

# **LES TECHNOLOGIES DE L'IMAGERIE AU SERVICE DE L'ANALYSE DU MOUVEMENT EN PÉDAGOGIE DU PIANO\***

**Pierre Payeur, Martin Côté**

École d'ingénierie et de technologie de l'information, Université d'Ottawa

**Gilles Comeau**

Faculté de musique, Université d'Ottawa

---

## **Résumé**

L'enseignement du piano est une activité complexe combinant la technique, la cognition et l'expression artistique dans un cheminement visant à développer des aptitudes psychomotrices très raffinées. Les méthodes traditionnelles d'enseignement qui sont toujours en pratique aujourd'hui font peu appel à la technologie, et ce malgré son évolution au cours du dernier siècle. Il y a toutefois lieu de croire que la mise au point d'instruments de mesure adaptés à la pratique du piano pourrait contribuer à un apprentissage plus efficace pour un plus grand nombre d'élèves et à la prévention de nombreuses blessures en évitant la répétition de gestes inadéquats, notamment pendant les longues heures de répétition requises pour atteindre la maîtrise technique de l'instrument.

Cette publication propose une revue historique et technique des méthodes d'évaluation du mouvement dans le contexte spécifique de la pédagogie du piano. Les techniques d'imagerie par ordinateur sont plus spécialement examinées puisqu'elles sont parmi les seules à offrir la possibilité d'éliminer toute interférence avec le pianiste. Les technologies les plus avancées dans le domaine sont mises à l'épreuve dans un contexte où les contraintes imposées sur l'environnement et les individus doivent demeurer minimales. Un ensemble de résultats expérimentaux sont présentés et analysés pour mettre en lumière les avantages et les limitations des méthodes actuelles et une stratégie innovatrice de segmentation des séquences vidéo est proposée comme une composante fondamentale d'un dispositif informatisé d'assistance à l'enseignement du piano.

## **INTRODUCTION**

En dépit des efforts déployés au cours du dernier siècle pour mesurer et analyser les mouvements complexes effectués par les musiciens et les athlètes, l'ère de l'informatique et de l'intelligence artificielle n'a toujours pas réussi à proposer des techniques suffisamment flexibles pour opérer sans contraintes dans les environnements dans lesquelles ces performances sont exécutées. Les outils couramment utilisés par les

---

\* Les auteurs désirent remercier l'Université d'Ottawa pour le support financier fourni par l'entremise du Programme de développement d'initiatives interdisciplinaires ainsi que le Conseil de recherche en sciences naturelles et en génie du Canada qui ont rendu ce travail de recherche possible.

professeurs, entraîneurs et spécialistes de la santé ne peuvent généralement pas fournir une évaluation et une comparaison complète des gestes effectués. Une telle source d'information est pourtant hautement désirable pour mesurer l'évolution des habitudes et des pratiques dans un contexte d'apprentissage ou d'entraînement, afin d'améliorer l'efficacité des efforts déployés et l'exactitude des gestes répétés, et pour détecter et corriger des situations problématiques qui trop souvent dégèrent en blessures chroniques (Comeau, Payeur, Desjardins, Keillor et Bressler, 2004).

Bon nombre des méthodes de mesure usuelles employées en pratique professionnelle dépendent encore massivement de technologies encombrantes qui exigent l'installation de marqueurs ou de câblages sur l'individu dont les mouvements doivent être évalués (Trappe, Parlitz, Katzenberger et Altmüller, 1998 ; Boulic, Silaghi et Thalmann, 2000). Très souvent les mesures doivent aussi être prises dans des environnements parfaitement contrôlés, souvent très différents du milieu avec lequel les individus sont familiers. Ces facteurs ont une influence directe sur l'attitude, les gestes et la performance du sujet et corrompent l'exactitude des données recueillies, en plus de rendre la prise de mesures fastidieuse, difficilement accessible et dispendieuse.

L'objectif des travaux de recherche multidisciplinaires entrepris par notre équipe consiste à mettre à profit les expertises complémentaires de spécialistes en pédagogie du piano et en technologie de l'information afin de contribuer à l'évolution des techniques de mesure du mouvement sur des sujets humains exécutant des tâches complexes, tout en éliminant toute interaction directe avec le sujet. Ces méthodes de mesure sont dites non invasives puisqu'elles éliminent l'interférence créée par les instruments ou l'environnement sur les habitudes et la performance du sujet. Ultiment, les techniques développées doivent permettre à un pianiste de compléter une performance dans son environnement de travail habituel sans devoir porter de marqueurs ou de capteurs quelconques, de sorte que son jeu ne soit nullement influencé par la procédure de mesure en cours. Les données recueillies pendant la performance pourront alors être sauvegardées, analysées, comparées ou transmises à distance afin de permettre une évaluation quantitative ou une reproduction exacte des mouvements.

Ces enregistrements très riches en information trouvent de nombreuses applications pour l'enseignement du piano grâce à la rétroaction directe ou assistée par ordinateur au professeur et à l'élève, le téléapprentissage, la mesure non subjective de l'évolution du mouvement lors de la performance, l'étude approfondie de la gestuelle et de la posture dans l'analyse comparative des styles et des approches techniques, la prévention des pratiques comportant des risques pour la santé, la prévention des blessures, et la réadaptation. Un des objectifs du projet est de mettre au point un assistant à l'enseignement et à l'apprentissage du piano informatisé pouvant interagir en temps réel avec le professeur et l'élève afin de favoriser l'apprentissage et de prévenir le développement de mauvaises habitudes de gestuelle qui peuvent nuire au développement d'une bonne technique et ainsi

affaiblir la performance, et pouvant éventuellement conduire à des blessures graves (Zaza, 1995). Ce même dispositif servira aussi à l'évaluation objective et comparative de la performance technique et fournira une source d'information quantitative pour l'analyse biomécanique du geste dans l'analyse et la prévention des blessures reliées au jeu pianistique.

L'impact social que pourront avoir de tels outils de mesure n'est pas négligeable. La disponibilité de méthodes d'évaluation quantitative peut contribuer directement à l'amélioration continue des méthodes d'enseignement. Ceci pourrait permettre de mieux développer les talents des individus et d'en assurer la pérennité en orientant davantage le processus d'apprentissage vers des approches sécuritaires et moins répétitives. En effet, des études récentes ont démontré que le nombre d'individus souffrant de séquelles permanentes suite à des blessures reliées à la pratique de leur instrument dépasse les 60% chez les musiciens professionnels de formation classique, ce chiffre atteignant même les 45% chez les élèves (Zaza, 1995 ; Bressler, 2000). Les problèmes associés à la posture et aux efforts répétés sont identifiés comme les principales causes de blessures dans une proportion allant jusqu'à 98% des cas (Rogers, 1999 ; Hsu, 1997). Les coûts médicaux et les pertes personnelles associés aux blessures subies par les pianistes sont aussi considérables. Les dépenses encourues pour le traitement (consultations médicales, radiographies, chiropractie, acuponcture, chirurgie) de chaque individu peuvent facilement atteindre plusieurs milliers de dollars, sans compter la perte de talents précieux. Avec plus de 100 000 élèves diplômés chaque année des grandes écoles de musique au Canada et aux États-Unis selon les plus récentes statistiques, la motivation à fournir aux professionnels de la pédagogie et de la santé de meilleurs instruments de mesure et de prévention est évidente. Et encore, ces statistiques ne prennent pas en compte l'impact tout aussi significatif de mauvaises pratiques chez les athlètes amateurs et professionnels, non plus que sur les nombreux travailleurs qui exécutent quotidiennement des tâches répétitives.

Cette publication propose une revue historique et technique des méthodes d'évaluation du mouvement, des motivations ayant conduit à la mise au point d'une panoplie d'outils plus ou moins sophistiqués au cours des siècles et des raisons qui poussent encore les chercheurs d'aujourd'hui à raffiner les méthodes disponibles. Les méthodes d'évaluation par vision artificielle s'imposant de plus en plus dans ce domaine, une évaluation expérimentale des techniques d'imagerie qui sont à la fine pointe de la technologie a été réalisée dans le contexte spécifique de la pédagogie du piano. Les résultats obtenus sont présentés et analysés. Ces derniers permettent de conclure sur le potentiel réel des techniques d'imagerie mais mettent aussi en lumière les limites importantes qui sont imposées par la nécessité de réduire les contraintes sur l'environnement et sur les individus, afin de mettre au point des méthodes de mesure purement et véritablement non invasives.

## D'HIER À AUJOURD'HUI – LA MESURE DU MOUVEMENT

### Mesurer le mouvement – une longue tradition

Ce besoin de mesurer tant le mouvement que l'effort fourni dans l'exécution d'une oeuvre musicale trouve ses origines aussi loin qu'au début du XX<sup>e</sup> siècle alors que différentes études utilisant les moyens techniques de l'époque sont documentées dans la littérature. Les motivations associées à ces études sont diverses, allant de l'étude de la sonorité jusqu'aux évaluations biomécaniques, en passant par l'apprentissage du piano et l'amélioration des performances.

Faisant figure de pionnière en la matière, Marie Jaëll propose dès 1897 une étude approfondie des effets du contact des doigts avec le clavier sur la sonorité produite par un piano (Jaëll, 1897). La répercussion du touché sur l'apprentissage plus ou moins fructueux du piano est revue dans un contexte physiologique suggérant que certains mouvements minimaux mais « bien pensés » et contrôlés produisent de meilleurs effets musicaux qu'un apprentissage mécanique axé principalement sur la répétition. Cette étude visant l'analyse et la compréhension des mouvements des doigts et de la main exploite un système de mesure original basé sur les empreintes digitales. De petites pièces de carton sont placées sur les touches du piano et les doigts du pianiste sont encrés avant la performance afin de laisser des traces plus ou moins nettes sur les cartons en fonction de la pression exercée sur chacune des touches. Une analyse manuelle des courbes et de la largeur des empreintes permettait alors d'évaluer l'angle du toucher et son intensité, conduisant à une évaluation de la sonorité produite en fonction de la position des doigts et des mains du pianiste.

