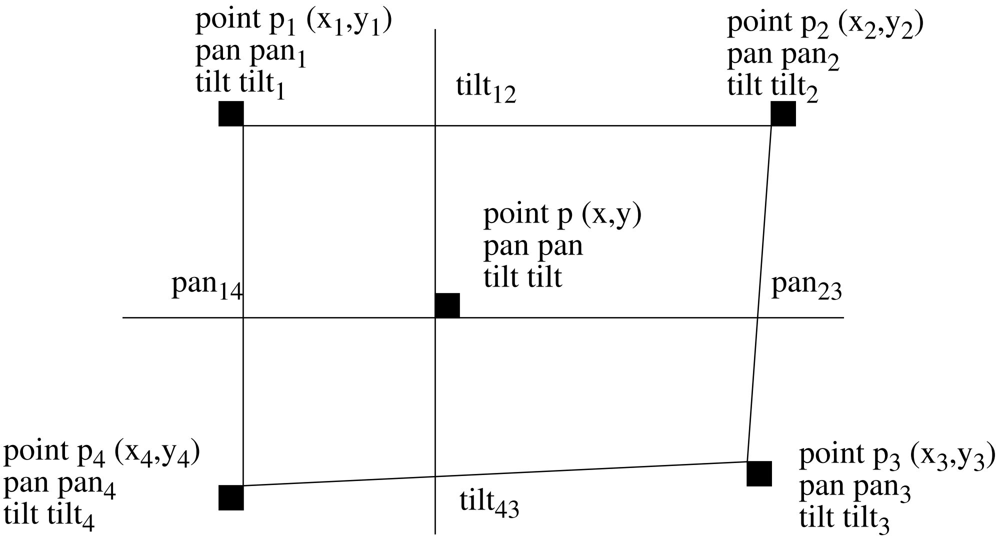

point p_1 (x_1,y_1)
pan pan_1
tilt $tilt_1$
$tilt_{12}$
point p_2 (x_2,y_2)
pan pan_2
tilt $tilt_2$
point p (x,y)
pan pan
tilt tilt
pan_{14}
pan_{23}
point p_4 (x_4,y_4)
pan pan_4
tilt $tilt_4$
$tilt_{43}$
point p_3 (x_3,y_3)
pan pan_3
tilt $tilt_3$

Figure 9 : Calcul de pan et tilt par interpolation

pan_{14} est le pan interpolé au point d'intersection (x_{14},y) entre la droite (p_1p_4) et la droite horizontale passant par p :

$$x_{14} = \frac{(x_1 - x_4) \cdot (y - y_4)}{(y_1 - y_4)} + x_4 \qquad (10)$$

$$pan_{14} = \frac{(pan_4 - pan_1) \cdot \sqrt{(x_{14} - x_1)^2 + (y - y_1)^2}}{\sqrt{(x_4 - x_1)^2 + (y_4 - y_1)^2}} + pan_1 \qquad (11)$$

A l'aide de pan_{14} et pan_{23}, on interpole le pan au point désiré :

$$pan = \frac{(pan_{23} - pan_{14}) \cdot (x - x_{14})}{(x_{23} - x_{14})} + pan_{14} \qquad (12)$$

L'interpolation du tilt au point désiré se fait d'une façon similaire, en utilisant les tilts interpolés $tilt_{12}$ et $tilt_{43}$ (Cf graphe précédent) :

$$tilt = \frac{(tilt_{43} - tilt_{12}) \cdot (y - y_{12})}{(y_{43} - y_{12})} + tilt_{12} \qquad (13)$$

Cette technique ne fonctionne évidemment que pour un point pouvant être entouré par quatre points du quadrillage. Le cas contraire, c'est-à-dire celui d'un point sur le bord du tableau, n'a pas été envisagé. La raison principale étant que l'utilisation qui est faite de cet algorithme de positionnement n'amène à traiter que des points "convenables".

L'obtention du zoom se fait grâce à la table des zooms et à la table des surfaces.

Tout d'abord, la table des surfaces nous permet d'approximer la surface du rectangle qui aurait été vue au zoom en plan serré au point considéré, grâce aux 4 points (et donc aux 4 surfaces associées) qui l'entourent :

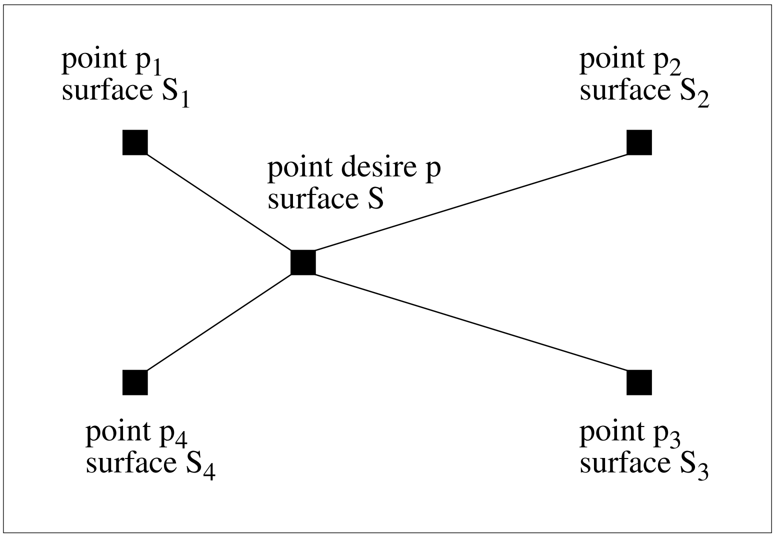


Figure 10 : Approximation de la surface S

Les surfaces S_1, S_2, S_3 et S_4 sont les surfaces, observées par la caméra, d'un même rectangle positionné aux points p_1, p_2, p_3 et p_4. La surface S est la surface approximée (par un barycentre des 4 surfaces qui l'entourent) de celle qui aurait été vue si le rectangle avait été visualisé au point p par la caméra. Une fois cette surface évaluée, on peut en déduire un ratio avec la surface, vue au même niveau de zoom, dans la table des zooms. Ce ratio représente donc la variation de surface vue entre deux rectangles identiques positionnés, l'un au centre du tableau, l'autre sur le point considéré.

Supposons que ce ratio soit constant quel que soit le zoom et notons le *r*. Notons $S_{\text{considérée}}$ la surface de la zone rectangulaire à centrer.

La table des zooms (TZ) ayant été constituée à l'aide d'un rectangle de surface S_{zooms}, il nous faut maintenant rechercher dans cette table le zoom maximal pour lequel la surface observée du rectangle de surface $S_{\text{considérée}}/S_{\text{zooms}}*r$ ne dépasse pas la taille de l'image. Ce zoom z est défini ainsi :

$$z = TZ(i, zoom) \tag{14}$$

où :

- soit i est l'entier tel que :
  la surface $[TZ(i,surface)*S_{\text{considérée}}/S_{\text{zooms}}*r]$ ne dépasse pas la taille de l'image et la surface $[TZ(i+1,surface)*S_{\text{considérée}}/S_{\text{zooms}}*r]$ dépasse la taille de l'image.

- soit i est le dernier élément du tableau et la surface [TZ(i,surface)* $S_{considérée}/S_{zooms}*r$] ne dépasse pas la taille de l'image.

4.4.3 PROBLÈMES RENCONTRÉS

Avant de parvenir à l'algorithme décrit ci-dessus, de nombreux tests ont été effectués.

Tout d'abord, nous avons pensé pouvoir approximer le pan et le tilt en une fonction linéaire des coordonnées (x,y) d'un point du tableau :

$$\mathrm{pan}(x, y) = a \cdot x + b \cdot y + c \qquad a,b,c \in \Re \tag{15}$$

$$\mathrm{tilt}(x, y) = d \cdot x + e \cdot y + f \qquad d,e,f \in \Re \tag{16}$$

Mais cette approximation s'est révélée trop forte. La précision n'était pas suffisante pour assurer le centrage d'un point quelconque du tableau.

Après cet échec, nous avons voulu analyser la forme des fonctions pan et tilt (en fonction des coordonnées x et y). Après avoir recueilli un ensemble de correspondances entre le pan (ou le tilt) et les coordonnées x et y, l'observation des courbes obtenues nous a permis de voir qu'il n'y avait probablement pas de modèle simple approximant d'une façon satisfaisante ces fonctions.

Nous avons donc fait le choix de calibrer la caméra avec un grand nombre de données “exactes” (à la précision de la caméra et du projecteur près) et d'utiliser ces données pour approximer simplement (linéairement) leurs valeurs en d'autres points. Ce choix s'est révélé payant, la précision désirée étant au rendez-vous. Le problème de cette technique étant la phase de calibration relativement lente, nous avons passé beaucoup de temps sur l'algorithme pour diminuer sa durée. Au final, environ 8 minutes sont nécessaires pour calibrer la caméra avec un quadrillage de 9*7 points.

4.4.4 ÉTUDE DES PERFORMANCES

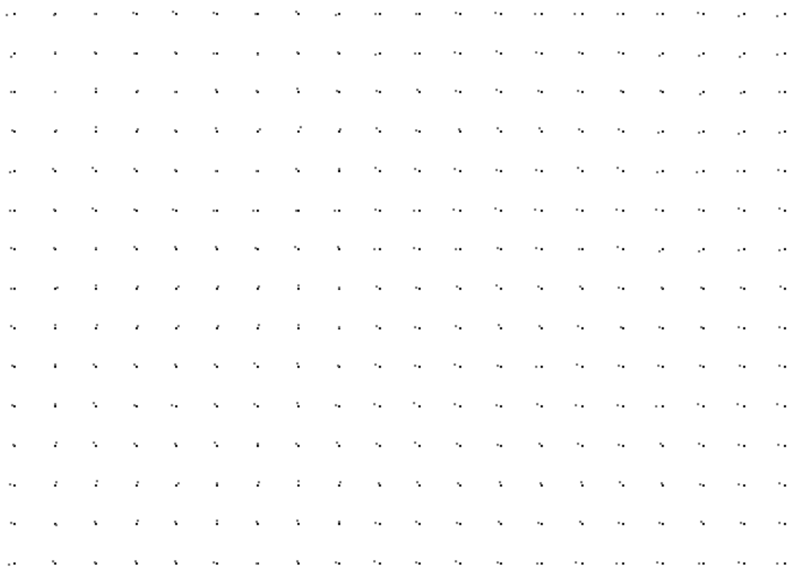


Figure 11 : Précision du positionnement de la caméra.

La figure 11 illustre les performances du positionnement (pan et tilt).

- Les points noirs représentent les endroits que la caméra devait pointer.
- Les points gris clairs représentent les endroits que la caméra a effectivement pointés.

On constate que la précision est acceptable pour l'utilisation que nous faisons du positionnement.

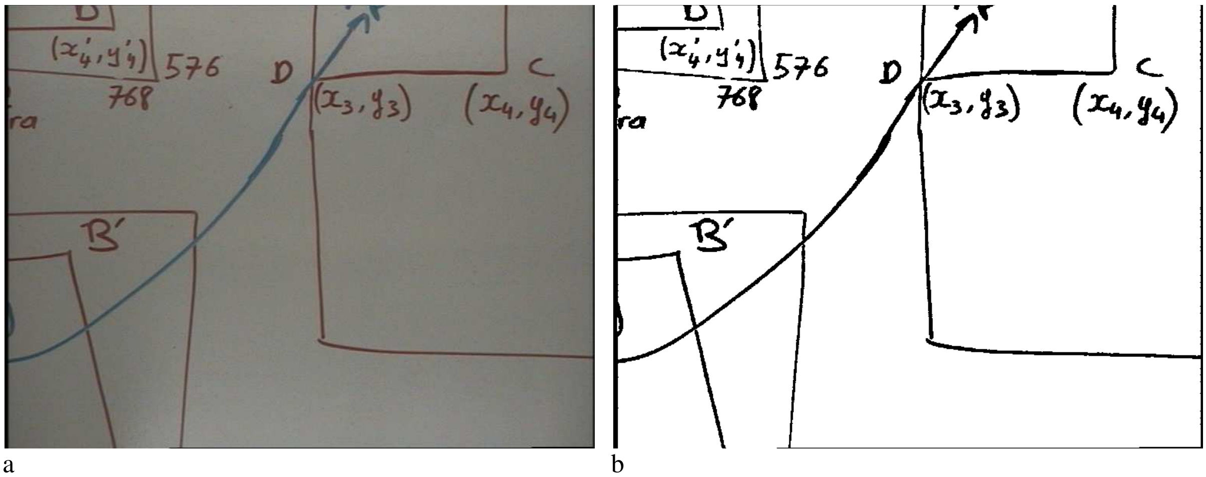
Concernant la fiabilité du positionnement du zoom, il est difficile de réaliser un test probant. Aussi, nous contenterons-nous de dire que sa précision, là encore, est satisfaisante pour l'utilisation qui en est faite dans le cadre du Tableau Magique.

4.5 NETTOYAGE D'IMAGES

4.5.1 PRINCIPE

L'image du tableau blanc acquise par la caméra est une image couleur contenant beaucoup de bruit (reflets de lumière sur le tableau, poussières d'encre, etc.). Pour distinguer l'encre du fond du tableau, l'image doit être nettoyée.

Le but est donc, à partir d'une image capturée par la caméra, d'obtenir une image noir et blanc où les points noirs correspondent à l'encre.



a b

Figure 12 : (a) Image capturée par la caméra, (b) Image nettoyée.

4.5.2 MISE EN OEUVRE

Deux techniques étaient à notre disposition pour effectuer ce nettoyage.

La première se base sur la construction d'un histogramme de couleurs et la recherche de différents modes (gaussiennes) sur celui-ci (en général 2 modes, le plus important correspondant aux pixels du fond du tableau et l'autre à l'encre). Ensuite, à partir de ces modes, un seuil est calculé qui sera le niveau de luminosité en-dessous duquel un pixel sera considéré comme représentant de l'encre.

Cette technique avait déjà été utilisée par des personnes du laboratoire; nous connaissions donc ses performances, à savoir qu'elle donne de bon résultats mais dans des temps relativement longs (de l'ordre de 2 secondes pour une image 768*576). Ceci n'était pas compatible avec l'interactivité souhaitée pour le système. Nous nous sommes donc tournés vers un second algorithme, a priori beaucoup plus rapide mais dont nous ne connaissions pas les performances.

TECHNIQUE DU «SEUILLAGE ADAPTATIF» DE P. WELLNER

Le principe de base est de calculer, pendant le parcours de l'image, un seuil qui soit la moyenne des intensités des derniers pixels rencontrés. Quand l'intensité d'un pixel est significativement inférieure à ce seuil, on le considère comme représentant de l'encre.

		p_{n-s}	...	p_{n-2}	p_{n-1}	p_n			

P_n représente l'intensité du pixel n dans une ligne de l'image qui doit être nettoyée.

Soit $f_s(n)$ la somme des intensités des s derniers pixels au point n,

$$f_s(n) = \sum_{i=0}^{s-1} p_{n-i} \tag{17}$$

L'intensité T(n) du point n correspondant dans l'image nettoyée est 0 (noir) ou 255 (blanc) suivant que p(n) est, respectivement, t % plus sombre ou non que la moyenne des intensités des s derniers pixels. Soit :

$$T(n) = \begin{cases} 0 \text{ si } p_n < \left(\frac{f_s(n)}{s}\right) \cdot \left(\frac{100-t}{100}\right) \\ 255 \text{ sinon} \end{cases} \tag{18}$$

La moyenne $f_s(n)$ peut être approximée en retranchant, pour chaque nouveau point, 1/s à l'ancienne moyenne et en ajoutant la luminosité de ce point.

L'approximation $g_s(n)$ de $f_s(n)$ s'écrit donc :

$$g_s(n) = g_s(n-1) - \frac{g_s(n-1)}{s} + p_n \tag{19}$$

D'après Wellner, cette approximation est meilleure que la valeur exacte puisqu'elle prend plus en considération les points proches du point n.

Il ne reste plus qu'à définir $g_s(0)$ pour chacune des lignes de l'image. Nous avons choisi de lui donner la valeur $s*p_0$ (p_0 étant la luminosité du premier pixel de la ligne considérée). Ce choix semble naturel quand on considère l'algorithme et, de toute façon, son influence se borne aux premiers pixels de l'image nettoyée.