Quelques années plus tard, Otto Ortmann apporte également une contribution majeure à l'examen du fonctionnement et des effets des mécanismes psychologiques associés au jeu pianistique (Ortmann, 1929). Ses études approfondies sur la modélisation de l'anatomie du bras, de la main et des doigts mènent à une analyse quantitative de l'application des forces sur le clavier pour la production sonore. Les lois classiques de la mécanique sont appliquées pour comprendre le mouvement des bras et des mains pendant l'exécution. Cela conduit Ortmann à construire des prototypes mécaniques du bras et de la main humaine ainsi que des instruments de mesure originaux pour enregistrer les déplacements des doigts et des mains ainsi que la dépression des touches du clavier. L'enregistrement des déplacements latéraux est rendu possible grâce à un mécanisme de pantographe attaché par un élastique au poignet du pianiste. Le pantographe reproduit ainsi les mouvements à échelle réduite sur une pièce de papier placée latéralement et en retrait du clavier afin de minimiser les contraintes sur le mouvement. L'appareil nécessite toutefois des réglages minutieux et présente des faiblesses lorsque le poignet du pianiste bouge en profondeur dans le sens des touches du clavier.

Réalisant aussi la nécessité de mesurer des mouvements plus fins et en fonction du temps afin de percevoir les subtilités du jeu, Ortmann conçoit un second appareil de mesure, celui-ci dédié à l'estimation du déplacement des doigts. Ce mécanisme consiste essentiellement en un ensemble de cinq tiges d'aluminium ayant chacune une extrémité attachée à un doigt du pianiste. L'autre extrémité des tiges est placée en contact avec un cylindre rotatif recouvert de papier et sur lequel le mouvement des doigts est reproduit sous forme graphique à mesure que le cylindre tourne à vitesse constante. Comme le mouvement des doigts tend à être vertical alors que le cylindre tourne horizontalement, les variations en vitesse sont mesurées par l'amplitude de la déflexion des pointeurs sur le cylindre. Toutefois, ce mécanisme étant attaché tout près des points de contact des doigts avec le clavier, la résistance mécanique qu'il introduit s'avère problématique et nécessite un ajustement méticuleux pour limiter l'interférence ressentie par le pianiste.

Le développement primitif de ces outils ingénieux a permis de comprendre comment rendre possibles certains mouvements complexes tout en évitant les conditions conduisant à la fatigue musculaire excessive. Quelques décennies plus tard, Schultz poursuit dans la même lignée en utilisant aussi une approche « scientifique » pour étudier le jeu pianistique et l'effet du toucher (Schultz, 1949). Il est intéressant de noter que ces visionnaires signalent déjà dans leurs écrits l'existence d'une certaine réticence de la part des artistes et des professeurs à utiliser des méthodes techniques et scientifiques pour aider à comprendre les processus physiques, physiologiques et psychologiques complexes de l'apprentissage du piano.

### **Diversité d'instruments et de pratiques – unicité de la problématique**

Plus récemment, un nombre impressionnant de stratégies destinées à la mesure des gestes effectués par des artistes (Parlitz, Trappe, Dreschner et Altenmüller, 1999 ; Visentin et Shan, 2003) ou des athlètes (Sullivan, Ericksson et Carlsson, 2002) ont été décrites dans la littérature. Trappe et al. s'intéressent aux mécanismes sensoriels et à la motricité permettant à des percussionnistes d'apprendre des patrons de mouvement répétitifs (Trappe, Parlitz, Katzenberger et Altenmüller, 1998). L'étude est conduite sur des musiciens possédant différents niveaux d'expérience et cherche à identifier des stratégies développées par la répétition des mouvements chez les percussionnistes pour minimiser le mouvement et l'effort tout en améliorant la performance. La mesure des mouvements est concentrée entre l'épaule et l'extrémité de la baguette tenue par la main droite et s'effectue à l'aide d'un dispositif de vision thermographique. Ce système emploie de petits émetteurs de lumière infrarouge qui sont installés sur des points choisis du bras du percussionniste et de la baguette. Un capteur photosensible particulièrement réceptif au signal généré par les émetteurs fournit des images des points lumineux qui doivent ensuite être combinés pour recréer une estimation de la position du bras et de la baguette dans l'espace à trois dimensions. Ce type de capteur par vision active est particulièrement encombrant car

chaque émetteur est relié à une source d'alimentation par un câble. La mise en place du système est également relativement laborieuse.

Lammers et al. étudient pour leur part l'ampleur et la vitesse des mouvements du bras droit effectués par des trombonistes professionnels et des élèves, avec l'objectif de contribuer à la compréhension des mécanismes du jeu du trombone et à l'amélioration des méthodes pédagogiques reliées à cet instrument (Lammers, Kruger, Stoner, Fuller et Allyn, 1996). Afin de mesurer les mouvements du bras et du trombone, diverses technologies sont combinées. La coulisse du trombone est dotée d'un vernier mesurant le déplacement, la vitesse et l'accélération de la pièce mobile de l'instrument. Un marqueur circulaire est aussi attaché à l'extrémité de l'instrument et ses déplacements sont estimés par un détecteur de mouvement infrarouge. Des capteurs magnétiques sont attachés sur le bras du tromboniste pour mesurer les changements relatifs dans la position du coude et du poignet. Finalement, la technique d'électromyographie de surface est employée pour mesurer l'activité musculaire du musicien. La combinaison d'autant de capteurs procure une quantité d'information inégalée mais ajoute à l'encombrement imposé au musicien. Une bonne part des mesures du mouvement étant fournie par les capteurs montés sur l'instrument, la pose du bras doit alors être interpolée à partir de mesures indirectes, ce qui tend à réduire la précision des données.

Tarabella et al. proposent une façon originale de générer de la musique électroniquement en suivant le rythme imposée par les mouvements d'un chef d'orchestre (Tarabella, Carosi et Bertini, 1994). Dans un premier temps, le chef d'orchestre tient une baguette dotée d'une pointe lumineuse qui est perçue par une caméra vidéo. La marque lumineuse laissée sur l'image est extraite, ce qui permet d'estimer le déplacement latéral et vertical de la baguette et ainsi d'en extraire le rythme. Dans une seconde expérience, le chef d'orchestre porte des gants réflecteurs qui réagissent particulièrement bien à la lumière d'un projecteur dirigé vers le chef d'orchestre. Ces gants produisent des marques claires dans les images prises par une caméra vidéo, rendant l'extraction des mains relativement aisée.

Ces expériences illustrent une approche très répandue en imagerie par ordinateur où des marqueurs ou des pièces de vêtement très contrastant avec le reste de l'environnement sont manipulés ou portés par les individus dont on veut mesurer les mouvements. Ce type de solution peut fournir de bons résultats dans des environnements bien contrôlés, mais seulement à la condition que les participants puissent et acceptent de travailler avec ces éléments imposés. L'objectif principal de la présente recherche vise à éliminer ces contraintes, de sorte que le mouvement d'un individu puisse être mesuré librement, peu importe l'instrument considéré ou la fonction exercée.

### Méthodes de mesure modernes

Les techniques de mesure du mouvement humain décrites dans les exemples précédents et les méthodes couramment employées en pratique professionnelle tendent encore à être très restrictives, invasives et encombrantes étant donné que la majorité d'entre elles exigent la mise en place de marqueurs réfléchissants ou de capteurs câblés sur l'individu. De tels appareils imposent des limites physiques sur le sujet, ne sont pas confortables et impliquent une période de préparation relativement longue pour la mise en place. Le système *VICON* (Vicon Peak, 2005 ; Jobbagy, 2004) bien connu dans le milieu médical pour l'étude de la démarche humaine, et particulièrement utile en réadaptation et en orthothérapie, est un excellent exemple. Il utilise une approche dite de vision active qui consiste à installer de petites billes réfléchissantes en des points sélectionnés sur l'individu et à filmer leurs déplacements à l'aide de caméras vidéo dont les lentilles sont entourées de diodes photoémettrices dont le signal se réfléchit particulièrement bien sur les billes. Cette stratégie élimine le besoin de câblage directement attaché au patient mais limite l'observation aux points marqués par les billes qui doivent être installées préalablement. Palmer et al. ont fait appel à ce type de dispositif pour étudier la gestuelle et le processus d'acquisition de la technique dans le cas du jeu pianiste (Palmer et Drake, 1997 ; Palmer et Pfordresher, 2000). Une variante passive et beaucoup plus économique de cette approche est illustrée par Drouin et al. qui mesurent les mouvements d'un sujet vêtu d'une combinaison noire sur laquelle des balles de couleur orange sont attachées (Drouin, Hébert et Parizeau, 2003). Une caméra couleur standard permet alors de suivre le déplacement des balles hautement contrastantes avec la combinaison.

Lorsque de tels marqueurs ponctuels sont utilisés, la stratégie habituelle consiste à tenter de retrouver la position dans l'espace de chacun des membres du sujet en comparant la projection des marqueurs sur l'image avec les points qui résulteraient du mouvement d'un squelette virtuel sur lequel les marqueurs seraient installés aux mêmes endroits (Albrecht, Haber et Seidel, 2003). Des modèles cinématiques du squelette humain sont alors définis afin de simuler toutes les positions possibles de l'individu et de retrouver celle qui correspond le mieux à sa situation actuelle telle que perçue dans l'image des marqueurs. Malheureusement, le fait que les marqueurs soient attachés aux vêtements, ou au mieux à la peau de l'individu, produit des déplacements qui ne correspondent pas exactement aux mouvements du squelette. Une certaine déviation entre la mesure du mouvement estimé et le déplacement réel est ainsi introduite (Boulic, Silaghi et Thalmann, 2000).

Une étude récente des méthodes classiques d'imagerie par ordinateur pour l'évaluation du mouvement humain a démontré que les techniques cherchant à détecter des points de repère directement sur le sujet sans employer de marqueurs fonctionnent bien dans des environnements contrôlés offrant une arrière-scène uniforme (Wang et Singh, 2003). Les recherches actuelles visent à étendre ces méthodes de détection à des environnements non contraints où l'arrière-scène n'aura pas à être de couleur uniforme, où l'éclairage pourra

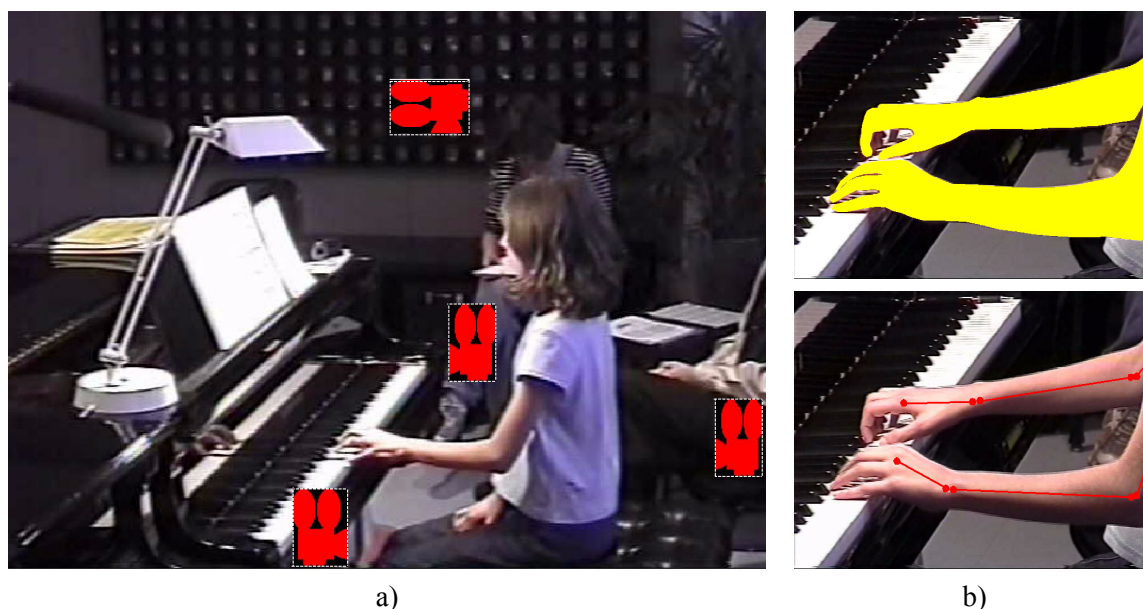
varier d'une région à l'autre de l'espace et où le sujet n'aura pas à porter de vêtements d'une couleur spécifique. La présente recherche s'inscrit dans ce contexte et vise à fournir au domaine de la pédagogie du piano des outils opérationnels qui élimineront les contraintes couramment imposées sur la liberté et l'étendue des mouvements du musicien tout en minimisant les manipulations requises par le professeur et l'élève pour la mise en place et l'opération du système de mesure.

## **SUR LA PISTE DES MÉTHODES DE MESURE NON INVASIVES**

### **Les défis**

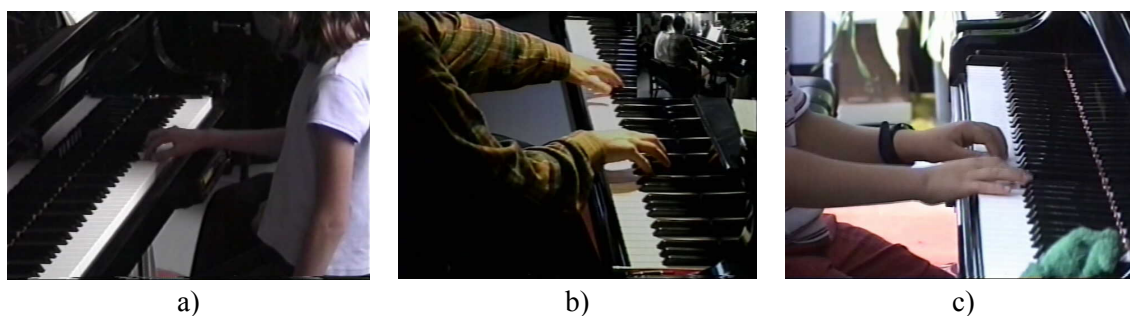
Parmi les techniques de mesure du mouvement considérées à ce jour dans la littérature, les approches basées sur l'imagerie passive se révèlent clairement être les moins invasives puisqu'elles permettent de capter le mouvement tout en gardant l'équipement à distance de l'individu. Pour être à la fois efficaces et utiles ces technologies doivent par contre pouvoir s'adapter aux caractéristiques de l'environnement dans lequel elles sont appelées à opérer. Dans une application en pédagogie du piano, tel que considéré dans le cadre de cette recherche, le dispositif idéal doit être en mesure de fonctionner dans divers lieux d'enseignement et de pratique aux caractéristiques différentes, notamment le studio d'enseignement et la salle de répétition, mais aussi la résidence de l'élève où la majorité des heures de pratique ont lieu. La figure 1 illustre la configuration typique d'un système étendu tel que proposé pour un studio d'enseignement. Le dispositif compte deux caméras disposées de part et d'autre du clavier pour estimer le mouvement des bras et des mains en élévation et en profondeur, une caméra au-dessus du clavier pour mesurer le déplacement latéral et une dernière caméra située à l'arrière du pianiste pour évaluer la posture du dos et de la tête. Le dispositif vise à permettre l'extraction des membres du sujet afin d'en déterminer la pose dans l'espace tridimensionnel grâce à une recherche des intersections entre les contours des zones d'intérêt perçues par chacune des caméras. Une version simplifiée mesurant un ensemble de paramètres restreint est envisagée pour des applications résidentielles.





**FIGURE 1.** a) Configuration du dispositif d'imagerie pour la mesure du mouvement et b) extraction des régions occupées par la pianiste dans la vue prise à partir du côté gauche

Dans un tel contexte les principaux éléments à prendre en compte sont donc : la capacité d'opérer dans des environnements changeants où l'arrière-scène est de couleur quelconque et contient un ou plusieurs objets de couleur et de forme variées ; la capacité de détecter les membres du pianiste et d'en mesurer le déplacement sans imposer le port de vêtements ou de marqueurs spécifiques ; une mise en place simple et rapide nécessitant un minimum de connaissances techniques ; l'utilisation d'équipements abordables, tel que des caméras couleur standards largement disponibles sur le marché et un ordinateur de puissance moyenne. La nécessité de se limiter à des équipements peu dispendieux élimine aussi tout recours aux capteurs sophistiqués présentés précédemment et impose que l'effort de traitement soit concentré au niveau algorithmique. Une telle flexibilité implique le développement de méthodes de traitement des images et d'estimation de la pose des membres qui soient très robustes à de nombreux paramètres comme le montre la figure 2 où l'éclairage, les vêtements, la complexité de l'arrière-scène et les ombrages ne doivent pas interférer avec la détection des membres du pianiste.



**FIGURE 2. Conditions difficiles dans lesquelles un système typique doit opérer : a) éclairage non uniforme ; b) objets de couleurs complexes ; c) arrière-scène encombrée et ombrages**

Nos expériences préliminaires basées sur les méthodes classiques de traitement des images ont démontré que ces approches ne répondent absolument pas aux exigences complexes de l'application en pédagogie du piano. Les méthodes de base de segmentation des images, c'est-à-dire la mise en évidence du sujet dans une image par rapport à l'arrière-plan, reposent sur l'observation et la reconnaissance de couleurs ou de niveaux d'intensité lumineuse caractérisant la ou les régions d'intérêt dans l'image, dans notre cas la partie supérieure du corps et les bras du pianiste. Une identification guidée manuellement par un opérateur permet d'estimer les paramètres chromatiques (couleurs) et d'intensité d'un ou de plusieurs groupes de pixels (points dans l'image) appartenant au pianiste. L'algorithme de segmentation automatisé se limite alors à rechercher dans le reste de l'image les pixels ayant approximativement les mêmes caractéristiques et les identifient comme appartenant au sujet alors que les pixels dont l'intensité ou la couleur diffèrent sont classifiés comme faisant partie de l'arrière-scène.

Ces méthodes peuvent être raffinées en incluant un mode de poursuite du mouvement lorsqu'elles sont appliquées sur des séquences vidéo. La position future des pixels d'intérêt est alors anticipée à partir de l'hypothèse que les mouvements sont continus dans l'espace. Les régions où devraient se retrouver les membres du pianiste dans les images futures sont calculées à partir de l'historique récent des déplacements. Ceci permet d'accélérer la recherche des pixels d'intérêt sans avoir à balayer la totalité de l'image. Dans tous les cas, les techniques de segmentation recherchées doivent aussi fournir une résolution suffisamment élevée pour permettre la reconstruction en trois dimensions, c'est-à-dire le calcul des régions occupées par le sujet dans l'espace (Cheung, Baker et Kanade, 2003), tout en demeurant stables face à des variations considérables de l'éclairage, des textures et des couleurs, des ombrages et des déplacements.