Valeur des paramètres s et t

Wellner conseille de choisir les valeurs suivantes :

$$\begin{gathered} s = \frac{1}{8} \cdot \text{largeur image} \\ t = 15 \end{gathered} \tag{20}$$

A l'utilisation, on s'aperçoit que la valeur de t optimale est dépendante des conditions lumineuses.

4.5.3 PERFORMANCES

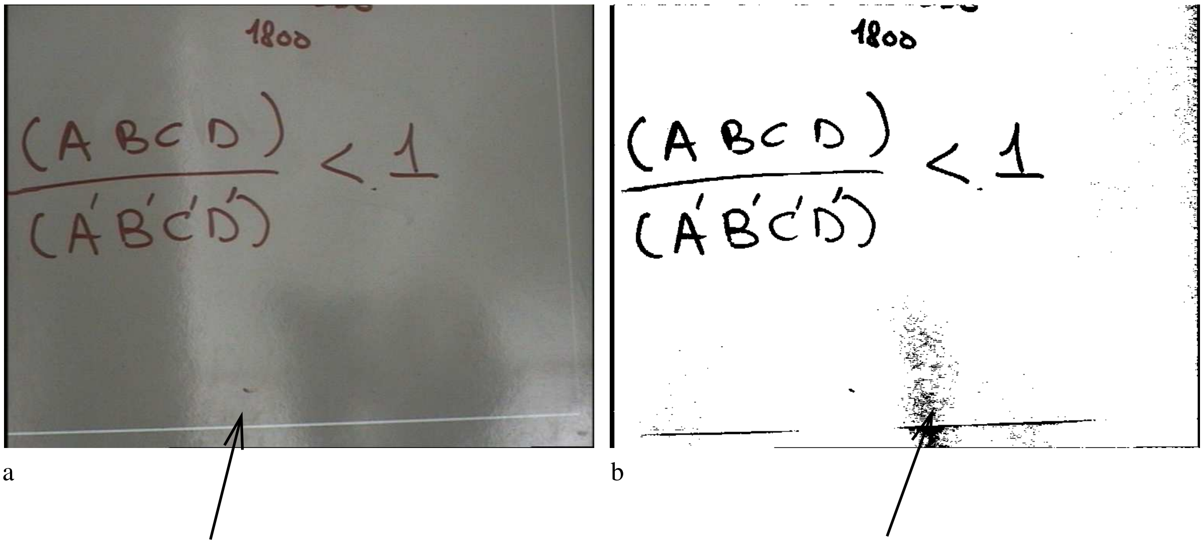
Cet algorithme est très simple à implémenter et est très rapide car l'image n'est parcourue qu'une seule fois.

Les résultats obtenus sont satisfaisants lorsque les conditions de luminosité sont bonnes. Cependant, quelques défauts apparaissent à l'utilisation.

Le premier problème intervient quand une trop longue zone d'encre est présente sur l'image. Dans ce cas, la fin de cette zone n'est plus considérée comme de l'encre. Cette faiblesse s'explique directement par l'algorithme lui même. En effet, lors du parcours d'une ligne sur cette zone d'encre, le seuil descend et finit par devenir trop faible pour que les points suivants soient vus comme de l'encre (ils ne sont plus t% inférieurs au seuil).

Ce problème peut être simplement résolu en effectuant deux parcours de l'image à nettoyer: un premier, comme décrit précédemment, suivant les lignes et l'autre suivant les colonnes (en gardant le même algorithme). Le résultat final est la somme des deux images nettoyées (un point de l'image finale sera noir s'il l'était dans l'une des deux images). Ainsi, une grande zone horizontale d'encre sera détectée par le parcours vertical et une grande zone verticale d'encre par le parcours horizontal.

Le second défaut apparaît quand il y a des taches claires sur l'image à nettoyer (dans notre cas, elles sont dûes à des reflets lumineux sur le tableau). Du bruit apparaît alors en limite droite de ces zones, comme le montre la figure 13.



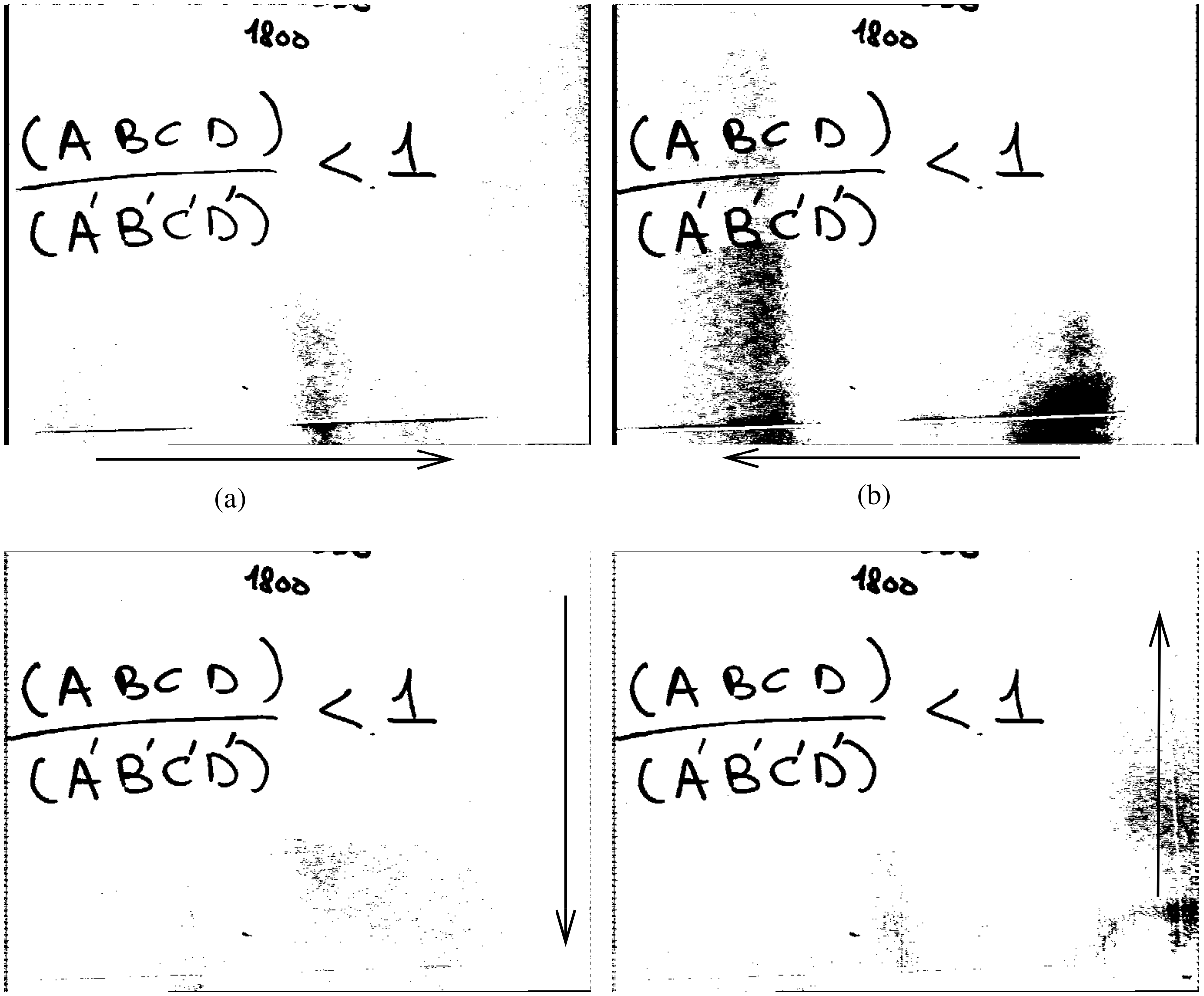
zone plus claire (reflet)
sens de parcours : de la gauche vers la droite

bruit apparu à cause du reflet
présent dans l'image à nettoyer

Figure 13 : Limitations de l'algorithme de Wellner. (a) image originale, (b) image traitée.

Ce bruit peut être fortement réduit en combinant les résultats du nettoyage de la même image selon quatre parcours différents.

La figure 14 illustre le résultat du nettoyage de la même image selon la technique de Wellner mais parcourue dans des sens différents.



(a) (b) (c) (d)

Figure 14 : Résultat de l'algorithme de Wellner selon les 4 parcours possibles :
de gauche à droite (a), de droite à gauche (b),
de haut en bas c), et de bas en haut (d).

Pour obtenir l'image nettoyée finale nous effectuons l'opération suivante : un point de l'image finale est noir s'il est noir dans les deux images résultant des parcours horizontaux (figure 14a et figure 14b) ou s'il est noir dans les deux images résultant des parcours verticaux (figure 14c et figure 14c).

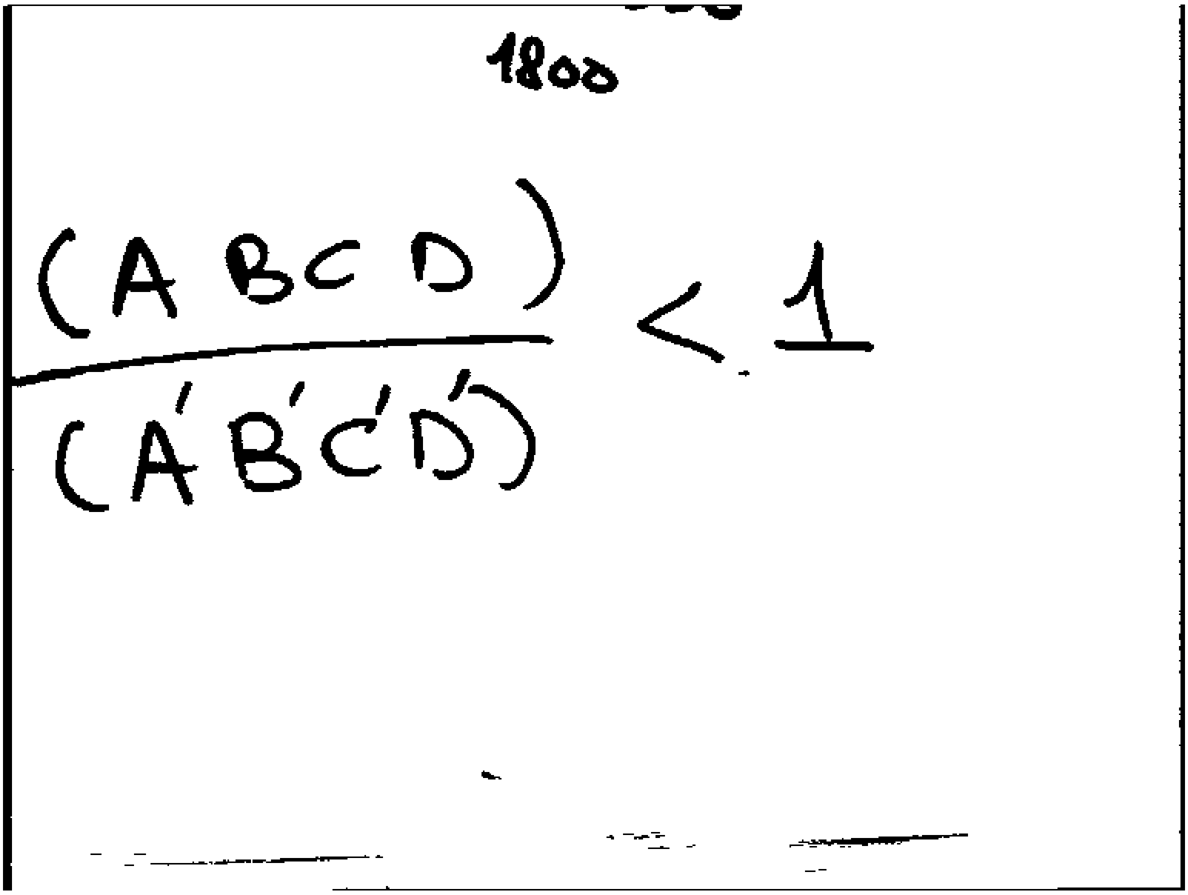


Figure 15 : image finale résultant de la combinaison des 4 images nettoyées

Comme on peut le voir, l'image obtenue par la combinaison de quatre images nettoyées selon des parcours différents est de meilleure qualité. Dans le cadre restreint du Tableau Magique, ce résultat nous convient parfaitement.

TEMPS DE CALCUL

Malgré les 4 parcours nécessaires au nettoyage de l'image, la méthode utilisée reste largement plus rapide que celle basée sur les histogrammes de couleurs.

Temps de calcul moyen pour nettoyer une image 768*576 sur le PowerMac : 0,6 secondes.

4.6 CALIBRAGE CAMÉRA

4.6.1 PRINCIPE

Le but du calibrage de la caméra est de mettre en correspondance les deux repères que sont le repère du projecteur et le repère de la caméra, c'est-à-dire de connaître la position d'un point dans l'un des deux repères à partir de ses coordonnées dans l'autre. Un exemple d'utilisation de cette correspondance est la projection sur le tableau d'un retour d'information de la position du doigt (cette position n'étant connue que dans le repère caméra).

On suppose que la fonction qui permet de passer d'un repère à l'autre est une transformation projective plane, caractérisée par une matrice de projection 3*3 H = [h_{ii}] telle que :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{22} \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \qquad (21)$$

(x',y') étant un point dans le repère caméra et (x,y) le point correspondant dans la repère projecteur. H est la matrice de transformation du repère caméra dans le repère projecteur; pour obtenir la transformation inverse, il suffit de travailler avec H^{-1}.

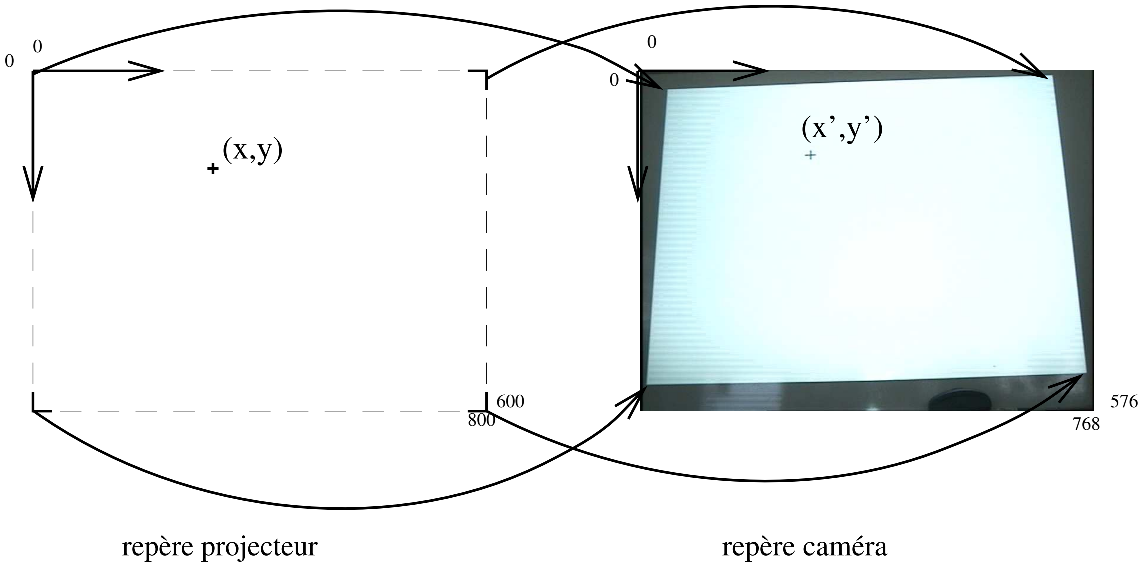


Figure 16 : Effet de la transformation projective sur une image de mire.

Pour déterminer cette matrice de projection, nous avons besoin de connaître la correspondance pour quatre points au minimum. En effet, chaque correspondance fournit deux équations linéaires utilisées dans le calcul de la matice H (équation 22).

$$x' = \frac{h_{11} \cdot x + h_{12} \cdot y + h_{13}}{h_{31} \cdot x + h_{32} \cdot y + h_{33}}, \qquad y' = \frac{h_{21} \cdot x + h_{22} \cdot y + h_{23}}{h_{31} \cdot x + h_{32} \cdot y + h_{33}} \tag{22}$$

soit, après transformation sous forme matricielle :

$$\begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -x' \cdot x & -x' \cdot y & -x' \\ 0 & 0 & 0 & x & y & 1 & -y' \cdot x & -y' \cdot y & -y' \end{bmatrix} \cdot h = 0 \tag{23}$$

où h = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}) est la matrice H écrite comme un vecteur.

Pour quatre points, nous obtenons donc le système linéaire de 8 équations à 9 inconnues suivant :

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1 \cdot x_1 & -x'_1 \cdot y_1 & -x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y'_1 \cdot x_1 & -y_1 \cdot y_1 & -y'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x'_2 \cdot x_2 & -x'_2 \cdot y_2 & -x'_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y'_2 \cdot x_2 & -y_2 \cdot y_2 & -y'_2 \\ x_3 & y_2 & 1 & 0 & 0 & 0 & -x'_3 \cdot x_3 & -x'_3 \cdot y_3 & -x'_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y'_3 \cdot x_3 & -y_3 \cdot y_3 & -y'_3 \\ x_3 & y_2 & 1 & 0 & 0 & 0 & -x'_4 \cdot x_4 & -x'_4 \cdot y_4 & -x'_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y'_4 \cdot x_4 & -y_4 \cdot y_4 & -y'_4 \end{bmatrix} h = 0 \tag{24}$$

La solution h de ce système A.h=0 est le noyau de A qui, si les quatre points ont été choisis correctement (non alignés), a pour dimension 1.

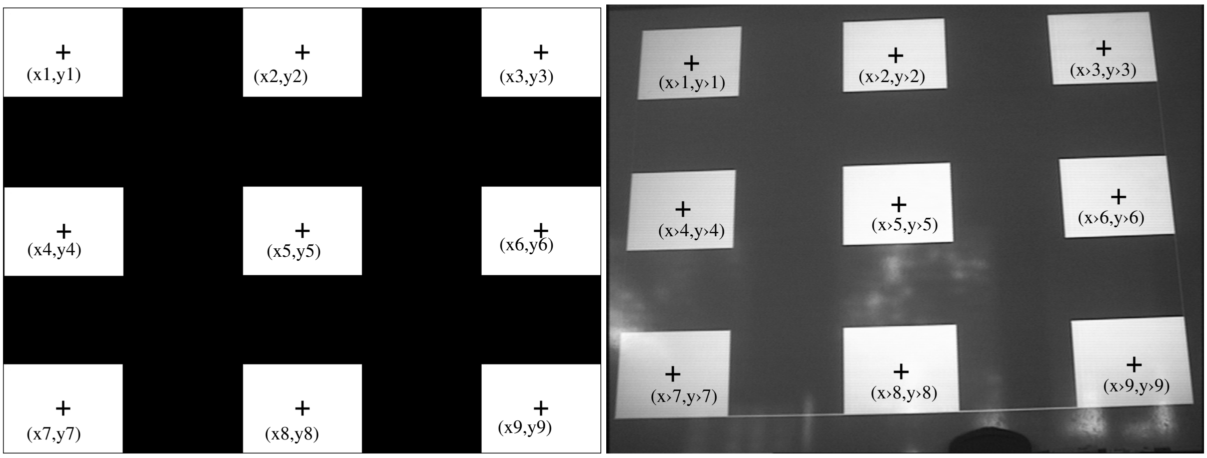
Il est possible d'utiliser plus de quatre points. Dans ce cas, A devient une matrice 2n*9 où n est le nombre de points choisis. Il n'existe alors plus de solution à l'équation A.h=0. Nous essayons donc de trouver h tel que $\|A.h\|$, ou $h^t(A^t.A)h$, soit minimum avec $\|h\| = 1$. Le h sastisfaisant cette condition est le vecteur propre correspondant à la plus petite des valeurs propres de $A^t.A$.

Dans le cas où n=4, la plus petite des valeurs propres est 0 et le vecteur propre est le noyau de A.

4.6.2 MISE EN OEUVRE

Le problème n'est pas la résolution de ce système linéaire A.h=0 ou la recherche de la plus petite des valeurs propre de $A^t.A$, mais plutôt la détermination des coefficients de A. En effet, il nous faut trouver la correspondance pour au moins 4 points.

Notre solution est de projeter sur le tableau une mire composée de 9 points, de coordonnés ((x1,y1), ... (x9,y9)), et de calculer ensuite leurs coordonnées ((x'1,y'1), ... (x'9,y'9)) sur l'image capturée par la caméra.

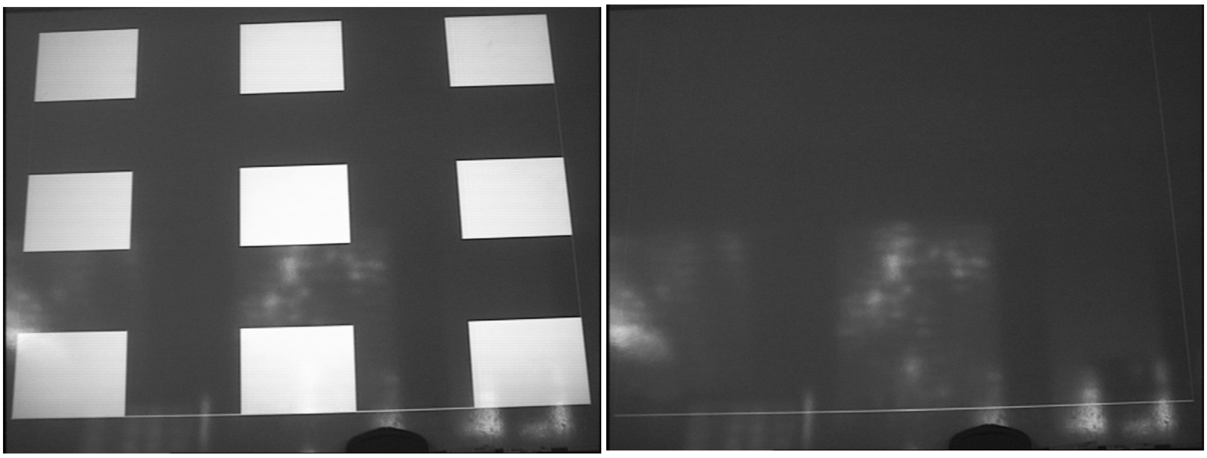


a b

Figure 17 : Mire projetée (a), image de la mire capturée (b).

Le calcul de ces coordonnées se décompose ainsi :

- Acquisition d'une image avec la mire (figure 18a)
- Acquisition d'une image sans la mire (figure 18b)



a b

Figure 18 : Image capturée, mire présente (a) et image capturée, mire absente (b).

- Calcul de la différence entre ces 2 images. Il en résulte une image ne contenant que du noir et du blanc, le blanc correspondant aux points où une différence a été détectée (figure 18).
- On procède ensuite à la recherche des 9 zones connexes les plus grandes correspondant, a priori, aux 9 rectangles projetés vus par la caméra. On calcule ensuite le barycentre de ces 9 zones. A ce stade, nous avons donc déterminé les coordonnées de tous les points de la mire dans le repère caméra mais nous ne savons pas encore à quels points du repère projecteur ils correspondent.

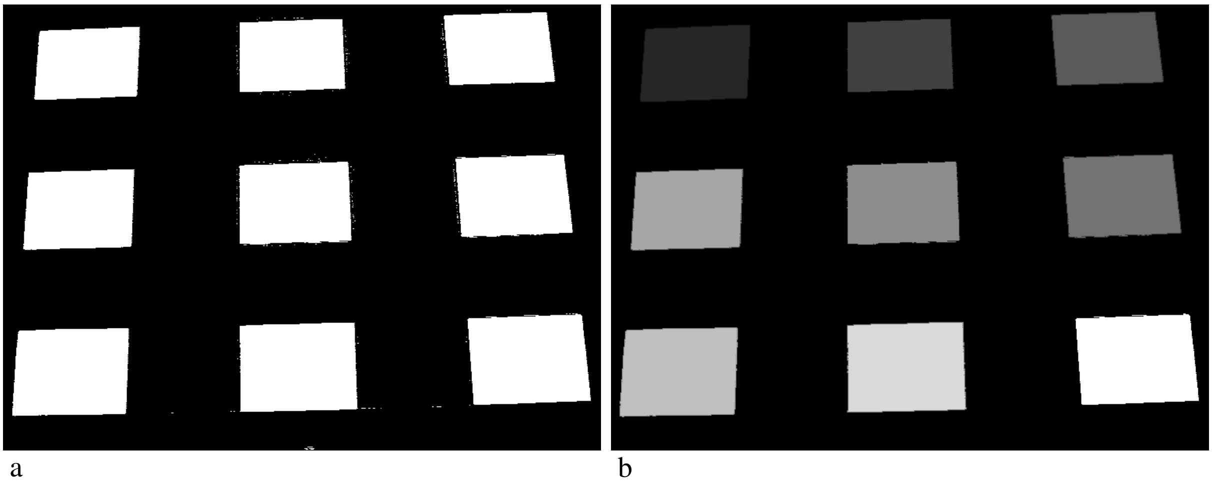


Figure 19 : Résultat de la différence d'images (a) et résultat de l'analyse de connexité (b) : chaque niveau de gris correspond à une zone connexe.

- On trie donc les point obtenus de façon à savoir à quels points de la mire projetée chacun correspond. Le tri utilisé est spécifique à la répartition des points dans la mire.
- La matrice A est maintenant connue, c'est une matrice 18*9 car nous utilisons 9 points. Nous somme dans le cas où, pour trouver la matrice de projection H, nous devons rechercher le vecteur propre correspondant à la plus petite des valeurs propres de $A^t.A$. Pour ce calcul nous utilisons les bibliothèques de Numerical Recipes.

Remarque : il est possible avec cette technique de calibrer la caméra sur n'importe quelle zone du tableau en projetant une mire sur cette zone et en positionnant la caméra pour qu'elle la visualise en plan serré.

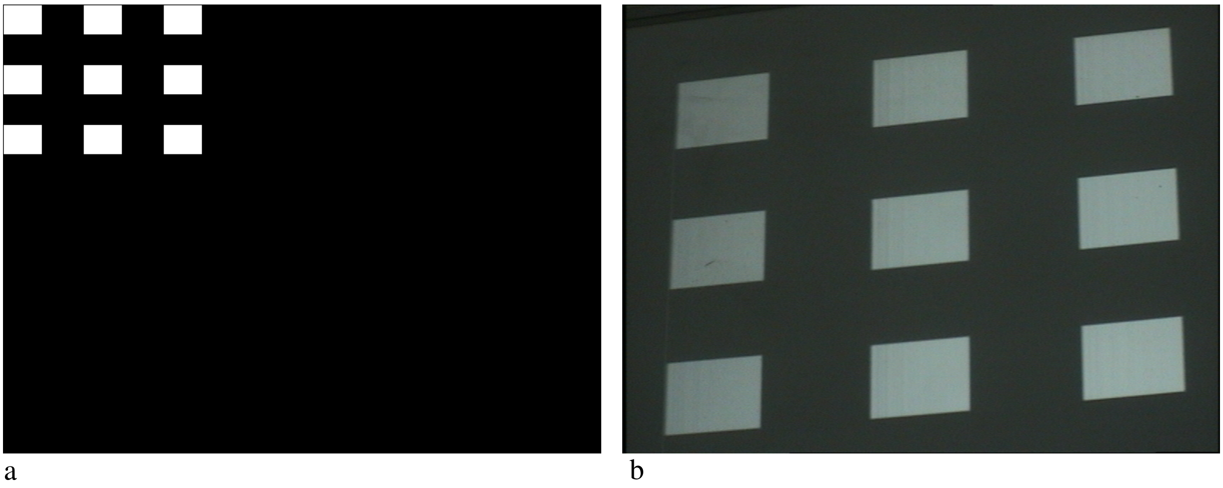


Figure 20 : Calibrage de la caméra sur une zone du tableau. Mire projetée sur la zone visée (a), et image caméra en plan serré sur cette zone (b).

Quelques remarques sur la mire utilisée :

- Nous utilisons une mire de 9 rectangles blancs sur un fond noir équitablement répartis sur la zone à calibrer. Les couleurs noires et blanches ont été choisies pour qu'il y ait un maximum de contraste et qu'ainsi la différence d'images donne le meilleur résultat possible.
- Les points de la mire sont de “gros” rectangles, afin que le calcul des coordonnées du barycentre se fasse sur de grande zones connexes et soit donc plus précis.
- Nous laissons un espacement assez important entre chaque rectangle afin d'éviter que deux rectangles ne soient considérés comme une seule zone connexe après la différence d'images.
- Le fait d'utiliser plus de quatre points dans la mire permet de vérifier la cohérence des résultats obtenus. En effet si l'on utilise 4 points, la matrice de projection est exacte même si la localisation des points dans le repère caméra est mauvaise. Et on aura toujours $p'_1=H.p_1$, $p'_2=H.p_2$, $p'_3=H.p_3$, $p'_4=H.p_4$. Par contre, en utilisant 9 points, la matrice calculée ne transformera aucun des points parfaitement et il sera donc possible d'évaluer la validité de la transformation calculée en comparant les coordonnées des points de la mire dans le repère caméra et ceux calculées par la transformation. On peut ainsi détecter de trop grands écarts et les considérer comme des erreurs s'il existe i tel que $||p'_i - H.p_i||$ est supérieur à un seuil défini.

4.6.3 ETUDE DES PERFORMANCES

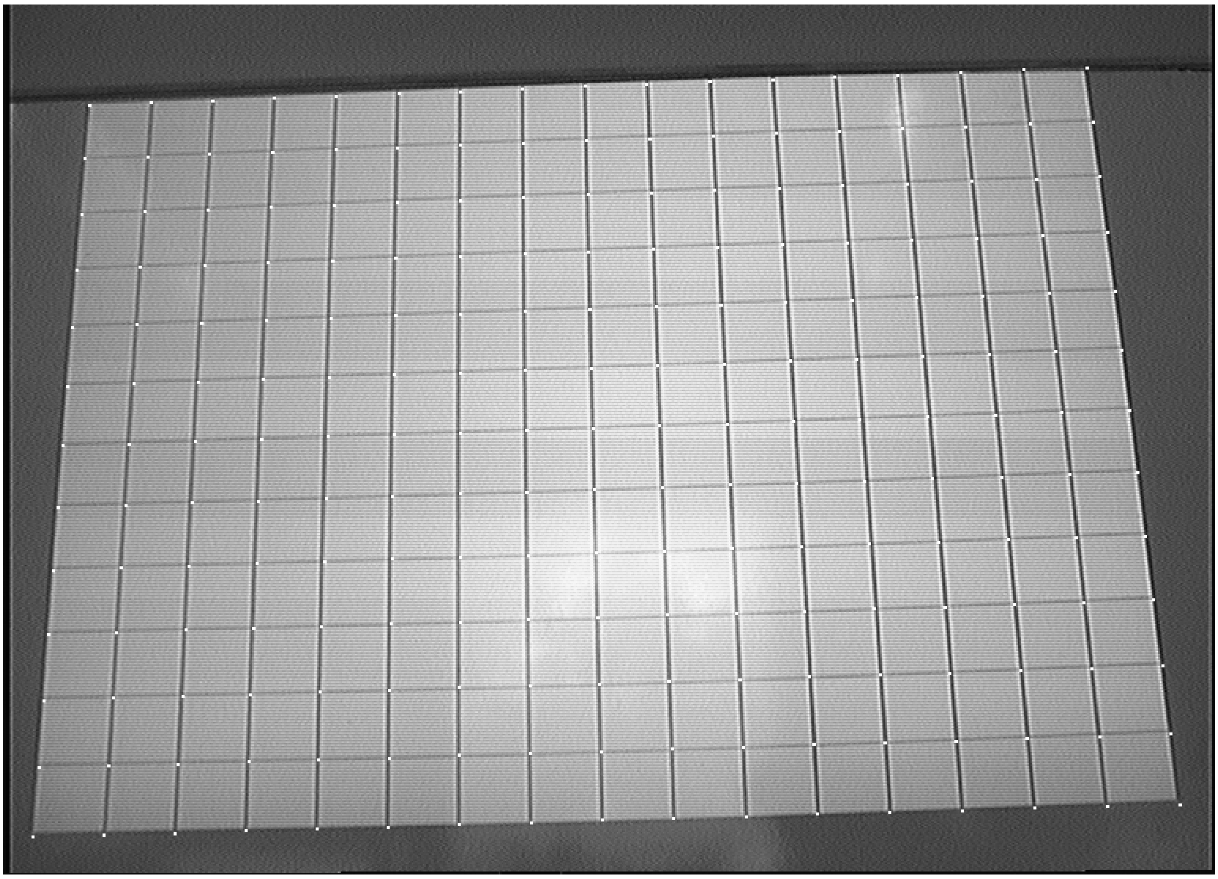


Figure 21 : Précision du calcul de la transformation projective.
Les points blancs correspondent au calcul de la projection des intersections de la grille.