C'est sur cette base que nous avons entrepris d'analyser, de comparer et de raffiner des techniques de traitement des images à la fine pointe de la technologie afin de surmonter les nombreux défis présentés par l'application de la vision artificielle dans un contexte bien

réaliste qu'offre la pratique du piano. Malgré l'évolution considérable de la science en intelligence artificielle, bien peu de travaux se sont encore penchés sur le développement de systèmes pouvant travailler dans des environnements aussi peu contrôlés et sans imposer de contraintes spécifiques sur le sujet. La prochaine section passe en revue les méthodes modernes de segmentation qui ont été considérées jusqu'à maintenant alors qu'une section subséquente propose une analyse comparative expérimentale de ces méthodes dans le contexte spécifique de la pédagogie du piano.

### **Analyse des technologies de vision par ordinateur de pointe**

Parmi les méthodes les plus récentes proposées dans la littérature, les filtres probabilistes basés sur les caractéristiques chromatiques de la peau et la technique des combinaisons de distributions gaussiennes offrent des approches relativement simples pour éliminer certaines contraintes imposées sur l'uniformité de l'arrière-scène. Malheureusement, elles comportent aussi des limitations importantes pour la poursuite dynamique des membres dans des séquences d'images, ce qui réduit leur potentiel. Face à ces constatations, nous avons initié une nouvelle direction de recherche qui vise à combiner deux autres techniques modernes connues sous les noms de *Continuous Adaptive Mean Shift* et *Receptive Field Histograms*, afin de permettre la détermination de caractéristiques invariantes dans une image qui peuvent ensuite être recherchées dans les images subséquentes d'une séquence vidéo avec plus de stabilité. Nos premières expériences ont démontré que cette technique combinée offre une bonne robustesse aux variations d'éclairage, à la couleur des vêtements, ainsi qu'aux déplacements. Cette solution ouvre la voie à un système d'analyse des images robuste, capable d'opérer à partir d'un ensemble de capteurs visuels passifs par rapport au sujet et réduisant les contraintes à imposer sur l'environnement. Les principes d'opération de ces techniques sont ici résumés.

#### *Filtres probabilistes basés sur les caractéristiques chromatiques de la peau*

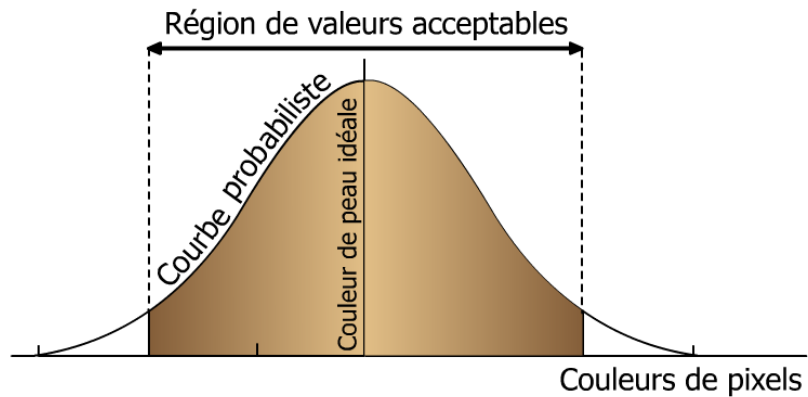
La première approche analysée dans le but de déterminer visuellement la posture d'un pianiste est similaire aux approches chromatiques classiques décrites précédemment mais se spécialise sur l'identification et la détection des couleurs dominantes faisant partie d'un individu. Les couleurs sont considérées comme étant une propriété importante des objets. Il peut donc être avantageux d'en tirer profit dans la mesure où elles demeurent constantes peu importe les transformations physiques effectués par un objet (Bayoumi, Fouad et Shaheen, 2003). En d'autres mots, si un objet est soumis à une rotation, une translation ou à une déformation dans l'espace, ses couleurs demeurent approximativement identiques du point de vue d'une caméra. Plusieurs techniques tirent avantage de ceci pour pouvoir identifier les humains à partir de la couleur de leur peau. Évidemment, une telle approche n'est applicable que dans la mesure où une surface de peau suffisante est exposée au

champ de vision de la caméra, ce qui peut s'avérer intéressant pour la détection des mains et des bras mais s'avère moins pratique pour le reste du corps.

Du et Li (2000) ont exploité la couleur de la peau pour réaliser une interface machine pouvant reconnaître certains gestes humains. Jones et Rehg (1999) ont conçu des filtres probabilistes pour la détection de site Web contenant potentiellement des photographies pornographiques. L'application de ce type de filtres a aussi été introduite dans le contexte de l'indexage d'images et l'identification de la présence d'êtres humains (Bayoumi, Fouad et Shaheen, 2003). Dans toutes ces recherches, les auteurs ont identifié plusieurs difficultés associées avec la segmentation des couleurs. Les ombrages, l'imprécision des capteurs, la direction et la couleur de l'éclairage influencent tous la représentation des couleurs d'un objet (Sigal, Scarloff et Athitsos, 2004). Par conséquent, une couleur est rarement représentée par des valeurs constantes sur toute la durée d'une séquence vidéo. Afin de pouvoir gérer les changements de couleurs, l'environnement doit aussi être considéré.

Les images numériques étant souvent représentées comme une combinaison de valeurs trichromatiques, typiquement le rouge, le vert et le bleu, il est possible de limiter l'influence des variations de l'éclairage sur une image en utilisant une base trichromatique plus appropriée telle qu'un encodage de type YCrCb (Yang, Lu et Waibel, 1998). Dans ce mode de représentation des couleurs, les valeurs Cr et Cb représentent la couleur et Y indique le niveau auquel une couleur est éclairée, c'est-à-dire la luminosité d'un objet. En utilisant une telle représentation, les changements aux valeurs Y ont un impact minimal sur la couleur d'une scène. Les principales sources de variation dans la couleur de la peau se limitent alors aux erreurs engendrées par les capteurs et à la pigmentation naturelle des individus.

Il est possible de représenter la couleur d'un pixel d'une image par sa probabilité d'être localisé dans une région correspondant à de la peau. Le modèle probabiliste gaussien, illustré à la figure 3 est généralement adéquat pour prendre en compte les variations possibles sur les caractéristiques chromatiques de la peau résultant de la prise d'image par une caméra et de la pigmentation. La courbe gaussienne indique la probabilité que la valeur des paramètres d'un pixel corresponde en effet à une couleur de peau. La courbe est centrée sur la couleur considérée idéale, c'est-à-dire la plus représentative de la peau du sujet et s'étend sur une gamme de couleurs dites acceptables.



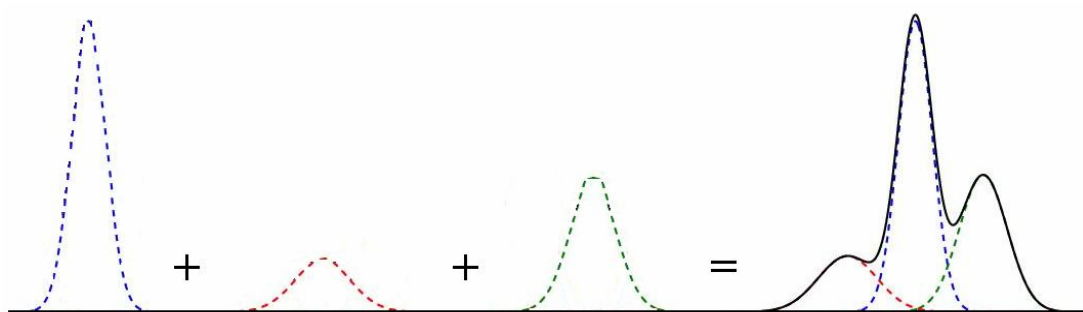
**FIGURE 3. Distribution probabiliste de la couleur de peau humaine**

La principale limitation de cette technique provient du besoin d'initialiser le modèle probabiliste. Une séquence vidéo a le potentiel de contenir une gamme de couleurs de peau mesurées sous différentes conditions environnementales. Afin de pouvoir opérer de manière fiable, le système doit avoir une connaissance préalable des couleurs dans une séquence vidéo, ou doit être assisté par un opérateur qui identifie un échantillon représentatif. Dans une scène de complexité arbitraire, il se peut que plusieurs objets aient une couleur similaire à la peau humaine. Sans un modèle très précis, le système n'arrive pas à bien classer les objets. Ses performances souffrent donc des mêmes limitations que les méthodes de segmentation chromatique classiques lorsqu'elle est appliquée à l'analyse du mouvement chez les pianistes.

#### *Combinaison de distributions gaussiennes*

Une extension possible à la méthode des filtres probabilistes pour la détection de la peau consiste à produire une segmentation plus complète du sujet en introduisant la capacité d'identifier non seulement les sections de peau mais toutes les composantes d'un individu. Une classe importante d'algorithmes de segmentation tente de classer les pixels d'une image comme faisant partie de l'avant ou de l'arrière-scène à l'aide d'un mélange de distributions gaussiennes. Plutôt que de limiter la description de la chromaticité des zones d'intérêt dans les images à une seule distribution probabiliste centrée autour de l'unique couleur recherchée, Stauffer et Grimson (1999) ont proposé de combiner plusieurs distributions gaussiennes qui permettent d'identifier une plus large gamme de couleurs. Ces travaux sont à la base de plusieurs applications reliées à la poursuite des mouvements du corps humain (Wren, Azarbayejani, Darrell et Pentland, 1997) et à l'apprentissage d'activités dans le contexte de la surveillance de la circulation automobile (Stauffer, et Grimson, 2000).

La technique proposée implique que chaque pixel d'une image soit représenté par une ou plusieurs distributions probabilistes gaussiennes. Ceci contraste avec la technique précédente qui attribuait seulement une distribution pour la totalité de l'image. L'initialisation des distributions probabilistes est effectuée à partir du contenu de la première image d'une séquence, c'est-à-dire que la première couleur de chaque pixel définit le centre d'une courbe gaussienne distincte qui s'étend le long d'une région de couleurs arbitraires. Stauffer et al. ont pris en compte le fait que certains pixels de l'arrière-scène peuvent présenter des comportements complexes par rapport à leur couleur. Par exemple, un pixel représentant un feu de circulation peut potentiellement prendre la couleur verte, jaune ou rouge. L'algorithme attribue donc aux pixels une distribution gaussienne pour chaque région de couleur distincte tel qu'illustré à la figure 4.



**FIGURE 4.** Comportement des pixels représenté par un mélange de distributions gaussiennes

Pour chaque image d'une séquence vidéo, le système recherche des distributions gaussiennes qui décrivent le mieux la valeur courante des pixels. Cette recherche s'effectue parmi la liste des distributions associées à chaque pixel. Si la recherche est concluante, la distribution est raffinée afin de mieux intégrer les caractéristiques chromatiques courantes, sinon une distribution supplémentaire est ajoutée à la liste du pixel en question. Le système fait l'hypothèse que les couleurs de l'arrière-scène demeurent plus ou moins constantes. Chaque fois qu'une distribution probabiliste est associée à un pixel lors de l'analyse d'une image, une mesure de correspondance est attribuée à la distribution. En d'autres termes, plus souvent un pixel est représenté par une même distribution, plus il est probable que cette distribution probabiliste représentant une couleur spécifique soit classifiée comme faisant partie de l'arrière-scène. Les distributions associées aux objets en mouvement ont une mesure de correspondance moins élevée puisque les pixels de l'image y sont associés moins fréquemment. Les publications récentes de Horprasert, Hardwood et Davis (2000) ainsi que celle de Atev, Masoud et Papanikolopoulos (2004) proposent de nouveaux mécanismes qui prennent aussi en considération les effets d'ombrage et les changements d'éclairage.

En dépit des gains considérables réalisés par rapport à la recherche de caractéristiques chromatiques uniques, l'utilisation des combinaisons de distributions gaussiennes comporte plusieurs désavantages pour une application en pédagogie du piano. En effet, dans des séquences vidéo montrant des pianistes, les mouvements effectués par le dos et le torse de l'individu peuvent être très subtils. L'amplitude réduite de ces mouvements fait en sorte que ces régions, potentiellement importantes pour l'analyse de la gestuelle, ont tendance à être considérées comme faisant partie de l'arrière-scène. De plus, puisque le système est généralement initialisé avec un individu déjà présent dans la scène, le système n'a pas l'opportunité d'apprendre les caractéristiques de la partie de l'arrière-scène qui est cachée par le pianiste. Lorsque des mouvements plus brusques ou de plus grande ampleur sont produits, des régions d'arrière-scène subitement révélées à la caméra sont souvent faussement classifiées comme faisant partie de l'avant-scène puisque leurs caractéristiques correspondent à de nouvelles distributions gaussiennes non présentes dans le modèle ou peu visitées pendant les phases de recherche précédentes.

#### *Technique du « Continuously Adaptive Mean-Shift »*

De nombreux travaux de recherche rapportés dans la littérature tendent à démontrer que dans l'application des procédés de bas-niveau en vision artificielle des informations de plus haut niveau doivent souvent être employées pour guider le traitement (Comaniciu et Meer, 2002). Dans le cadre du développement de méthodes non invasives de mesure du mouvement pour la pédagogie du piano, ceci suggère que les algorithmes tentant de classifier ce qui appartient à l'avant et à l'arrière-scène en ne considérant que les caractéristiques chromatiques ne soient pas appropriés. Étant donné la complexité de la scène et l'absence souhaitable de contraintes sur l'environnement et le pianiste, il semble plus prometteur de rechercher aussi une description géométrique des objets d'intérêt et de segmenter les images à partir de ces informations.

Tel que démontré précédemment, les images et leurs couleurs peuvent facilement être décrites en fonction de modèles probabilistes. L'algorithme du *Mean-Shift* introduit par Comaniciu et Meer (2002) considère les couleurs comme étant une représentation probabiliste d'une image et permet à une fenêtre (en l'occurrence un groupe de pixels délimitant une région dans l'image) de converger et de s'adapter à la forme d'une distribution quelconque, un membre du pianiste par exemple. Avec l'assistance d'un opérateur humain, l'algorithme peut être initialisé à partir d'une approximation des distributions autour desquelles il doit converger. L'algorithme du *Mean-Shift* a déjà été exploité dans des systèmes de poursuite d'objets dans des images (Comaniciu, Ramesh et Meer, 2002) et de détection de navires (Tek, Comaniciu et Williams, 2001).

Les variations subies par les couleurs peuvent s'opérer de façon subtile ou apparaître graduellement dans une séquence vidéo. L'algorithme du *Mean-Shift* a été adapté par les chercheurs de la compagnie Intel afin de pouvoir répondre à ces variations (Bradski, 1998).

Le nouvel algorithme résultant de leurs travaux se nomme le *Continuously Adaptive Mean-Shift (CAMSHIFT)*. Cette approche débute en définissant une fenêtre sur les régions identifiées par un opérateur. Après avoir complété cette convergence autour de la zone d'intérêt, la fenêtre de chaque région s'oriente et s'agrandit en fonction de la distribution des couleurs. Lorsqu'une nouvelle image est présentée au système, l'algorithme prend en considération les modèles probabilistes représentant les régions de convergence précédentes. En considérant ces nouvelles valeurs, l'algorithme *CAMSHIFT* adapte sa description des modèles pour inclure les derniers changements qui ont été introduits à la scène.

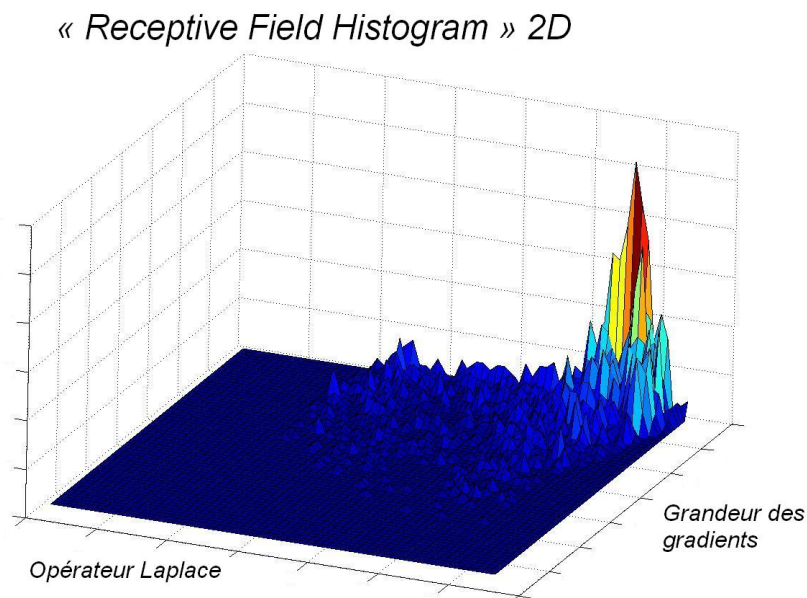
L'avantage de cet algorithme est sa simplicité. Les modèles probabilistes sont représentés sous la forme d'histogrammes qui sont très faciles à manipuler et à traiter. L'algorithme *CAMSHIFT* offre aussi l'avantage de converger vers les sommets des distributions probabilistes (identifiant ainsi les régions les plus fiables) sans nécessiter une connaissance préalable de la forme des objets. En contrepartie, l'algorithme fonctionne mieux pour les distributions unimodales, c'est-à-dire celles qui ne comportent qu'un seul sommet dans leur distribution probabiliste. En permettant à un opérateur d'identifier les régions d'intérêt, le système risque d'introduire des distributions multimodales à partir desquelles la segmentation robuste et fiable est plus difficile à effectuer. De plus, lorsqu'une fenêtre converge sur une distribution, la conformité entre la région définie par la convergence et la distribution probabiliste réelle n'est jamais totale. Il y a donc introduction d'erreurs dans les versions subséquentes du modèle. Ces erreurs influencent les convergences subséquentes et tendent à s'accumuler. Ceci constitue une limitation majeure à la technique puisque qu'il est fortement probable qu'à la fin d'une séquence vidéo les distributions probabilistes ne reflètent plus les objets d'intérêt et conduisent systématiquement à des segmentations erronées.

#### *Technique des « Receptive Field Histograms »*

La technique *CAMSHIFT* se base principalement sur les informations de couleur des régions d'intérêt tel qu'introduites par Swain et Ballard (1990). Dans le but de minimiser les difficultés identifiées avec la méthode *CAMSHIFT*, les histogrammes utilisés par cette technique peuvent être remplacés par des descriptions plus élaborées des régions d'intérêt. Il est possible d'extraire beaucoup plus d'informations d'une image que seulement ses couleurs. Schiele et Crowley (2000) utilisent ce principe dans leurs travaux de recherche pour obtenir une nouvelle méthode de description des objets. Leur technique utilise des histogrammes multidimensionnels nommés *Receptive Field Histograms (RFHs)* qui incluent diverses informations au-delà des caractéristiques chromatiques. Ils représentent une extension des histogrammes de couleurs qui incluent de l'information concernant les contours, les textures et les gradients d'une image.