La figure 21 illustre la précision du calcul de la matrice projective : on a projeté un quadrillage noir sur le tableau (une ligne tous les 50 pixels). Ce quadrillage définit 221 points dans le repère projecteur ((0,0), (0,50), (0,100), ..., (800,600)). Les points blancs correspondent à la transformation de tous ces points dans le repère caméra par la matrice de projection calculée dans cette disposition.

On peut remarquer que les points blancs sont très proches des intersections de la grille. Au maximum, il y a un pixel d'écart en abscisse et en ordonnée. On peut donc conclure que le modèle de la transformation projective plane est un modèle valide et que la technique utilisée pour trouver la correspondance entre les points des deux repères est fiable.

4.7 LE MOSAÏCING

4.7.1 PRINCIPE

Le mosaïcing est une technique de recomposition d'une image globale à partir de plusieurs images. Nous utilisons cette technique pour obtenir une image haute résolution et nettoyée du tableau avec son encre.

Dans l'exemple qui va suivre, nous avons découpé le tableau en 9 zones égales (3*3). Nous capturons une image de chacune de ces zones avec la caméra en plan serré, puis nous l'insérons dans l'image haute résolution globale.

4.7.2 MISE EN OEUVRE

Les étapes suivantes sont répétées pour chacune des 9 zones du tableau:

- La caméra est positionnée en plan serré sur la zone du tableau considérée (cf pilotage de la caméra).
- Elle est calibrée sur cette partie du tableau. La matrice calculée est la matrice de projection entre le repère caméra et le repère de l'image haute résolution globale.
- L'image vue par la caméra est capturée (nous appelons cette image l'image 1).
- Elle est nettoyée (le resultat est appelé image 2).
- Les points correspondant à de l'encre sont ensuite projetés dans l'image haute résolution globale. Pour cette dernière étape, on parcourt l'image nettoyée. Quand un pixel (x',y') correspond à de l'encre, on calcule ses coordonnées (x,y) dans le repère de l'image haute résolution globale à l'aide de la matrice de projection H obtenue lors du calibrage. Ce point (x,y) est affecté en fonction de la luminosité du pixel (récupérée dans l'image 1).

4.7.3 ETUDE DES PERFORMANCES

TAILLE DE L'IMAGE HAUTE RÉSOLUTION

Cette taille dépend du découpage du tableau. Pour un découpage n*n choisi, la taille maximum de l'image haute résolution à laquelle on peut aboutir est n*768*n*576 (la taille de l'image capturée par la caméra étant 768*576).

En pratique, étant donné que la caméra ne peut pas cadrer exactement la zone voulue, nous supposons que la surface utile n'est que de 600*450 au lieu de 768*576.

Ceci nous donne pour un découpage 3*3 une image de taille 1800*1350 et pour un découpage 6*6, une image 3600*2700. Ce dernier découpage est le plus grand autorisé par la caméra dont nous diposons. En effet, pour aller au- delà et continuer à augmenter la résolution, il faudrait rapprocher la caméra beaucoup trop près du tableau (moins de 1.5 m.).

Les figures des pages suivantes montrent le résultat d'une capture haute résolution par mosaïcing pour un découpage 3*3. On a colorisé l'image résultante suivant un modèle de couleurs (voir Annexe E : "Le modele de couleurs").

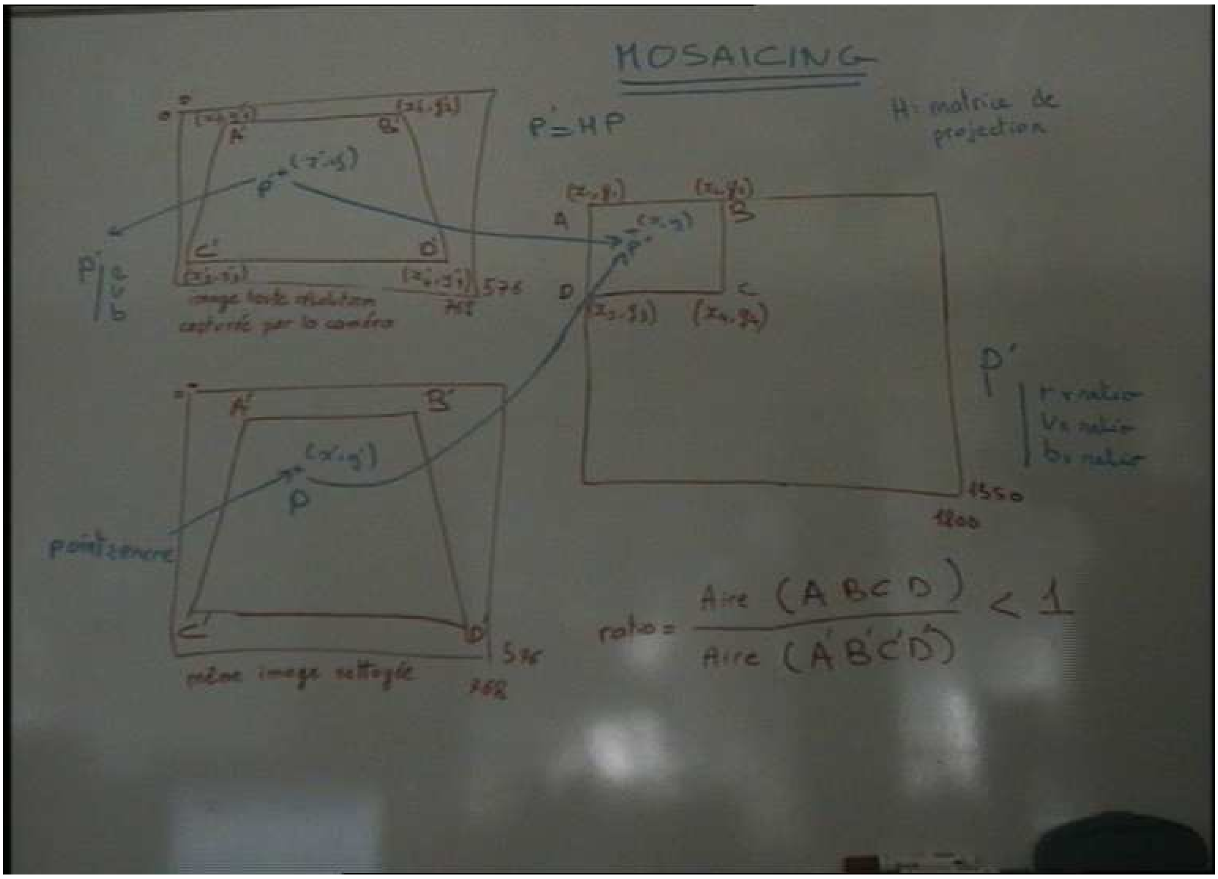

MOSAICING
P'=HP
H: matrice de projection
A'
B'
C'
D'
P'
576
768
image haute résolution capturée par la caméra
A
B
C
D
1550
1200
point zéro
même image nettoyée
ratio = Aire (ABCD) / Aire (A'B'C'D') < 1

Figure 22 : contenu du tableau vu par la caméra en plan large

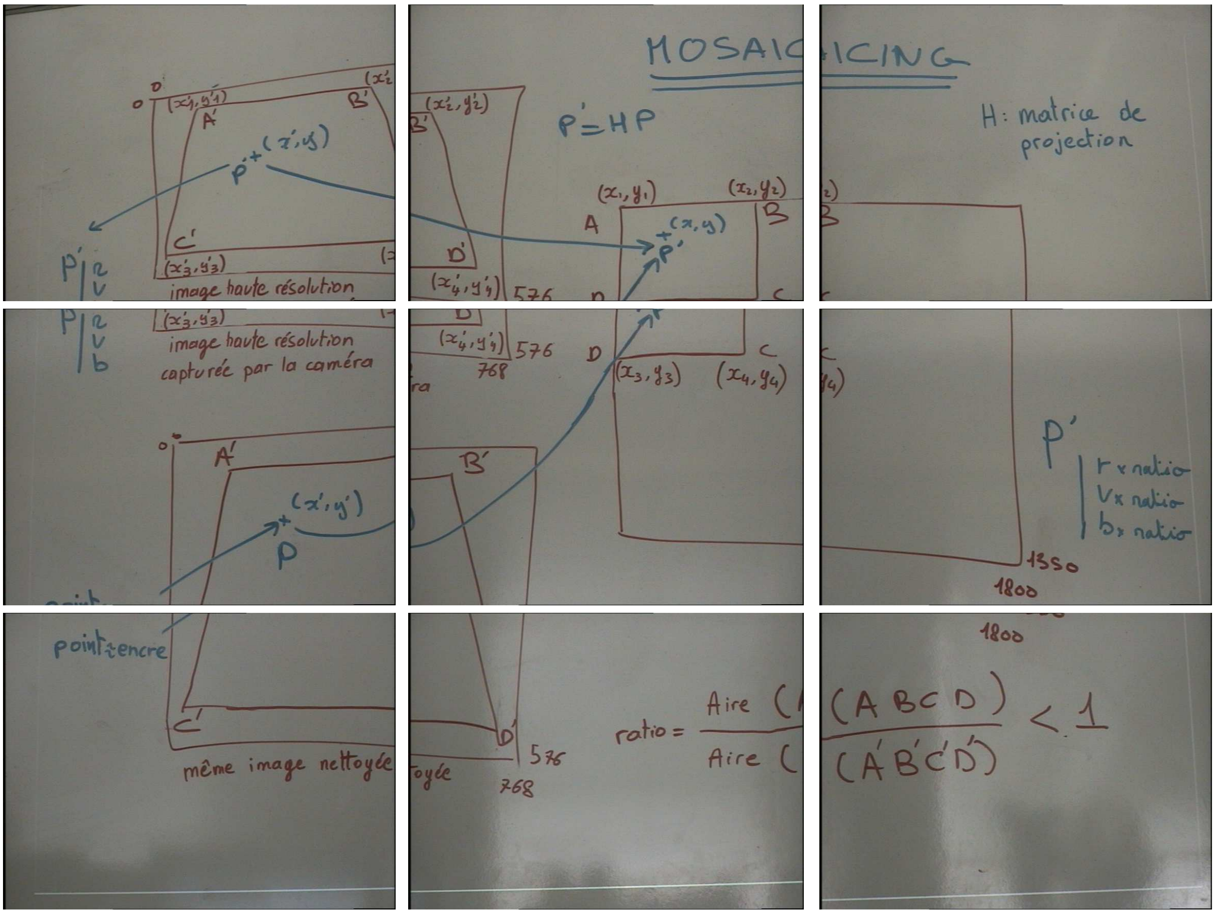

MOSAICICING
P'=HP
H: matrice de projection
(x1, y1)
(x2, y2)
A
B
C
D
(x, y)
P'
(x3, y3)
(x4, y4)
(x'1, y'1)
(x'2, y'2)
(x'3, y'3)
(x'4, y'4)
A'
B'
C'
D'
(x', y')
P
P'
r
v
b
image haute résolution capturée par la caméra
576
768
point encre
même image nettoyée
P'
r x ratio
V x ratio
b x ratio
1350
1800
ratio = Aire (ABCD) / Aire (A'B'C'D') < 1

Figure 23 : Captures haute résolution de chacune des 9 zones du tableau.

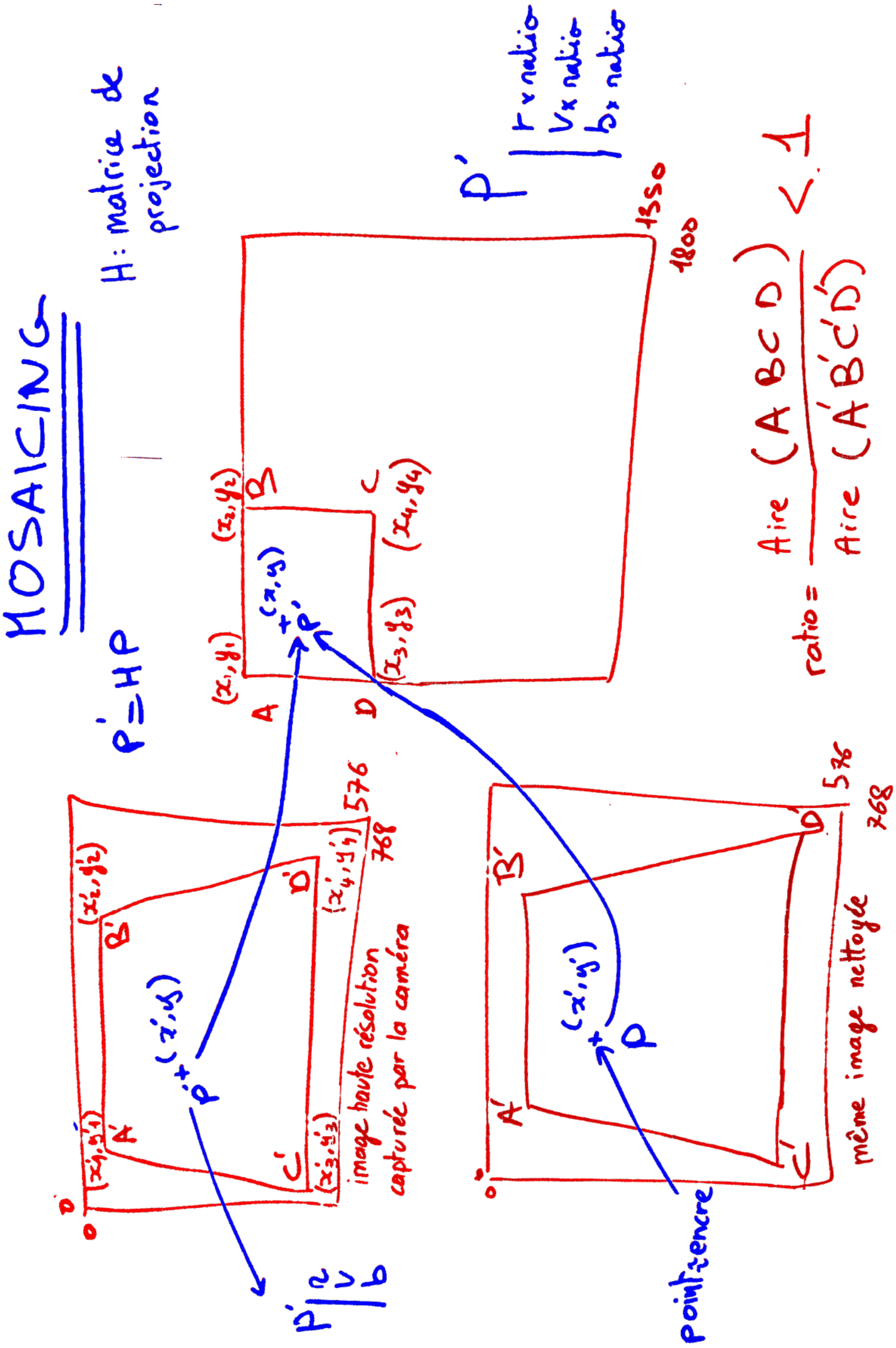

MOSAICING
H : matrice de projection
P' = HP
P'
r × ratio
V × ratio
b × ratio
1350
1800
A
B
C
D
(x1,y1)
(x2,y2)
(x3,y3)
(x4,y4)
(x,y)
P
ratio = Aire (A B C D) / Aire (A'B'C'D') < 1
576
768
O
A'
B'
C'
D'
(x'1,y'1)
(x'2,y'2)
(x'3,y'3)
(x'4,y'4)
(x',y')
P
image haute résolution capturée par la caméra
P'
r
v
b
même image nettoyée
point encre

Figure 24 : Résultat du moisaïcing du tableau en 36 parties.
La résolution de cette image est 3600*2700 points.