Lors de leur définition, il est avantageux de formuler ces *RFHs* en utilisant le plus d'informations invariantes possible. La figure 5 illustre un *RFH* à deux dimensions formulé à partir d'un opérateur de Laplace qui permet de faire ressortir le contour des objets, et de la grandeur absolue des gradients d'une image, c'est-à-dire les variations d'intensité locales. Les *RFHs* ont été appliqués avec succès dans un système d'identification de panneaux publicitaires diffusés lors d'événements sportifs (Pelisson, Hall, Riff et Crowley, 2003).



**FIGURE 5.** Exemple bidimensionnel d'un *Receptive Field Histogram*

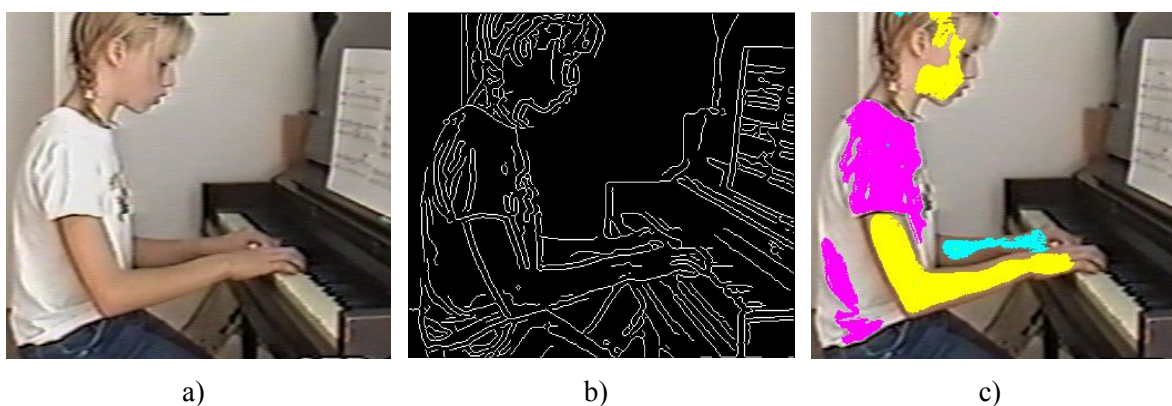
L'utilisation des *RFHs* comporte plusieurs avantages au-delà des histogrammes traditionnels employés dans la méthode *CAMSHIFT*. Avec l'intégration de plusieurs classes d'information, les *RFHs* fournissent des descriptions plus discriminatoires des régions d'intérêt. De plus, les problèmes associés avec les distributions multimodales sont minimisés en raison de l'inclusion d'informations complémentaires. Les *RFHs* peuvent être formulés pour répondre aux besoins du système. Par contre, cette nouvelle technique ne parvient pas à répondre à toutes les difficultés associées à l'analyse de la posture et des gestes d'un pianiste. Certaines composantes des images comportent des informations très complexes et difficiles à représenter par des histogrammes. En augmentant les dimensions de ces derniers, il est possible d'obtenir de meilleurs résultats mais cela engendre également une augmentation considérable de la puissance de calcul requise. Étant donné l'abondance d'information disponible dans les séquences vidéo d'un pianiste en action, l'obtention d'une représentation unique et invariante des régions d'intérêt en ne misant que sur les *RFHs* apparaît irréaliste.

## RÉSULTATS EXPÉRIMENTAUX

L'objectif de la première phase de cette recherche visait à analyser expérimentalement les techniques classiques de mesure du mouvement ainsi que les méthodes de pointe en imagerie par ordinateur, décrites précédemment. Ces méthodes ont été évaluées dans un environnement réel d'enseignement du piano à une clientèle de jeunes élèves à l'Université d'Ottawa. Les contraintes imposées par un tel environnement rendaient la tâche de détection du sujet pianiste fort complexe, mais ont néanmoins permis d'identifier les forces et les faiblesses des différentes stratégies et de mettre en lumière les approches les plus prometteuses pour le développement d'un dispositif de mesure, de quantification et d'évaluation de la gestuelle appliqué à la pratique du piano. Cette section présente les résultats expérimentaux obtenus avec les multiples techniques qui ont été décrites précédemment.

### Performance des techniques de mesure classiques

Les essais menés sur des séquences vidéo montrant un pianiste et utilisant les méthodes de traitement des images traditionnelles, reposant essentiellement sur la détection des variations abruptes d'intensité (détection des arêtes de l'image) et sur la correspondance de la couleur des pixels de l'image avec un modèle fixe et non distribué, ont clairement démontré leurs limitations. La figure 6 illustre l'application de telles techniques de segmentation sur une image d'une jeune pianiste pendant une séance de pratique dans un environnement non contrôlé qu'est sa résidence. Cette scène représente bien le contexte dans lequel les méthodes d'analyse du mouvement doivent être en mesure d'opérer pour avoir un impact réel sur la pédagogie du piano.

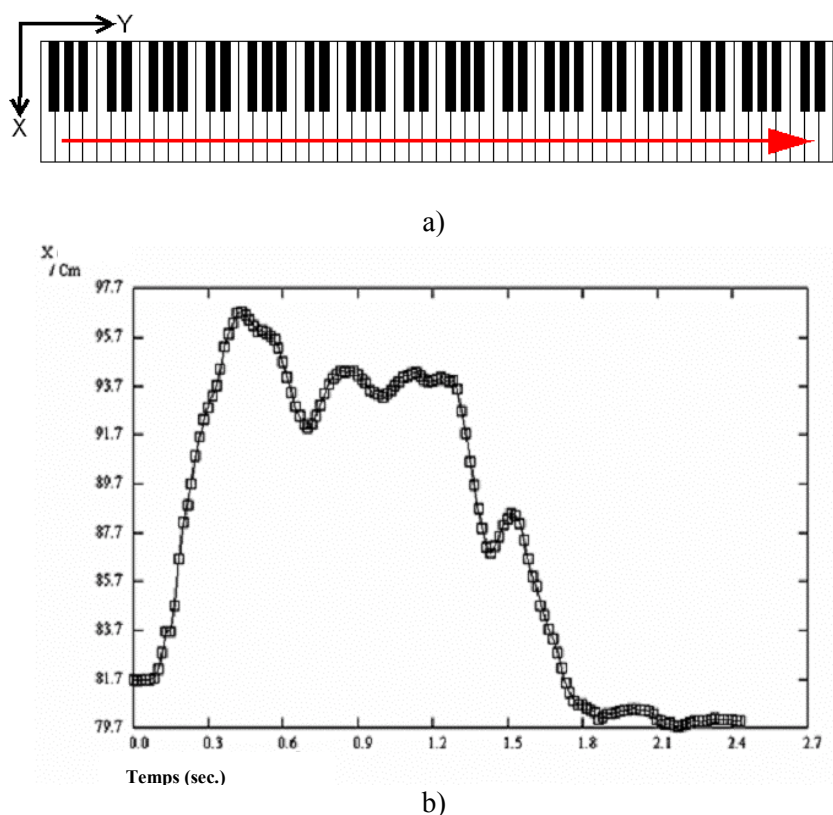


**FIGURE 6.** Évaluation des techniques de segmentation traditionnelles en imagerie dans le contexte de la pédagogie du piano : a) scène de la pianiste ; b) méthodes de segmentation classiques par arêtes ; c) méthodes de segmentation par chromaticité

Lorsque la segmentation traditionnelle par arêtes est employée (figure 6b), le contour de la pianiste est difficilement identifiable en raison du nombre considérable d'objets et de textures différentes dans l'environnement. La zone occupée par la pianiste elle-même étant composée de plusieurs couleurs se retrouve subdivisée en régions alors qu'elle devrait idéalement constituer un tout délimité par une bordure de forme libre.

Dans le cas de la segmentation par chromaticité (figure 6c), les effets d'ombrage et les variations d'intensité et de couleur qui surviennent pendant l'exécution des mouvements ainsi que la similarité entre les couleurs de certaines parties de l'arrière-scène et les vêtements de la pianiste conduisent à une segmentation imparfaite où des régions du mur sont classifiées comme si elles appartenaient à la pianiste. Les bras, la tête et le chandail sont identifiés mais ne couvrent pas la totalité de la surface correspondante, ce qui entraîne une déviation dans l'estimation de la pose des membres de la pianiste. Ces expériences ont démontré la nécessité d'avoir recours à des stratégies d'analyse des images plus sophistiquées qui sont capables d'apprendre et de s'adapter dynamiquement aux caractéristiques de l'environnement, tout en tenant compte des variations présentes dans les régions occupées par le sujet.