BILAN

5

Dans ce chapitre, nous exposons les résultats obtenus par rapport aux objectifs que nous nous étions fixés et notre contribution au projet.

5.1 OBJECTIFS ATTEINTS

Tous les objectifs dont l'ordre de priorité était l'obligation ont été atteints. Plus précisément :

- Le calibrage du système est suffisamment précis pour que le tableau soit utilisable.
- La capture et la numérisation du contenu (partiel ou total) du tableau en couleurs est opérationelle (la reconnaissance des couleurs n'était pas un objectif prioritaire mais a été implémentée). L'image numérisée peut être sauvegardée dans un fichier (format TIFF).
- Le suivi de doigt a un niveau de robustesse acceptable et une bonne vitesse de suivi. La pause est reconnue.
- Les outils de base que sont la sélection, la copie et le déplacement d'une zone rectangulaire du tableau sont disponibles. Ils permettent de manipuler aussi bien les inscriptions physiques que leurs équivalents virtuels.
- Deux systèmes distants peuvent communiquer et échanger des informations (les images de chaque tableau, les matrices de projection). Le protocole social basé sur les tours de parole a effectivement été établi.

5.2 CONTRIBUTION AU PROJET

Dans cette section, nous énumérons de manière détaillée tous les développements effectués lors de ce projet. L'ensemble de ces développements constitue notre contribution à ce projet. L'ordre d'énumération suit le niveau d'abstraction auquel se situent les éléments énumérés (les composants de bas niveau sont cités en début de liste, les composants de haut niveau vers la fin).

- Implémentation d'un algorithme de calcul de la matrice représentant la transformation projective (mise en correspondance des repères cartésiens de l'acquisition et de la restitution vidéo).

- Implémentation de l'algorithme de Wellner (voir “Nettoyage d'images” page 31) avec la possibilité d'effectuer plusieurs parcours de l'image à nettoyer (selon 4 sens : de haut en bas, de bas en haut, de gauche à droite en de droite à gauche).
- Développement d'un pilote de caméra mobile fonctionnant sur un port série, avec une API Tcl de haut niveau. Elaboration d'une méthode de positionnement automatique de la caméra pour le mosaïcing, quelle que soit la position de départ de la caméra (voir “Contrôle de la caméra” page 24).
- Implémentation d'un élément de synchronisation de scripts Tcl par rendez-vous.
- Développement d'un détecteur de mouvements par différence d'images. Ce composant permet de détecter les mouvements ayant lieu dans une zone donnée d'une image et de générer des événements en conséquence. Le seuil de détection est calculé de manière autonome en fonction des conditions lumineuses lors d'une phase de calibrage prévue à cet effet.
- Implémentation d'un composant de haut niveau effectuant un suivi d'objet par corrélation, avec gestion d'événements (mouvement, pause) (voir “Nettoyage d'images” page 31).
- Développement d'un système de création de menus virtuels projetés sur le tableau. Ce système permet de créer des menus graphiques projetés sur le tableau, utilisables avec les doigts, et d'associer des traitements à chacun des boutons composant le menu.
- Développement d'une API Tcl de haut niveau permettant de gérer des communications entre deux tableaux distants via un réseau.

5.3 PLANNING SUIVI

Dans cette section, nous rappelons brièvement le planning que nous avions prévu et nous exposons le planning effectivement suivi en justifiant les modifications vis-à-vis de nos prévisions.

La figure suivante illustre le planning prévisionnel. La ligne verticale indique la date prévue pour la présentation du premier incrément.

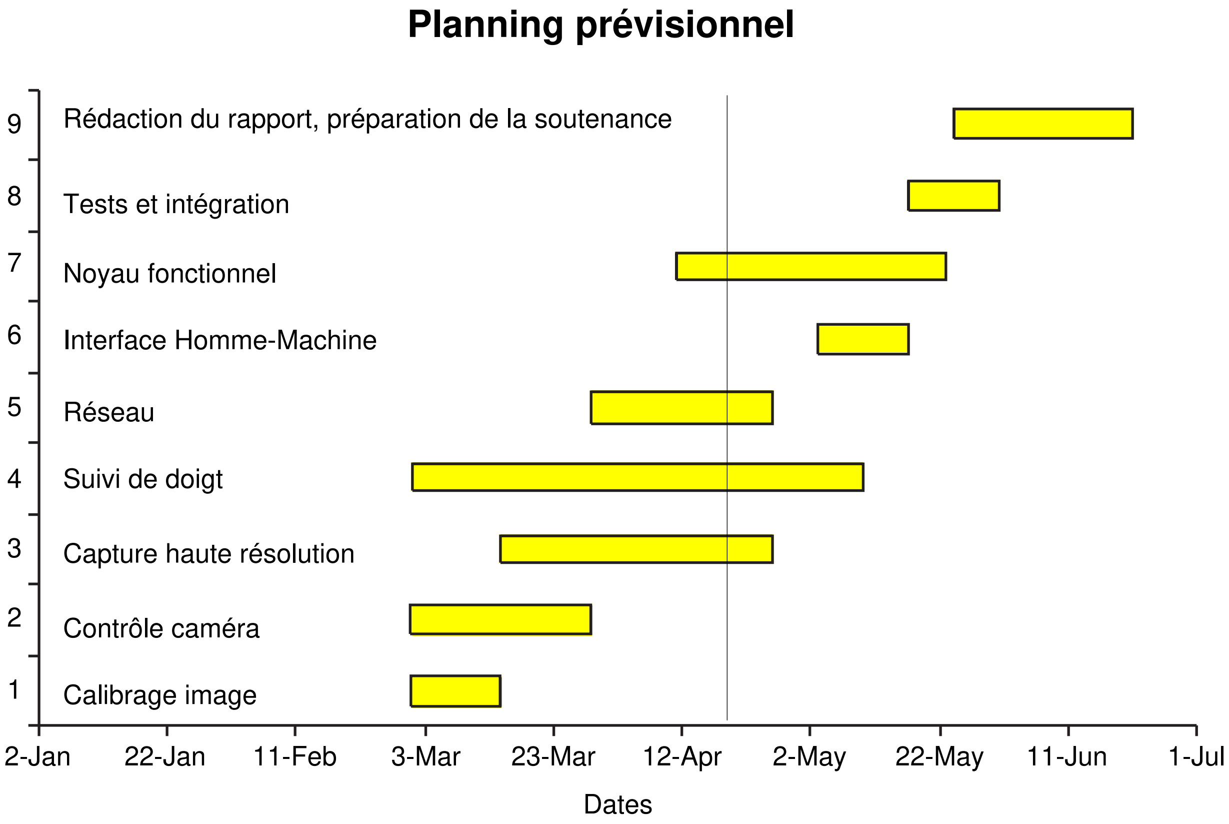


Figure 25 : Planning prévisionnel

Par rapport au planning prévisionnel, le planning effectivement suivi a subi les modifications suivantes :

- La date prévue pour la présentation de la Version 0 a du être repoussée d'une semaine.
- Le temps nécessaire à la réalisation de la partie réseau avait été surestimé. En effet, la réutilisation d'une bibliothèque de compression JPEG et l'utilisation des fonctions d'utilisation du réseau fournies par Tcl ont permis de réduire le temps de développement de la partie réseau de moitié par rapport au temps que nous nous étions imparti.
- Le temps initialiement prévu pour l'optimisation du suivi de doigt a été raccourci au profit du temps passé à l'élaboration de l'interface Homme-Machine, cette tâche s'étant avérée plus lourde et plus délicate à réaliser que ce qui avait été prévu.
- Pour les autres tâches, le planning prévisionnel a été respecté.

Finalement, le planning suivi est illustré par la figure suivante :

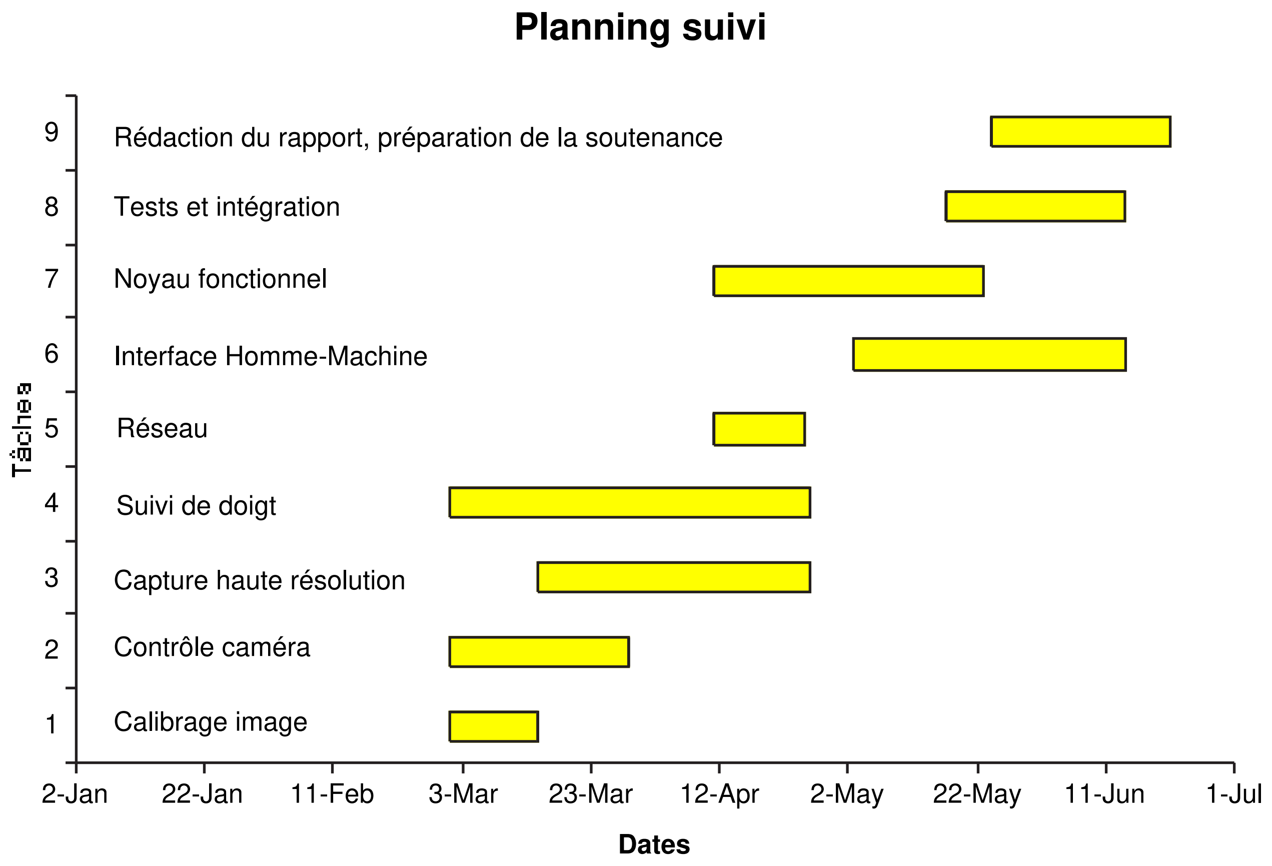


Figure 26 : Planning suivi.

CONCLUSION : LEÇONS ET PERSPECTIVES

6

6.1 LEÇONS DE L'EXPÉRIENCE

Le projet “Tableau Magique” a été l'occasion pour nous d'expérimenter pour la première fois dans des conditions proches de la réalité industrielle, la conception d'un projet informatique dans sa totalité. De cette expérience nous retenons les leçons suivantes :

- Le temps nécessaire à la rédaction des documents ne doit pas être négligé. En effet, si la rédaction de documents tels que le cahier des charges pouvait a priori sembler une perte de temps, cette phase d'analyse nous a beaucoup aidé par la suite et a permis de clarifier les points essentiels que sont les objectifs à atteindre et le planning à suivre
- Les objectifs que nous nous étions fixés étaient en réalité très ambitieux par rapport au temps disponible pour la réalisation du projet. Nous avons obtenu de bons résultats par rapport à ces objectifs; cependant, nous avons du limiter le temps consacré au développement de certains aspects du projet pour pouvoir livrer une maquette complète. La maquette du Tableau Magique réalisée est donc complètement fonctionnelle, certains composants n'ayant toutefois pas pu être optimisés vu le manque de temps.
- Le développement d'une maquette multi-plateforme semblait a priori une contrainte difficile à respecter. A posteriori, nous nous sommes aperçu que le fait de développer sur deux plateformes différentes nous a beaucoup servi. En effet, nous avons exploité les avantages de chacune de ces plateformes : l'architecture du PowerMacintosh nous offrait une puissance de calcul remarquable (ce qui est crucial lorsque l'on fait du traitement d'images), mais le système d'exploitation présentait des défauts que ne présentait pas IRIX (notamment, absence de mémoire partagée). Ainsi, l'une des plateformes a été dediée à la mise au point (Silicon Graphics O2) et l'autre à l'exécution (PowerMacintosh).
- Le choix des outils de développement a été déterminant. Le choix d'un langage interprété tel que Tcl pour la réalisation des composants de haut niveau s'est avéré bénéfique, surtout en ce qui concerne la mise au point : nous avons certainement gagné beaucoup de temps durant la phase de mise au point et de test grâce à Tcl, même si un langage interprété a toujours des performances plus faibles qu'un langage compilé.

6.2 PERSPECTIVES

La maquette du Tableau Magique que nous avons réalisée constitue une base qui est destinée à être enrichie dans les années à venir. La principale perspective que suscite la réalisation de ce projet est le développement de nouveaux paradigmes d'interaction utilisant cette technologie. Ces développements futurs pourront s'orienter selon deux axes principaux :

- L'intégration des concepts d'interaction Homme-Machine classiques : le Tableau Magique offrirait donc les services usuels proposés par un ordinateur de bureau. Pour cela, la reconnaissance de caractères et le développement d'une interface plus élaborée sont les améliorations à apporter en priorité.
- Le développement de méthodes d'interaction propres au Tableau Magique.

ANNEXES

LA DIFFÉRENCE D'IMAGES

A

La technique de différence d'images apour but de fournir une mesure du degré de différence entre deux images. Ce degré de différence peut être mesuré de deux manières :

- L'énergie de la différence, qui est une valeur numérique.
- Une image de différence représentant le degré de différence pour chaque point des deux images.

L'énergie de la différence entre une image I et une image J de même taille (m*n) est calculée comme suit :

$$E = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} (I(u, v) - J(u, v))^2 \qquad (25)$$

Comme le montre cette formule, l'énergie de la différence entre deux images est d'autant plus grande que les valeurs de deux pixels correspondant sont éloignées, et d'autant plus grande que les images sont de taille importante.

L'image de différence est une image en niveaux de gris, dont les valeurs des pixels sont calculées de la manière suivante (on note D l'image de différence) :

$$\forall(u, v) D(u, v) = |I(u, v) - J(u, v)| \qquad (26)$$

Le calcul de la différence entre deux pixels peut être fait soit sur la luminosité du point, soit sur un canal de couleur particulier (rouge, vert ou bleu).

ANALYSE FONCTIONNELLE

B

Dans cette annexe, nous présentons une analyse fonctionnelle du Tableau Magique. Nous détaillons ensuite les principaux composants et leur fonctionnement. La figure suivante illustre la décomposition en unités de traitement ou tâches.

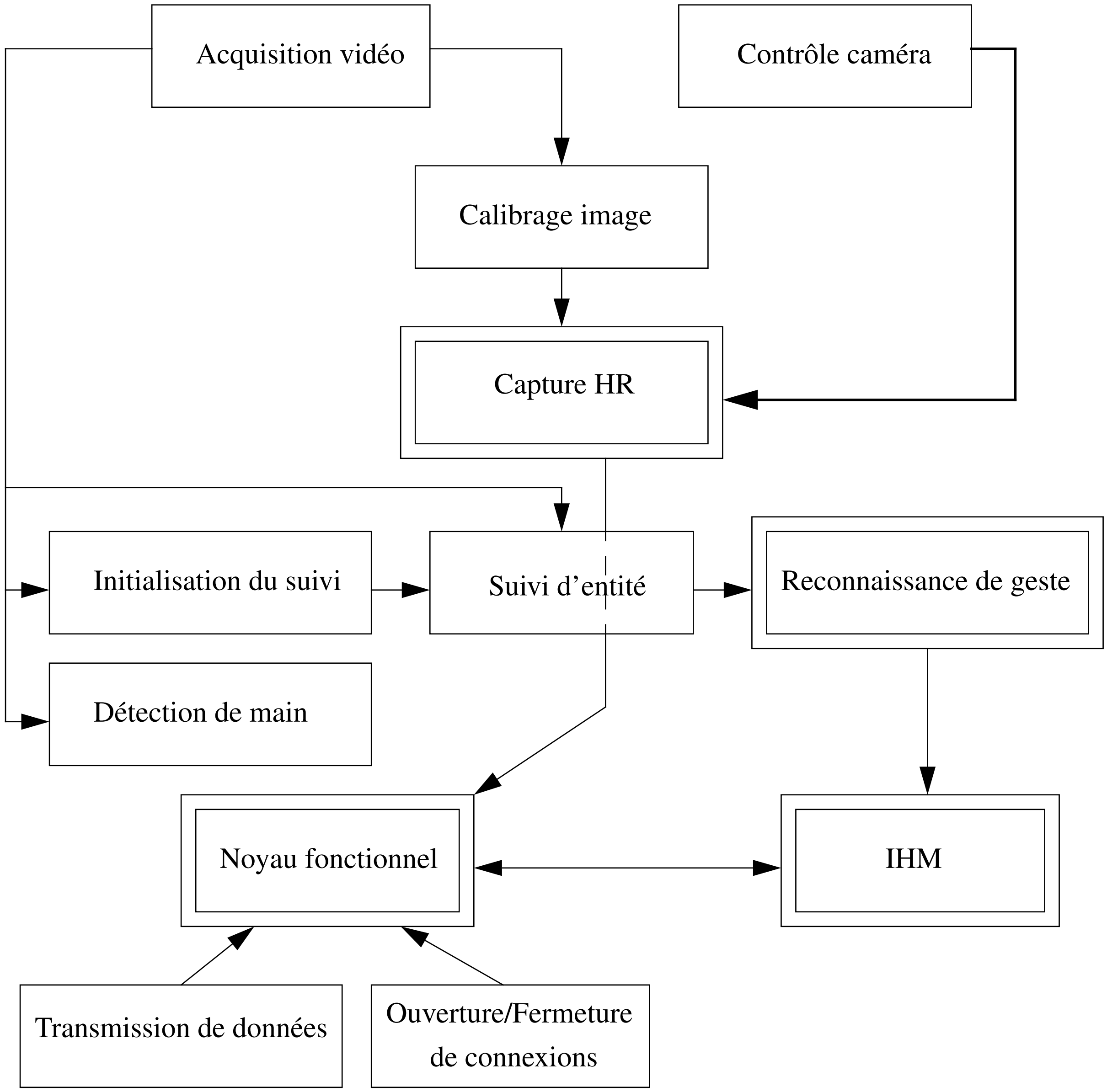

Acquisition vidéo
Contrôle caméra
Calibrage image
Capture HR
Initialisation du suivi
Suivi d'entité
Reconnaissance de geste
Détection de main
Noyau fonctionnel
IHM
Transmission de données
Ouverture/Fermeture de connexions

Figure 27 : Décomposition fonctionnelle

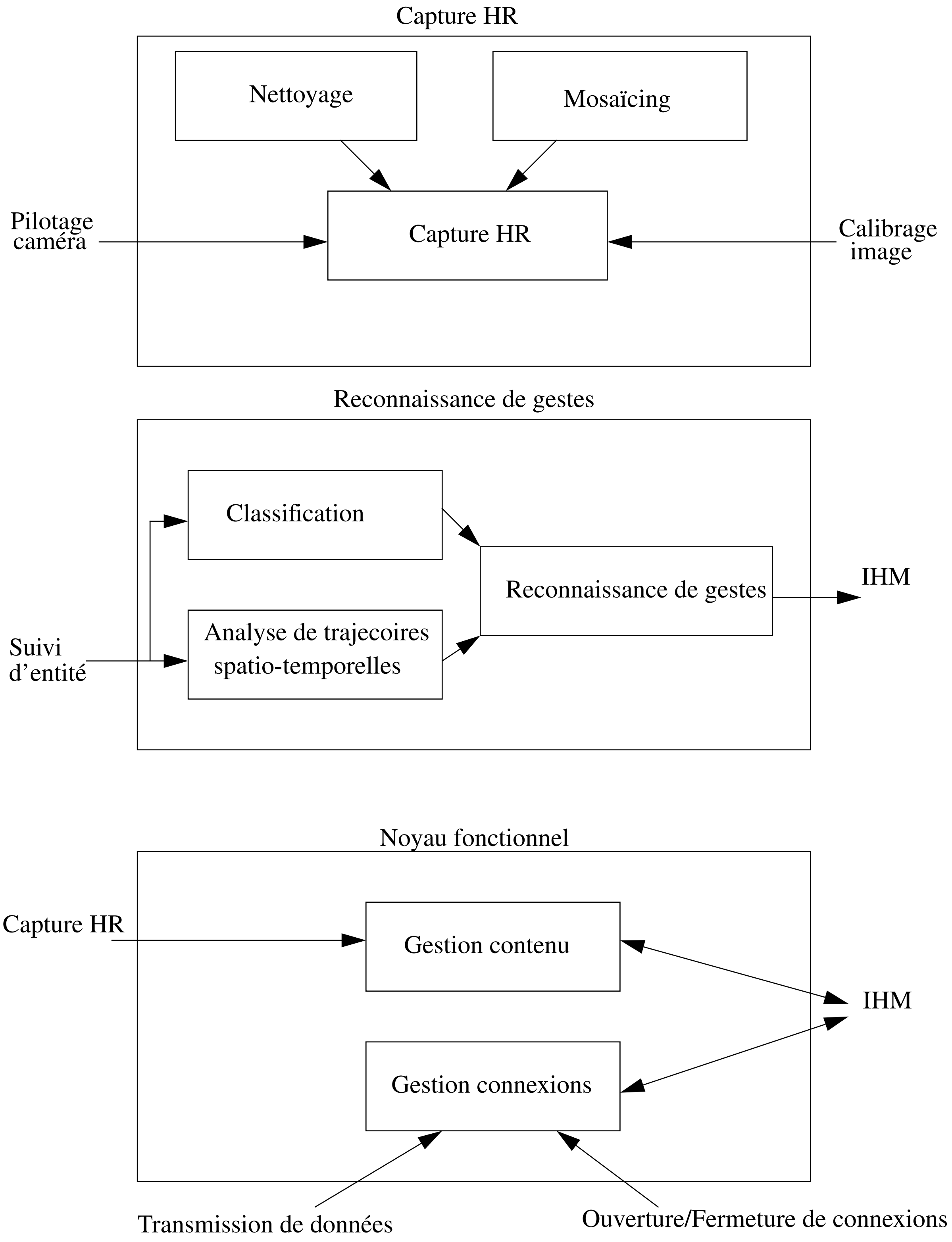

Capture HR
Nettoyage
Mosaïcing
Pilotage caméra
Capture HR
Calibrage image
Reconnaissance de gestes
Classification
Reconnaissance de gestes
IHM
Suivi d'entité
Analyse de trajecoires spatio-temporelles
Noyau fonctionnel
Capture HR
Gestion contenu
IHM
Gestion connexions
Transmission de données
Ouverture/Fermeture de connexions

Figure 28 : Détail de la décomposition fonctionnelle (capture haute résolution, reconnaissance de gestes, et noyau fonctionnel).

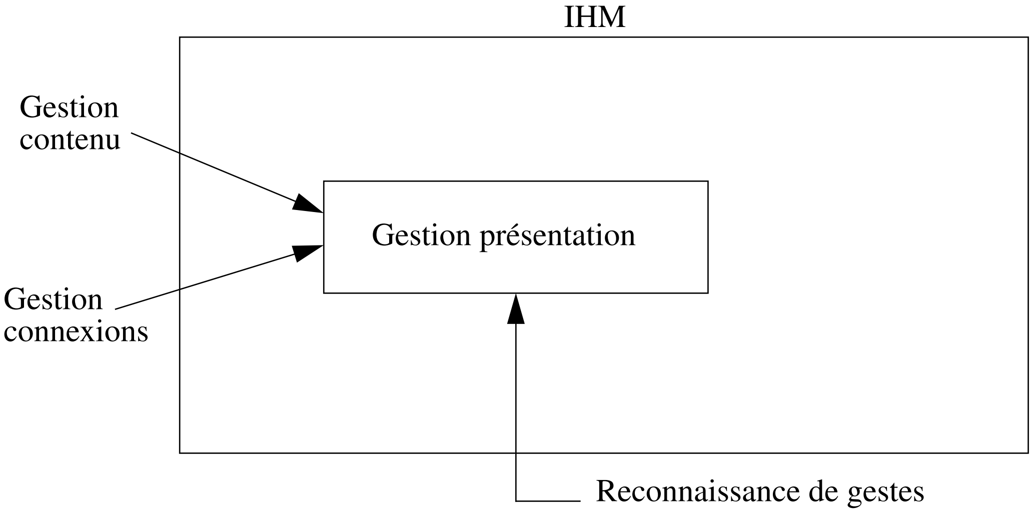


Figure 29 : Détail de la décomposition fonctionnelle (interface homme-machine).

La caméra produit des images que stocke le composant “acquisition vidéo”. Ces images sont consommées à la demande par les composants "Calibrage image" (dont le but est la mise en correspondance des repères cartésiens de l'acquisition et de la restitution vidéo), “Suivi d'entité”, “Initialisation du suivi” et “Détection de main”. Ces trois derniers composants fournissent les informations nécessaires à la reconnaissance de gestes, qui est le composant central sur lequel s’appuie le fonctionnement du Tableau Magique. Le fonctionnement de ces composants est expliqué en détail dans le chapitre concernant la réalisation.

La fonction de capture à haute résolution du contenu du tableau peut être décomposée en deux sous-fonctions que sont le "Nettoyage" et le "Mosaïcing" : En effet, le tableau a été “découpé” en plusieurs zones de taille et de positions connues par le système. Pour obtenir une image à haute résolution du contenu du tableau, on fait l'acquisition d'une image par zone (la caméra est en plan serré lors de l'acquisition, nous avons donc des images haute résolution). Après avoir nettoyé ces images (c'est-à-dire identifié les pixels correspondant à l'encre physique et les pixels correspondant au tableau lui-même, et supprimé de fait les impuretés dues au tableau), on recompose une image du tableau initial en rassemblant les images nettoyées. La capture haute résolution utilise les fonctions de pilotage de la caméra et de calibrage image.

Les images haute résolution sont gérées par le noyau fonctionnel, dont le rôle est la gestion du contenu virtuel du tableau et des connexions entre les divers tableaux distants. Les services de base qu'offre le réseau (l'ouverture et la fermeture des connexions et la transmission de données) sont directement exploités par le noyau fonctionnel. Il dialogue avec l'Interface Homme-Machine vers laquelle il génère des retours d'information (pour avertir l'utilisateur) et de laquelle il reçoit des informations relatives aux ordres de l'utilisateur.

L'interface Homme-Machine gère la présentation des divers services offerts par le Tableau Magique. Le principal moyen d'interaction avec le système sont les gestes de l'utilisateur, ce qui impose à cette interface de s’appuyer sur le composant de reconnaissance de gestes. Ce dernier fait appel au suivi d’entité qui fournit en permanence la position de l’objet suivi. Le composant “Reconnaissance de gestes” effectue une analyse de la trajectoire spatio-temporelle de cet objet et peut donc générer des événements à destination de l’IHM, qui, en collaboration avec le noyau fonctionnel, exploite ces événements pour mettre à disposition des utilisateurs les fonctions du Tableau Magique.

Tous les composants du Tableau Magique pouvant être considérés comme des services de base en vision par ordinateur on été implémentés en C++ et intégrés à la bibliothèque d'outils de vision TclVision (extension de Tcl/Tk développée au CLIPS par François Bérard). Tous les autres composants, plus spécifiques au Tableau Magique, tels que l'interface Homme-Machine et le noyau fonctionnel, ont directement été élaborés à partir de scripts Tcl.

LE NOYAU FONCTIONNEL

C

Le noyau fonctionnel du Tableau Magique a deux fonctions principales : la gestion du contenu du tableau virtuel et la gestion des connexions entre deux systèmes distants (dans le cas d'un fonctionnement en réseau). Nous exposons ici le fonctionnement de ce noyau fonctionnel.

GESTION DU CONTENU DU TABLEAU VIRTUEL

Le tableau virtuel est composé de deux images : une image basse résolution qui est projetée en permanence sur le tableau physique, et une image haute résolution qui sera celle sauvegardée en fin de session si l'utilisateur le désire. Le noyau fonctionnel est chargé de la mise à jour de ces deux images.

Le noyau fonctionnel reçoit des directives de l'interface Homme-Machine, en fonction des souhaits de l'utilisateur. Il exploite ces directives pour effectuer les mises à jour nécessaires sur les deux images (basse et haute résolution), et génère un retour d'information à destination de l'IHM pour informer l'utilisateur de l'état dans lequel il se trouve.

Le noyau fonctionnel peut être vu comme un automate d'états finis. Les transitions correspondent aux ordres issus de l'IHM, et à chaque état correspond l'envoi d'un retour d'information particulier vers l'IHM. Il nous faut cependant distinguer deux types d'ordres : ceux issus du suivi de doigt, et ceux issus du menu (voir IHM -

Annexe). La figure ci-dessous représente les transitions dues aux ordres envoyés par le suivi de doigt.

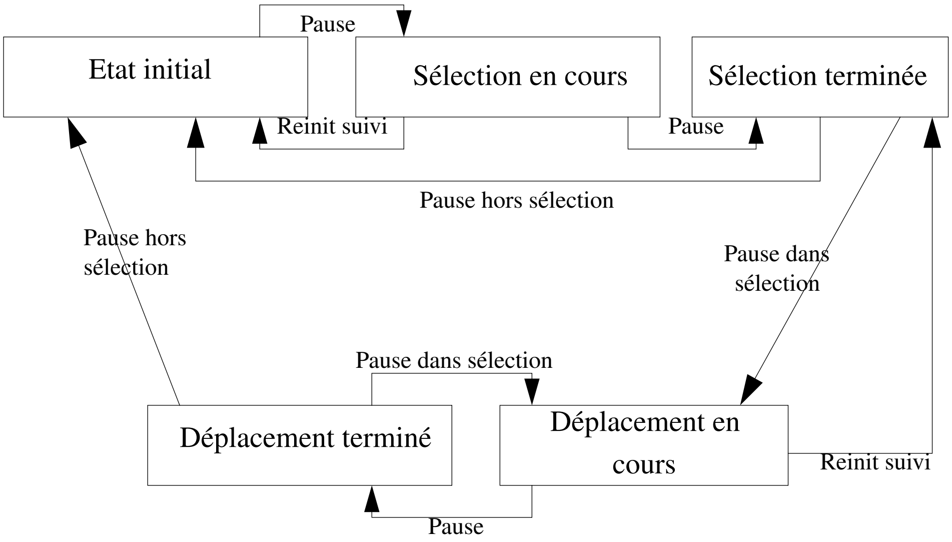

Pause
Etat initial
Sélection en cours
Sélection terminée
Reinit suivi
Pause
Pause hors sélection
Pause hors sélection
Pause dans sélection
Pause dans sélection
Déplacement terminé
Déplacement en cours
Reinit suivi
Pause

Figure 30 : Noyau fonctionnel

Cette figure illustre le fait que si une ré-initialisation du suivi de doigt intervient alors qu'une sélection ou un déplacement est en cours, on annule cette sélection ou ce déplacement et on revient à l'état précédent.