L'intérêt à développer des méthodes non invasives de mesure du mouvement pour des applications en pédagogie du piano a aussi été démontré suite à des essais menés avec un capteur de position fonctionnant à partir de champs magnétiques. La version *Fastrak* du dispositif d'évaluation du mouvement de Polhemus (2002) a été évaluée pour suivre les mains d'un pianiste pendant une répétition. Ce type de capteur est composé d'une antenne qui irradie un champ magnétique dans une zone sphérique pouvant s'étendre jusqu'à environ 20 mètres de diamètre. De petits récepteurs sont attachés sur le sujet et captent une partie du signal magnétique. Une analyse de l'amplitude et de la phase du signal permet d'estimer la position et l'orientation du récepteur et donc la pose du sujet dans l'espace. Malheureusement, ces capteurs ont tendance à être très sensibles aux autres champs magnétiques présents dans l'environnement ainsi qu'à la présence de matériaux conducteurs qui peuvent induire de tels champs et brouiller le signal utile émis par l'antenne. Notre évaluation cherchait donc à déterminer si la présence du piano et de ses composantes conductrices (la structure, les cordes, les pièces métalliques du mécanisme) pouvait compromettre la mesure du mouvement. La figure 7 présente une courbe typique du mouvement tel que mesuré par le capteur magnétique alors que la main est déplacée en ligne droite le long du clavier. On note la sensibilité excessive du capteur aux perturbations introduites par l'environnement qui transforme un mouvement rectilinéaire en une courbe relativement complexe avec une déviation de plus de 15 centimètres. En plus de créer un encombrement considérable pour le pianiste qui doit supporter les récepteurs et les câbles qui les relient à l'instrument de mesure, ce type de capteur ne peut fournir d'estimés fiables de la position des membres dans le contexte de la pratique du piano.



**FIGURE 7.** Estimation expérimentale d'un mouvement longitudinal le long du clavier à l'aide du système *Fastrak* de Polhemus : a) mouvement parallèle au clavier ; b) déplacement rapporté par le capteur dans la direction des touches du clavier

### Résultats expérimentaux basés sur l'exploitation des technologies visuelles de pointe

Les techniques d'imagerie par ordinateur décrites préalablement ont été testées sur une séquence vidéo montrant une élève au piano pendant une séance de répétition (figure 8a). La séquence comporte des passages où les mouvements sont effectués à différentes vitesses et avec une amplitude variable. Il importe de noter qu'aucune contrainte n'a été imposée sur l'environnement autour du piano contrairement à ce que requièrent la plupart des techniques d'imagerie actuelles. L'éclairage de la scène est non uniforme, il y a plusieurs ombrages, la couleur de l'éclairage est modifiée par la présence de reflets et plusieurs caractéristiques de la pianiste sont cachées en raison de l'angle à partir duquel la caméra prend les images. Cette séquence a été choisie en raison de sa complexité et parce qu'elle offre une bonne base discriminatoire entre les algorithmes.

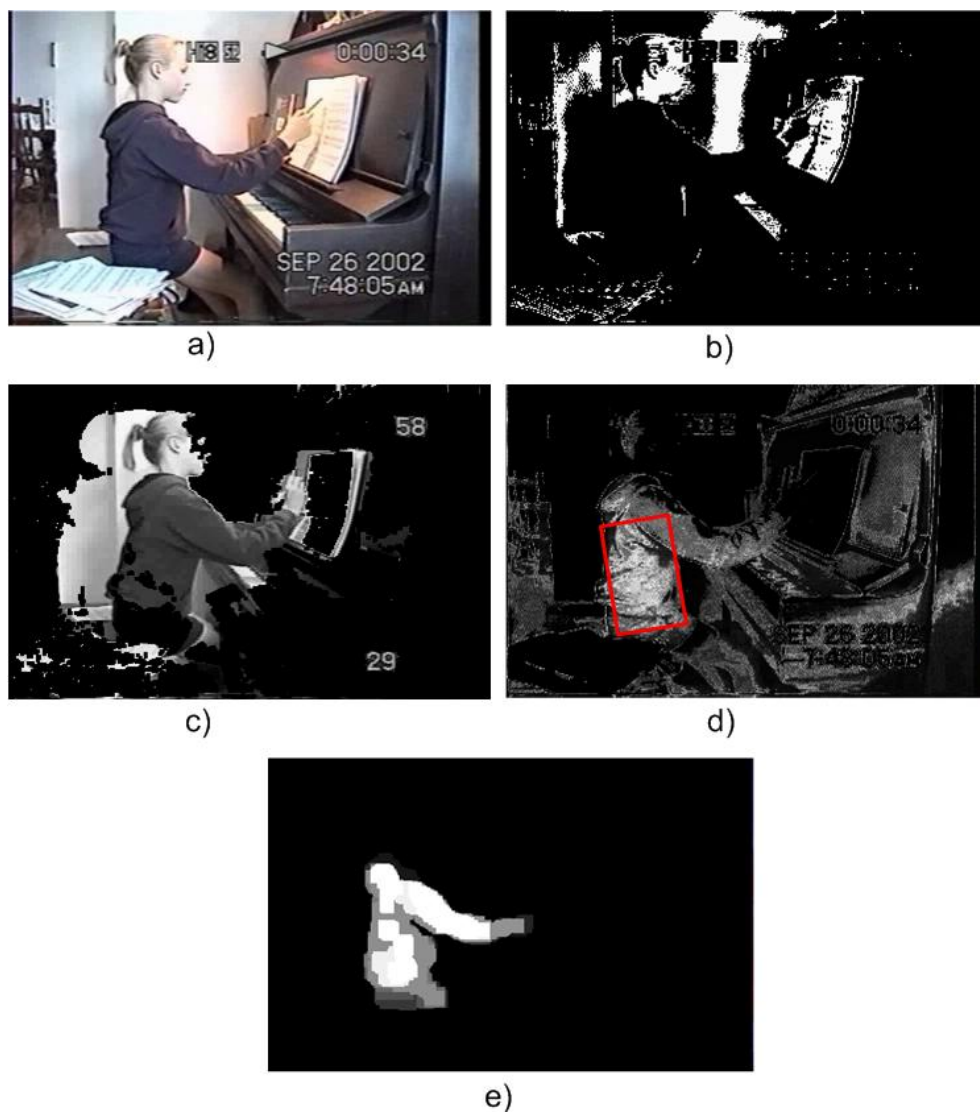
La figure 8b présente les résultats obtenus à l'aide d'un filtre probabiliste utilisant les caractéristiques chromatiques de la peau. On constate entre autres qu'en raison des

changements de couleur dans l'éclairage, la main de la pianiste n'est pas correctement représentée sur une longue période par la distribution gaussienne définie initialement pour la séquence. La segmentation devient rapidement inadéquate après quelques images à peine. Aussi, plusieurs régions de l'arrière-scène sont faussement identifiées comme étant des sections de peau, représentées par les zones blanches, puisque leur couleur s'apparente aux caractéristiques définies par la courbe gaussienne du filtre probabiliste. Une solution possible à cette limitation consisterait à élargir la définition des caractéristiques chromatiques de la peau afin d'y inclure plusieurs distributions de couleurs. Toutefois, cela augmenterait aussi la part de bruit présent dans la segmentation, conduisant à de nombreuses petites régions de l'espace faussement attribuées à un membre de la pianiste.

Dans le cas de la technique basée sur la combinaison de distributions gaussiennes, plusieurs distributions sont associées à chaque pixel de l'image. La classification de l'avant et de l'arrière-scène repose sur une évaluation de la correspondance entre les pixels examinés dans l'image et les distributions représentatives pour cette image. Mais lorsque le système est initialisé à partir d'une image où des composantes de l'avant-scène, ici la pianiste, sont présentes et cachent une partie de l'arrière-scène (le mur, le piano et le banc), la définition des caractéristiques de l'arrière-scène ne peut être complète. Ceci est mis en évidence, à la figure 8c, par la région d'arrière-scène découverte derrière la pianiste suite à son déplacement en direction du piano. Certaines régions le long du dos de la pianiste sont faussement classifiées comme faisant partie de l'avant-scène alors qu'elles appartiennent au mur. La principale cause de ce phénomène provient du manque de mouvements de grande amplitude dans la séquence et de l'apprentissage rapide du système.

La figure 8d représente la distribution probabiliste d'une fenêtre (en rouge) initialisée dans le contexte de l'algorithme *CAMSHIFT*. La fenêtre est concentrée le long du torse de la pianiste. L'utilisation d'histogrammes de couleur simples fait en sorte qu'il y a beaucoup de bruit dans la distribution. Les régions ayant des caractéristiques similaires sont aussi partiellement identifiées (les zones blanches) dans la distribution. Par exemple, le bras droit de la pianiste comporte des couleurs identiques au torse et donc fausse la conformité de la fenêtre. Dans ce cas, l'utilisation de formes géométriques introduit une limitation importante dans le système puisque l'être humain doit idéalement être perçu comme un objet déformable et ne peut que difficilement être représenté adéquatement par une géométrie simple.

Le dernier résultat illustré à la figure 8e a été obtenu suite à une adaptation de l'algorithme *CAMSHIFT* pour inclure des histogrammes multidimensionnels de type *Receptive Field Histograms*. L'algorithme a aussi été modifié afin de réduire sa dépendance à la conformité des formes géométriques. Les résultats obtenus suite à ces modifications aux approches originales sont très encourageants, notamment pour la segmentation du torse et du bras de la pianiste. Cependant, la représentation des régions plus complexes tel que la tête et les mains exige des raffinements supplémentaires.



**FIGURE 8.** a) Image originale ; b) résultat avec le filtre probabiliste de la peau ; c) résultat avec la combinaison de gaussiennes ; d) résultat avec l'algorithme *CAMSHIFT* ; e) résultat avec l'algorithme *CAMSHIFT* modifié incluant les *RFHs*

## CONCLUSION

Cette étude des méthodes de mesure du mouvement humain a permis de retracer les développements historiques de la technologie et de mettre en évidence le besoin prépondérant de définir des techniques de perception non invasives qui demeurent robustes aux variations continues des conditions de l'environnement. La mise au point de systèmes de mesure et d'analyse du mouvement qui pourront assister les enseignants et les

élèves dans la pratique du piano et le développement de méthodes de jeu efficaces et sécuritaires dans un contexte réaliste non soumis à d'importantes contraintes techniques passe par le raffinement des méthodes d'imagerie par ordinateur les plus avancées.