L'image basse résolution est mise à jour en permanence. Cela est nécessaire pour tenir l'utilisateur informé de l'état du système (sélections en cours, déplacement, suivi de doigt, initialisation du suivi ...). Le détail de ces retours d'information est donné en Annexe D : "L'interface Homme-Machine".

L'image haute résolution est mise à jour à chaque fois que l'utilisateur demande un nettoyage des sélections ou du tableau (aprés déplacement ou copie, les sélections sont affichées en basse résolution et sans couleur, elles servent uniquement à donner un aperçu de ce que sera le tableau après nettoyage). Lorsqu'un nettoyage est demandé, la technique de mosaïcing (voir "Le mosaïcing" page 43) est employée.

RÉSEAU

PRINCIPE

Le but de la version réseau du Tableau Magique est de permettre à deux personnes, situées en deux lieux distants, de travailler sur un même tableau virtuel.

Le protocole élaboré repose sur des tours de parole : à chaque instant, un des tableaux est en mode "rédacteur", l'autre en mode "lecteur". Un tableau en mode rédacteur peut être modifié par l'utilisateur. Son usage est exactement le même que pour un Tableau Magique s'exécutant simplement en local. En mode lecteur, le tableau ne peut être modifié. L'utilisateur se contente de regarder la vidéo en provenance du tableau distant. C'est l'utilisateur du tableau rédacteur qui choisit de passer la main. A cet instant, le contenu du tableau virtuel est transféré du rédacteur vers le lecteur et la parole passe à ce dernier : le rédacteur devient le lecteur et inversement.

MISE EN OEUVRE

La connexion entre les deux tableaux distants est établie grâce aux sockets TCP/IP de Tcl. L'implémentation du protocole décrit précédemment n'a posé aucun problème grâce aux fonctions réseau haut niveau de Tcl.

En ce qui concerne le transfert de la vidéo du rédacteur au lecteur, nous avons fait le choix de compresser chaque image avant de l'envoyer afin de limiter la charge du réseau. Cette compression est réalisée à l'aide d'une bibliothèque JPEG. Le fonctionnement du Tableau Magique nécessitant beaucoup de puissance de calcul, nous avons essayé de rendre “transparente” l'émission de la vidéo afin de ne pas altérer ses performances. La compression JPEG étant quelque chose de relativement coûteux, l'effectuer en une fois n'était pas possible car elle aurait accaparé la puissance de calcul trop longtemps. Nous avons donc découpé cette compression en “paquets de lignes” : quand elle peut prendre la main, un nombre donné de lignes de l'image est compressé puis le processeur est libéré. Une fois l'image entièrement compressée, elle peut être envoyée sur le réseau.

Côté lecteur, l'image est transformée à l'aide de la matrice de projection du tableau rédacteur (cf chapitre calibrage) reçue à l'ouverture de la connexion puis projetée sur le tableau. En prenant soin de faire cette transformation préalable, on assure que l'image coïncidera avec l'encre éventuellement présente sur le tableau.

PERFORMANCES

En divisant la compression JPEG en paquets de 5 lignes, nous avons pu la rendre invisible à l'utilisateur du Tableau Magique en mode rédacteur. La taille de l'image étant de 768*576, il faut 116 étapes pour compresser une image entièrement. Dans ces conditions, le débit vidéo atteint entre le rédacteur et le lecteur est d'environ 1 image toutes les 3 secondes.

Pour information, une image 768*576 compressée en JPEG avec une qualité de 25% occupe environ 10 kilo-octets. Non compressée, elle occupe environ 1,7 méga-octets. L'intérêt de la compression est donc évident quand on sait que l'on doit envoyer l'image par le réseau.

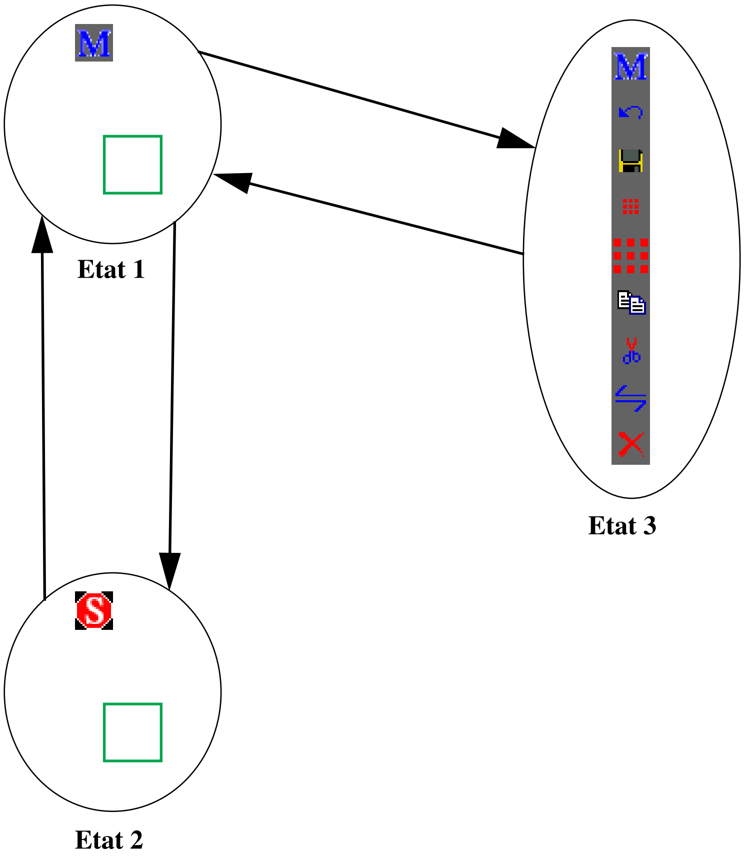
D

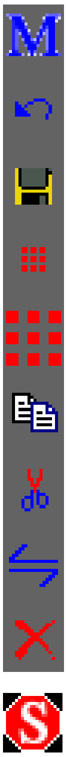
L'INTERFACE HOMME-MACHINE

Nous présentons dans cette annexe l'interface utilisateur du Tableau Magique.

LES DIFFERENTS AFFICHAGES SUR LE TABLEAU MAGIQUE

LES BOUTONS





- Bouton principal du menu (inactif quand le menu est déroulé)
- Annulation de la dernière opération de mise àjour
- Sauvegarde de l'image haute résolution du tableau
- Mise a jour des sélections
- Mise à jour du tableau dans son ensemble
- Copier
- Couper
- Changer le tour de parole (mode réseau uniquement)
- Fermer le menu
- Bouton STOP. Permet de couper un suivi de doigt ou d'afficher le bouton principal du menu

- **Transition de l'état 1 vers l'état 2** : Un suivi de doigt a été initialisé. Le menu disparait et laisse la place au bouton STOP.

- **Transition de l'état 2 vers l'état 1 :** Le bouton STOP a été activé : le bouton STOP disparait, le menu apparaît. Si un suivi de doigt était actif, le suivi est annulé et on reste dans l'état 2.
- **Transition de l'état 1 vers l'état 3 :** Le menu a été choisi : il se déroule. La zone d'initialisation du suivi de doigt disparait.
- **Transition de l'état 3 vers l'état 1 :** Un des boutons du menu a été choisi : la fonction associée est exécutée. Dés que cette dernière se termine, le menu réapparaît ainsi que la zone d'initialisation du suivi de doigt.

LES CURSEURS

Curseur normal
(positionnement gauche ou droite en fonction de la position des boutons sur le tableau)

Curseur "croix"
Une sélection est en train d'être créée.

Curseur "flèche"
Une sélection est en train d'être déplacée.

Curseur "sablier"
L'ordinateur est occupé.

L'INTERFACE DE DÉMARRAGE

Les deux figures suivantes illustrent l'interface via laquelle l'utilisateur configure la session au démarrage du Tabeau Magique.

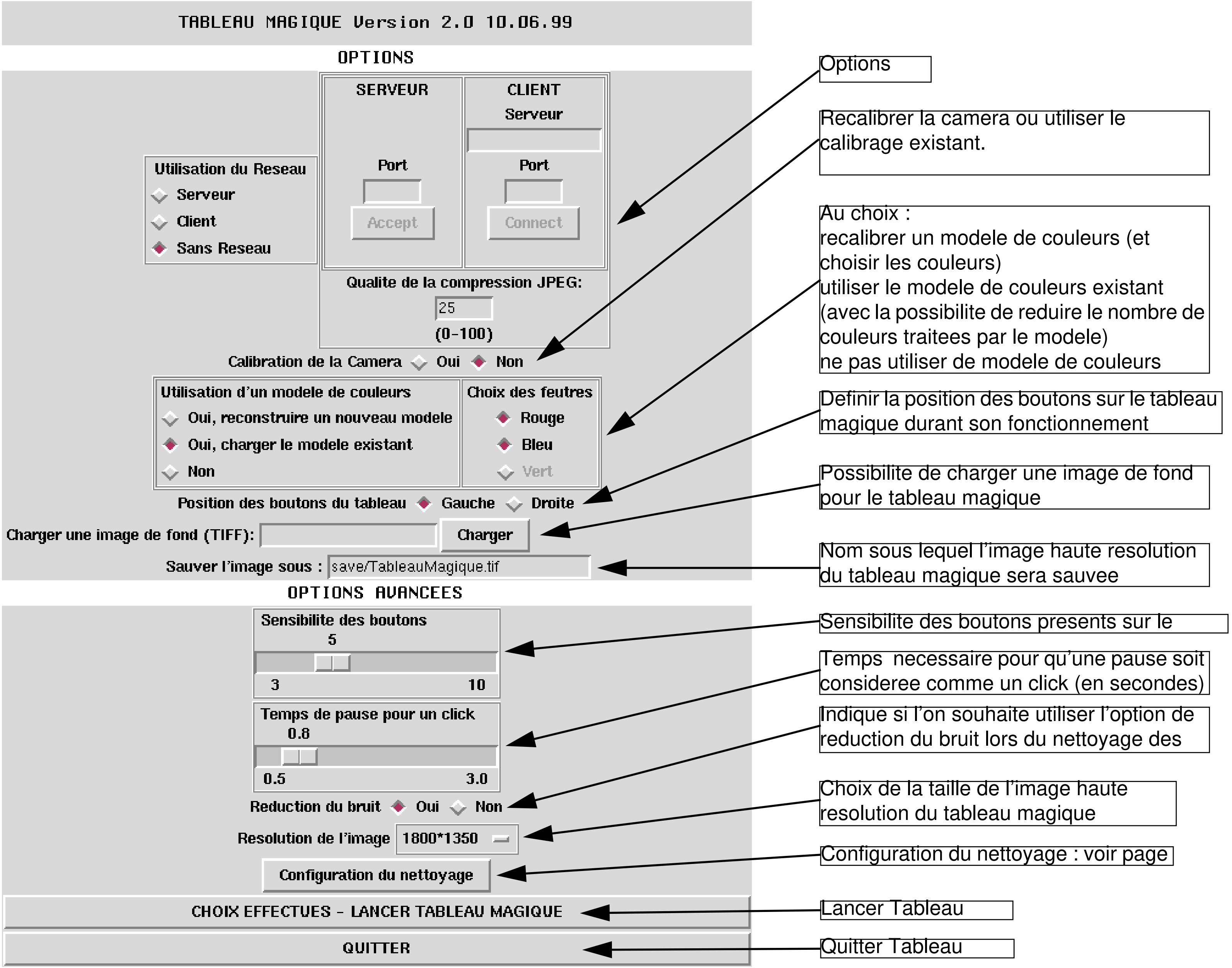


TABLEAU MAGIQUE Version 2.0 10.06.99

OPTIONS

Utilisation du Reseau
○ Serveur
○ Client
● Sans Reseau

SERVEUR
Port:
Accept

CLIENT
Serveur:
Port:
Connect

Qualite de la compression JPEG: 25 (0-100)

Calibration de la Camera ○ Oui ● Non

Utilisation d'un modele de couleurs
○ Oui, reconstruire un nouveau modele
● Oui, charger le modele existant
○ Non

Choix des feutres
● Rouge
● Bleu
○ Vert

Position des boutons du tableau ● Gauche ○ Droite

Charger une image de fond (TIFF):

Sauver l'image sous : save/TableauMagique.tif

OPTIONS AVANCEES

Sensibilite des boutons: 5 (3 – 10)

Temps de pause pour un click: 0.8 (0.5 – 3.0)

Reduction du bruit ● Oui ○ Non

Resolution de l'image: 1800*1350

Charger
Configuration du nettoyage
CHOIX EFFECTUES - LANCER TABLEAU MAGIQUE
QUITTER

Options

Recalibrer la camera ou utiliser le calibrage existant.

Au choix :
recalibrer un modele de couleurs (et choisir les couleurs)
utiliser le modele de couleurs existant (avec la possibilite de reduire le nombre de couleurs traitees par le modele)
ne pas utiliser de modele de couleurs

Definir la position des boutons sur le tableau magique durant son fonctionnement

Possibilite de charger une image de fond pour le tableau magique

Nom sous lequel l'image haute resolution du tableau magique sera sauvee

Sensibilite des boutons presents sur le

Temps necessaire pour qu'une pause soit consideree comme un click (en secondes)

Indique si l'on souhaite utiliser l'option de reduction du bruit lors du nettoyage des

Choix de la taille de l'image haute resolution du tableau magique

Configuration du nettoyage : voir page

Lancer Tableau

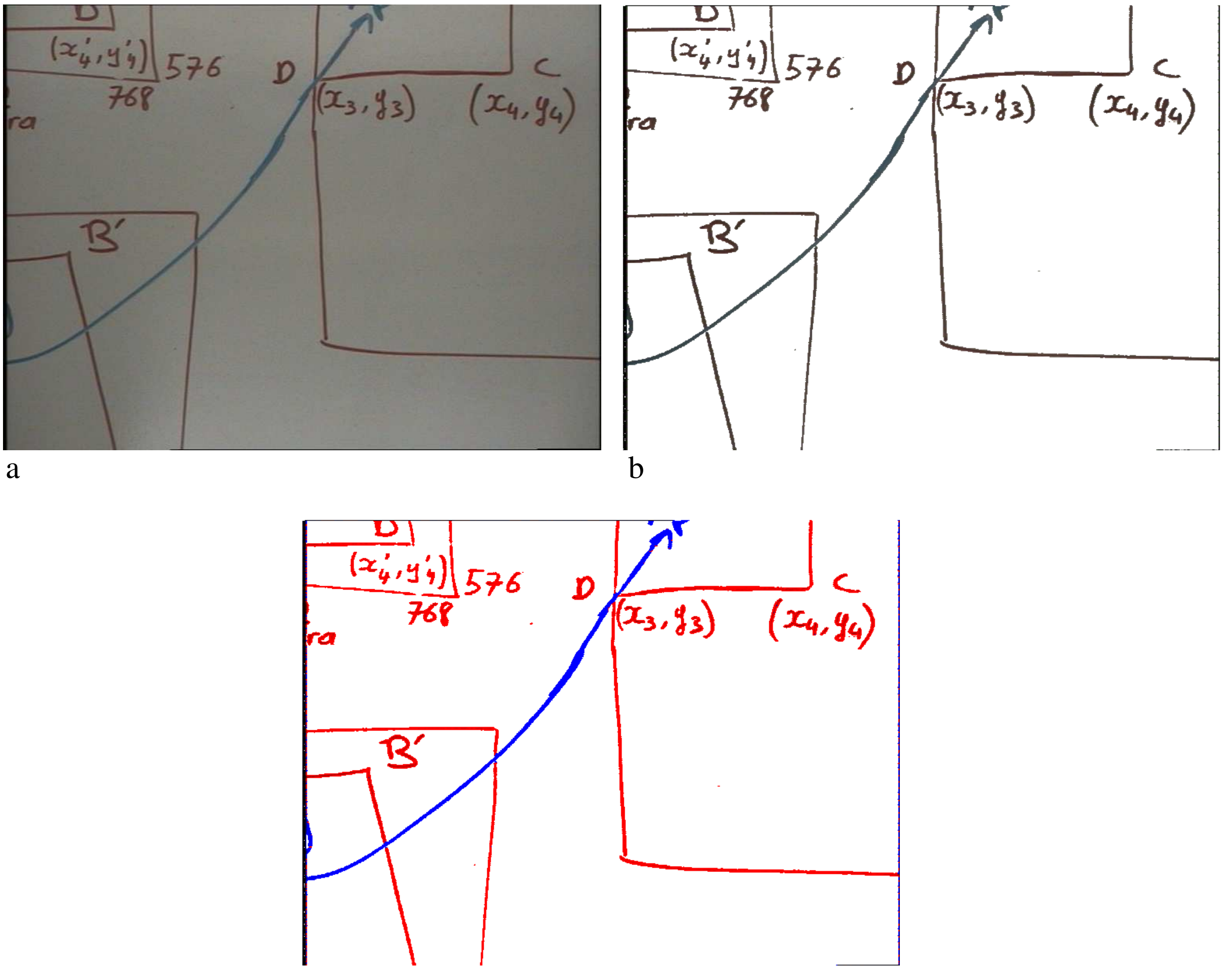
Quitter Tableau

E

LE MODELE DE COULEURS

PRINCIPE

Le but est, à partir d'une image couleur du tableau nettoyée sur laquelle on a remplacé le noir représentant l'encre par la couleur d'origine, d'obtenir une image avec des couleurs franches. Ceci est illustré sur la figure 31.



a b

c

Figure 31 : Application d'un modèle de couleurs.
(a) image capturée par la caméra, (b) image nettoyée avec les couleurs d'origine,
(c) image colorisée.