Une étude expérimentale des approches les plus prometteuses a été réalisée à partir de séquences vidéo très représentatives de l'environnement et des conditions dans lesquelles de jeunes musiciens développent leur talent et leur technique. Il en ressort que l'estimation en continu de la pose de la partie supérieure du corps d'un pianiste à partir d'instruments de mesure non invasifs représente un déficit de taille. Cette analyse a permis de conclure sur la nécessité de développer des techniques adaptatives capables d'apprendre en temps réel non seulement les caractéristiques chromatiques mais aussi les paramètres géométriques de plus haut niveau de la scène, afin d'arriver à suivre les variations omniprésentes dans une séquence vidéo enregistrée avec un minimum de contraintes physiques.

Une combinaison originale des méthodes du *Continuously Adaptive Mean-Shift* et des *Receptive Field Histograms* a permis d'obtenir des résultats très encourageants pour la segmentation de séquences vidéo de pianistes en répétition qui surpassent les résultats disponibles à partir des autres méthodes évoluées, en opération à ce jour. La suite de cette recherche permettra de raffiner davantage la nouvelle approche et de l'appliquer au dispositif de vision à points de vue multiples proposé qui fournira alors toute l'information requise pour réaliser la phase de reconstruction tridimensionnelle et d'estimation de la pose complète des membres du pianiste. Cette mesure de la pose des membres du sujet dans l'espace fournira une évaluation quantitative des mouvements d'un pianiste qui pourront ensuite être analysés autant dans une perspective pédagogique que biomécanique et médicale.

### Références

- Albrecht, I., J. Haber et H.-P. Seidel (2003). « Construction and animation of anatomically based human hand models ». *Proceedings of the Eurographics/SIGGRAPH Symposium on Computer Animation*.
- Atev, S., O. Masoud et N. Papanikolopoulos (2004). « Practical mixtures of gaussians with brightness monitoring ». *Proceedings of the 7<sup>th</sup> IEEE International Conference on Intelligent Transportation Systems*, p. 423-428, oct. 3-6.
- Bayoumi, F., M. Fouad et S. Shaheen (2003). « Based skin human detection in natural and complex scenes ». *Proceedings of the 46<sup>th</sup> IEEE International Midwest Symposium on Circuits and Systems*, vol. 2, p. 568-571, déc. 27-30.
- Boulic, R., M. Silaghi et D. Thalmann (2000). « Visualization of local movements for optical marker positioning ». *Lecture Notes in Computer Science 1899* : 133-144.
- Bradski, G. R. (1998). « Computer vision face tracking for use in a perceptual user interface ». *Intel Technology Journal 2* (2).
- Bressler, N. (2000). « Behind closed doors: a qualitative study exploring the pedagogical practices in piano ». Mémoire de maîtrise, McMaster University.
- Cheung, K. M. G., S. Baker et T. Kanade (2003). « Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture ». *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, p. 77-84.
- Comaniciu, D. et P. Meer (mai 2002). « Mean shift: a robust approach toward feature space analysis ». *IEEE Transactions on Pattern Analysis and Machine Intelligence 24* (5) : 603-619.
- Comaniciu, D., V. Ramesh et P. Meer (mai 2002). « Kernel-based object tracking ». *IEEE Transactions on Pattern Analysis and Machine Intelligence 25* (5) : 564-577.
- Comeau, G., P. Payeur, A. Desjardins, E. Keillor et N. Bressler (sept. 2004). « Challenging 300 years of piano teaching practices with 21<sup>st</sup> century technology: piano playing-related health problems ». *Canadian Medical and Biological Engineering Society Conference*, Québec, Canada.



- Drouin, S., P. Hébert et M. Parizeau (2003). « Simultaneous tracking and estimation of a skeletal model for monitoring human motion ». *Proceedings of Vision Interface Conference*, p. 81-88. Halifax, NS.
- Du, W. et H. Li (2000). « Vision based gesture recognition system with single camera ». *Proceedings of the 5<sup>th</sup> International Conference on Signal Processing*, vol. 2, p. 1351-1357, août 21-25.
- Horprasert, T., D. Hardwood et L. S. Davis (janv. 2000). « A robust background subtraction and shadow detection ». *Proceedings of the 4<sup>th</sup> Asian Conference on Computer Vision*, vol. 1, p. 983-988.
- Hsu, Y. P. (1997). « An analysis of contributing factors to Repetitive Strain Injury (RSI) among pianists ». Thèse de doctorat (Ph.D.), Teachers College, Columbia University.
- Jaëll, M. (1897). *Le mécanisme du toucher, l'étude du piano par l'analyse expérimentale de la sensibilité tactile*. Armand Colin et Cie, éditeurs.
- Jobbagy, A. (2004). « Analysis of finger-tapping movement ». *Journal of Neuroscience Methods*.
- Jones, M. J. et J. M. Rehg (1999). « Statistical color models with application to skin detection ». *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, p. 274-280, juin 23-25.
- Lammers, M., M. Kruger, L. J. Stoner, R. Fuller et D. Allyn (1996). « Biomechanics of musical performance: development of expertise in a natural task, playing the trombone ». *Proceedings of the Biomedical Engineering Conference*, p. 56-59.
- Ortmann, O. (1929). *The Physiological Mechanisms of Piano Technique, an Experimental Study of the Nature of Muscular Action Used in Piano Playing, and of the Effects thereof upon the Piano Key and the Piano Tone*. New York : E.P. Dutton & Co, Inc.
- Palmer, C. et C. Drake (1997). « Monitoring and planning capacities in the acquisition of music performance skills ». *Canadian Journal of Experimental Psychology* 51 : 369-384.
- Palmer, C. et P. Q. Pfordresher (2000). « From my hand to your ear: the faces of meter in performance and perception ». *Proceedings of the International Conference on Music Perception and Cognition*, C. Wood, G.B. Luck, R. Brochard, F. Seddon, J.A. Sloboda (Éd.), p. 1-10. Staffordshire, UK.

- Parlitz, D., W. Trappe, D. Dreschner et E. Altenmüller (1999). *Experimental Setup for a Combined Motion, Force and Sound Intensity Analysis of Professional Piano Technique*. Institute for Music Physiology and Performing Arts Medicine, University for Music and Theatre Hannover.
- Pelisson, F., D. Hall, O. Riff et J. L. Crowley (2003). « Brand identification using gaussian derivative histograms ». *Proceedings of the 3<sup>rd</sup> International Conference on Computer Vision Systems*, p. 429-501.
- Polhemus (2002). *Fast and Easy Digital Tracker*. Colchester, VM.
- Rogers, S. M. (1999). « Survey of piano instructors: awareness and intervention of predisposing factors to Piano-Related Injuries ». Thèse de doctorat (Ph.D.), Teachers College, Columbia University.
- Schiele, B. et J. L. Crowley (2000). « Recognition without correspondence using multidimensional receptive field histograms ». *International Journal of Computer Vision* 36 (1) : 31-52.
- Schultz, A. (1949). *The Riddle of the Pianist's Finger and its Relationship to a Touch-Scheme*. New York : Carl Fischer.
- Sigal, L., S. Scarloff et V. Athitsos (juillet 2004). « Skin color-based video segmentation under time-varying illumination ». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (7) : 832-877.
- Stauffer, C. et W. E. L. Grimson (1999). « Adaptive background mixture models for real-time tracking ». *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 246-252.
- Stauffer, C. et W. E. L. Grimson (août 2000). « Learning patterns of activity using real-time tracking ». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8) : 747-757.
- Sullivan, J., M. Ericksson et S. Carlsson (2002). « Recognition, tracking, and reconstruction of human motion ». *Lecture Notes in Computer Science* 2492 : 142-154.
- Swain, M. J. et D. H. Ballard (1990). « Indexing via color histograms ». *Proceedings of the 3<sup>rd</sup> International Conference on Computer Vision*, p. 390-393.
- Tarabella, L., P. Carosi et G. Bertini (1994). « A language and some interfaces for computer music live performances ». *Aarhus Symposium on Generative Grammars for Music Performance*, Johan Sundberg, editors.

- Tek, H., D. Comaniciu et J. P. Williams (2001). « Vessel detection by mean shift based ray propagation ». *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*.
- Trappe, W., D. Parlitz, U. Katzenberger et E. Altenmüller (1998). *3-D Measurements of Cyclic Motion Patterns in Drummers with Different Skills*. Institute for Music Physiology and Performing Arts Medicine, University for Music and Theatre Hannover.
- Vicon Peak (2005). *Vicon Motion Capture System*, Lake Forest, CA.
- Visentin, P. et G .B. Shan (2003). « Biomechanical characteristics of violin performance and their links to Repetitive Strain Injuries ». *Proceedings of the International Arts Conference*.
- Wang, J. J. et S. Singh (2003). « Video analysis of human dynamics – a Survey ». *Real-Time Imaging* 9 : 320-345.
- Wren, C. R., A. Azarbayejani, T. Darrell et A. P. Pentland (juillet 1997). « Pfunder: real-time tracking of the human body ». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (7) : 780-785.
- Yang, J., W. Lu et A. Waibel (1998). « Skin-color modeling and adaptation ». *Proceedings of the 3<sup>rd</sup> Asian Conference on Computer Vision*, vol. 2, p. 687-694.
- Zaza, C. H. (1995). « Musicians' playing-related musculoskeletal disorders: an examination of physical, psychological, and behavioural factors ». Thèse de doctorat (Ph.D.), University of Waterloo.