Cette opération est nécessaire car si l'on projette telle quelle l'image avec les couleurs d'origine, la distinction entre les couleurs n'est pas flagrante et les couleurs sont, de manière générale, trop sombres.

MISE EN OEUVRE

CONSTRUCTION D'UN «MODÈLE DE COULEURS»

Une phase d'initialisation est nécessaire avant de pouvoir obtenir une image comme celle ci dessus. Celle ci comprend la capture et le traitement d'un échantillon de chaque couleur à traiter.

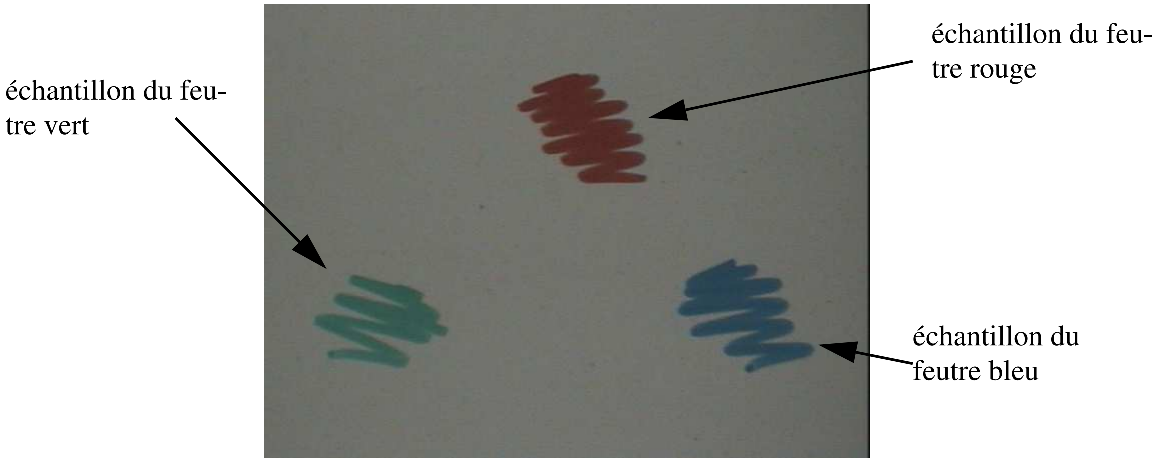

échantillon du feutre rouge
échantillon du feutre vert
échantillon du feutre bleu

Figure 32 : Image caméra sur laquelle on récupère les différents échantillons de couleur.

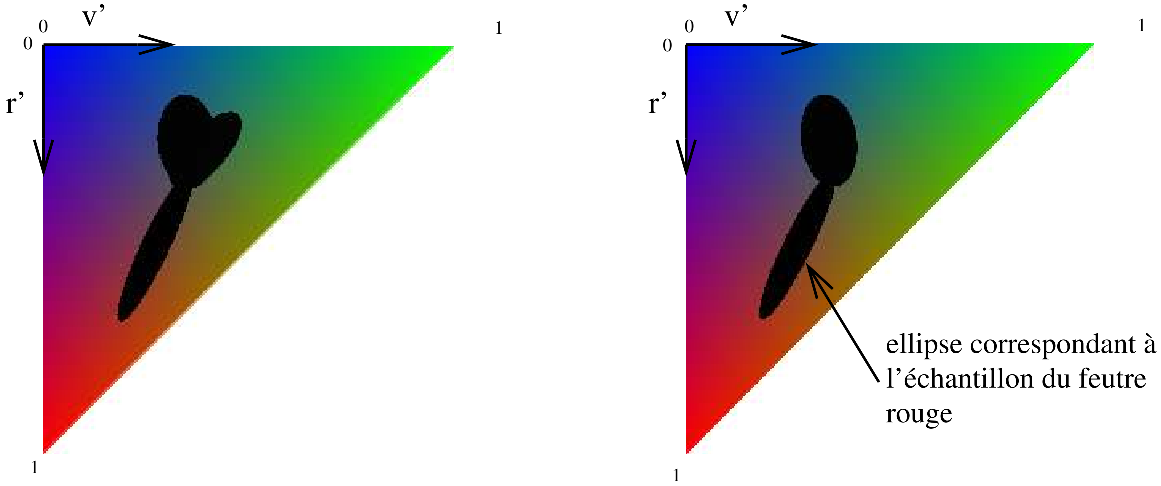
Pour chacun des points d'un échantillon, on récupére ses composantes rouges et vertes normalisées par la luminosité totale.

$$r' = \frac{r}{r+v+b} \qquad v' = \frac{v}{r+v+b} \tag{27}$$

On calcule ensuite, à partir de ces données, des informations statistiques sur chaque échantillon de couleur (ses moyennes $m_{r'}$ et $m_{v'}$ et sa matrice de covariance C) :

$$C = \begin{bmatrix} C_{r'} & C_{r'v'} \\ C_{r'v'} & C_{v'} \end{bmatrix} \tag{28}$$

Ces données statistiques constituent un modèle de couleur.



modèle de couleurs obtenu à partir des échantillons des 3 couleurs pris sur l'image précédente

même modèle mais construit avec les échantillons des feutres rouge et bleu uniquement

Figure 33 : Modèles de couleurs.

Les ellipses noirs correspondent à l'ensembles des point (v',r') pour lesquels la distance de Mahalanobis par rapport à l'un des échantillons est inférieure à 4.

$$\text{distance de mahalanobis}^2 = \begin{bmatrix}(r' - m_{r'})\\(v' - m_{v'})\end{bmatrix} \cdot C \cdot \begin{bmatrix}(r' - m_{r'})\\(v' - m_{v'})\end{bmatrix} \tag{29}$$

APPLICATION D'UN MODÈLE DE COULEUR À UNE IMAGE.

On parcourt l'image à coloriser et pour chaque pixel, on détermine de quel échantillon utilisé pour construire le modèle de couleurs il se rapproche le plus. On affectera alors ce pixel de la couleur choisie pour représenter l'échantillon correspondant.

L'échantillon le plus proche d'un pixel dont la couleur est caractérisée par ses deux composantes r' et v', est celui dont la distance de Mahalanobis au point (r',v') est minimale.

PERFORMANCES

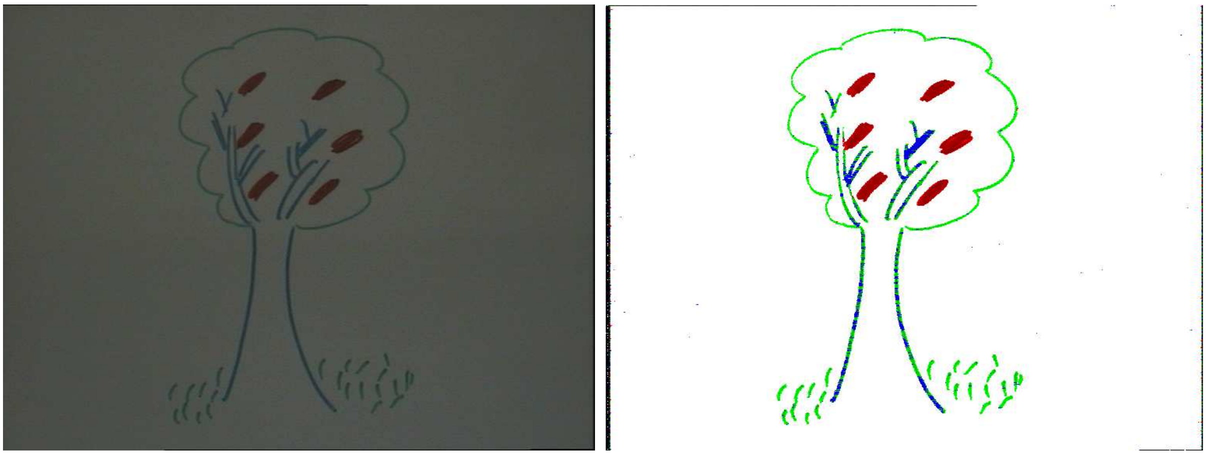


image caméra contenant de l’encre bleu, rouge et vert

même image nettoyée et colorisée avec le modèle de couleur précédent, construit avec 3 échantillons

Figure 34 : Résultat de l’application d’un modèle de couleur construit avec trois échantillons.

On constate qu’il y a beaucoup de pixels d’encre bleu qui ont été transformés en vert et l’inverse mais dans une moindre mesure. Cette mauvaise interprétation des pixels d’encre bleue est due au fait que, dans le modèle de couleurs appliqué (figure n°), les échantillons correspondant au feutres bleu et vert se superposent et certains pixels peuvent donc appartenir aux deux echantillons.

En pratique nous ne construisons donc le modèle de couleurs qu’avec deux échantillons de couleur (feutres rouge et bleu ou rouge et vert), on obtient ainsi de très bons résultats comme le montre la figure n°? au début de cette annexe.

DÉFINITIONS, ABRÉVIATIONS ET TERMINOLOGIE

F

- Adaptive Threshold : Algorithme de nettoyage d'image proposé par Pierre D. Wellner
- API : Application Programming Interface
- CLIPS : Communication Langagière et Interface Personne-Système. L'un des laboratoires de la fédération IMAG
- GRAVIR : Graphique, Vision, Image et Robotique. L'un des laboratoires de la fédération IMAG
- IHM : Interface Homme-Machine
- IIHM : Ingénierie de l'Interaction Homme-Machine
- IMAG : Informatique et Mathématiques Appliquées de Grenoble
- Mosaïcing : Technique d'acquisition d'image à haute résolution par découpage de l'image à capturer puis recomposition
- Pan : Positionnement horizontal de la caméra
- Suivi par corrélation : Algorithme de détection de forme en comparant par produit scalaire un motif de référence et les voisinnages d'une image
- TCP/IP : Transmission Control Protocol / Internet Protocol
- Tcl Vision : Bibliothèque d'outils de vision accessibles via une API Tcl
- Tilt : Positionnement vertical de la caméra

G

BIBLIOGRAPHIE

Bérard 94	F. Bérard, *Vision par ordinateur pour la Réalité Augmentée : Application au Bureau Numérique*, DEA d'Informatique, 1994.
Black 96	M.J. Black and A.D. Jepson, Eigen Tracking : *Robust Matching and Tracking of Articulated Objects using a View Based Representation*, ECCV 96, pp. 329-342, Springer Verlag, 1996
Card 83	S. Card, T. Moran, A. Newell : *The Pyschology of Human Computer Interaction*, Lawrence Erlbaum, 1983.
Cadoz 94	C. Cadoz, *Le geste, canal de communication Homme-Machine. La communication instrumentale*, Technique et Science de l'Information (TSI), Hermès Ed., vol 13, no. 1/1994, pp. 31-61.
Cipolla 96	R. Cipolla and N.J. Hollinghurst, *Human-Robot Interface by Pointing with Uncalibrated Stereo Vision, Image and Vision Computing*, vol. 14, pp. 171-178, mar. 1996
Cootes 93	T.F. Cootes, C.J. Taylor, A. Lanitits, D.H. Cooper and J. Graham, *Building and Using Flexible Models Incorporating Gray Level Information*, 4th ICCV, pp. 242-246, 1993.
Cootes 95	T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham, *Active Shape Models - Their Training and Applications, Computer Vision and Image Understanding*, vol 61, pp.38-59, jan. 1995.
Crowley 95	J.L. Crowley, F. Bérard and J. Coutaz, *Finger Tracking as an input Device for Augmented Reality*, FG'95. International Workshop on Automatic Finger and Gesture Recognition, Zurich, juillet 1995.
Crowley 97	J.L. Crowley and J. Martin, *Visual Processes for Tracking and Recognition of Hand Gestures*, International Workshop on Perceptual User Interfaces, Banff, Ca, October 1997.
Crowley97a	J.L. Crowley, *Mathematical Foundations of Navigation and Perception for an Autonomous Mobile Robot*, in *Reasoning and Uncertainty*, edite par L. Dorst, Springer Verlag, 1997.
Downton 91	A.C. Downton and H.Drouet, *Image Analysis for Model Based Sign Language Coding,* Proc. sixth Intl Conf on Image Analysis and Processing, pp. 637-644, 1991.
Gravila 95	D.M. Gravila and L.S. Davis, *Towards 3D Model Based Tracking and Recognition of Human Motion : A Multi-View Approach*, FGR 95. International Workshop on Automatic Face and Gesture Recognition, Zurich, juillet 1995.
Heap 96	T. Heap and D. Hogg, *Towards 3D Hand Tracking Using a Deformable Model*, FGR 96, Proc of Intl Conf on Automatic Face and Gesture Recognition, Killington, October 1996.

Koch 93	R. Koch, *Dynamic 3D Scene Analysis Through Synthetic Feedback Control*, IEEE-PAMI, Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 6, pp. 556-568, 1993.
Mackay 96	W. Mackay. *Réalité Augmentée : Le meilleur des deux mondes. Quand l'Informatique complète le mode réel au lieu de le remplacer*, numéro spécial La Recherche, no. 285, mars 1996, pp.32-37.
Maggioni 95	C. Maggioni, *GestureComputer - New Ways of Operating a Computer*, FG 95. International Workshop on Automatic Face and Gesture recognition, Zurich, juillet 1995.
Martin 94	J. Martin, *Techniques Visuelles de Détection et de Suivi de Mouvements*, Rapport de Magistère, LIFIA-IMAG, 1994.
Martin 97	J. Martin et J.L. Crowley, *An Appearence-Based Approach to Gesture Recognition*, 9th International Conference on Image Analysis and Processing, Florence, Italy, 17-19 Septembre 1997.
O'Rourke 80	O'Rourke and Badler, *Model Based Image Analysis of Human Motion Using Constraint Propagation*, IEEE-PAMI, vol. 2, pp. 522-536, 1980.
PD 97	P.D., *Un tableau blanc tactile*, Le Monde Informatique du 14 Novembre 1997.
Rekimoto 97	J. Rekimoto and N. Matsuhita, *Perceptual Surfaces : Towards a Human and Object Sensitive Interactive Display*, in Proc. 1rst Workshop on Perceptual User Interface, PUI97, Banff, 1997, pp. 30-32.
Saund 97	E. Saund, *Machine Perception of Instrumental Office Whiteboards*, in Proc. 1rst Workshop on Perceptual User Interface, PUI97, Banff, 1997, pp. 33-35.
Thalman 90	M. Magnenant-Thalman et D. Thalman, *Computer Animation : Theory and Practice*, Springer Verlag, 1990.
Wren 96	C. Wren, A. Azarbayejani, T. Darrel and A. Pentland, Pfinder : *Real Time Tracking of the Human Body*, FGR96, Proc of Int; Conf on Automatic Finger and Gesture Recognition, October 1996.