

## Numéros / n° 1 - automne 2011

# « Vision par ordinateur pour la reconnaissance des gestes musicaux des doigts »

**Sotiris Manitsaris**

Résumé

La vision par ordinateur est la seule technologie permettant la reconnaissance des gestes musicaux des doigts, sans interférence entre le musicien et son instrument, et à un faible coût. Le système PianOrasis propose deux approches pour la reconnaissance : a) la reconnaissance statique, ou reconnaissance des doigtés, et b) la reconnaissance dynamique. La reconnaissance statique s'applique à chaque image de la vidéo, en comparant les caractéristiques de l'image avec le modèle déterministe du geste. La reconnaissance dynamique présuppose la modélisation stochastique du geste et elle s'applique à un ensemble de séquences d'images.

Abstract

Computer vision is the only technology that permits the recognition of the gestures, without interference between the pianist and his instrument, and at a low cost. The system PianOrasis proposes two approaches for the recognition: a) the static recognition, or the fingering retrieval, and b) the dynamic recognition. The static recognition is applied in every video frame. It rests on the analysis and interpretation of the features of the frame, comparing them with the deterministic model of the gesture. The dynamic recognition presupposes the stochastic modelling of the gesture and it is applied in a set of frame sequences.

## 1. Introduction

L'interprétation musicale est le fruit de la symbiose entre le musicien et son instrument de musique. Cette symbiose prend la forme d'une relation interactionnelle et gravitationnelle. Le musicien est à la fois un élément déclencheur et émetteur, reliant la perception, le geste et la connaissance. La recherche présentée dans cet article contribue à l'analyse et la compréhension des mécanismes combinant le geste cinétique et le geste sonore, voire le geste musical, afin de modéliser et de reconnaître des gestes musicaux dans une séquence d'images.

L'analyse et la reconnaissance des gestes musicaux dans une vidéo permettent la compréhension approfondie de l'expressivité musicale. L'art d'interagir musicalement est pour l'artiste un espace d'expression et d'interprétation libre. Le musicien devient une source intarissable de sentiments, exprimés au travers des gestes (Decroux, 1994). En parallèle, les modèles stochastiques peuvent à la fois décrire et interpréter le geste musical et ses éléments structurels, sans tenir compte des sentiments. La vidéo, en tant que séquence d'images, peut contribuer à révéler la structure du geste (Bérard, 2000). L'information de l'expression musicale doit être extraite par la vidéo, et modélisée à l'aide de méthodes stochastiques. Informatique appliquée et mathématiques peuvent ainsi mettre leurs méthodes et techniques au service de l'expressivité musicale.

## 2. État de l'art

Les systèmes de capture optique des mouvements (SCOM), tels que Vicon Peak ou Optitrack, ont déjà été appliqués dans l'analyse de la marche, la rééducation des handicapés ainsi que dans la réalisation d'effets

spéciaux pour le cinéma d'animation en 3D (Vicon Peak, 2005). Palmer a attaché des marqueurs réfléchissants sur le vêtement d'un pianiste afin de capter ses mouvements expressifs par mesure de déplacement des marqueurs (Palmer, 2000).

Dans un autre cas, des chercheurs ont mené une recherche sur la mesure optique de capture des mouvements des violonistes en 3D, en utilisant le système Vicon 460 (Rasamimanana *et al.*, 2009 ; Demoucron *et al.*, 2008). L'objectif de cette recherche a été la modélisation de l'interprétation musicale en obtenant des informations sur les mouvements du violoniste. Ces systèmes sont souvent utilisés pour la mise en oeuvre d'une analyse du geste dans un temps différé par rapport à celui de l'interprétation musicale. Cette méthodologie se base sur des plateformes commerciales spécifiques, présupposant un matériel cher et restrictif pour le musicien.

L'information gestuelle délivrée en temps réel par les capteurs embarqués (Aylward *et al.*, 2006; Coduys *et al.*, 2004), comme dans le cas de la manette Wii (Grunberg, 2008), est sans doute de très haut niveau. Même si cette technologie est souvent utilisée pour la reconnaissance des gestes effectués dans l'espace, il serait pourtant pratiquement impossible qu'elle soit appliquée dans la reconnaissance des gestes des doigts sur une surface ou un objet, puisque les musiciens se sentiraient extrêmement contraints.

Avec ou sans fil, le coût de la technologie des capteurs embarqués a sensiblement diminué ces dernières années, contribuant ainsi au développement de l'analyse du geste dans les arts du spectacle. Boukir et Chenevière ont mené des recherches pour la reconnaissance d'un ensemble des gestes dansés de ballet contemporain, basée sur les trajectoires des mouvements fournies par des SCOM (Boukir *et al.*, 2004). L'Institut de recherche et de coordination acoustique/musique (IRCAM, France) a développé un Réseau de capteurs sans fil (RCSF) pour le suivi continu et la reconnaissance des gestes dansés et musicaux en temps réel (Bevilacqua *et al.*, 2007). Selon cette méthode, deux types d'informations sont mises en évidence de façon continue : (a) la similarité (*vraisemblance*) du geste effectué avec d'autres gestes préenregistrés et (b) la progression temporelle du geste effectué (Bevilacqua *et al.*, 2010).

Les études entreprises autour de l'extraction des doigtés portent sur quatre axes : (a) le prétraitement à l'aide de l'analyse de la partition (Radicioni *et al.*, 2004) ; (b) l'analyse en temps réel basée sur la technologie de MIDI (Verner, 1995) ; (c) le post-traitement du signal sonore (Traube, 2004) et (d) les méthodes de la vision par ordinateur (VpO) (Burns, 2006). Toutes ces approches reconnaissent l'effet du mouvement du doigt en tant qu'événement discret, autrement dit, il s'agit d'une reconnaissance statique des gestes des doigts. Elles ne tiennent pas compte de la nature stochastique du geste et elles ne peuvent pas ainsi être appliquées dans les interprétations vivantes.

### 3. Objectifs

L'objectif général de cette recherche a été la proposition d'une méthodologie de vision par ordinateur et le développement du système PianOrasis (associant les vocables Piano et Orasis ? signifiant « vision » en grec ?) pour la reconnaissance des gestes musicaux des doigts. La méthodologie suit plusieurs objectifs spécifiques. Elle doit être : (a) *Capable* de calculer tous les paramètres définissant les gestes des doigts ; (b) *Vision-orientée* vers l'image du musicien, sans analyse préliminaire ; (c) *Non intervenante*, permettant au musicien de se sentir libre, sans exigence d'équipement spécifique ; (d) *Accessible* et à faible coût, permettant l'utilisation à grande échelle (Manitsaris, 2010).

## 4. Méthodologie et PianOrasis

L'objectif principal de la méthodologie est la reconnaissance des gestes musicaux des doigts effectués sur un instrument de musique classique (piano, clarinette etc.), ou sur un clavier numérique (synthétiseur) ou bien dans l'espace, sans instrument de musique à partir d'une vidéo.

### 4. 1. Détection de la peau

Afin de rendre le système capable de détecter la peau dans une vidéo, un modèle de la peau (MP) a été

développé (figure 4. 1. 1.).

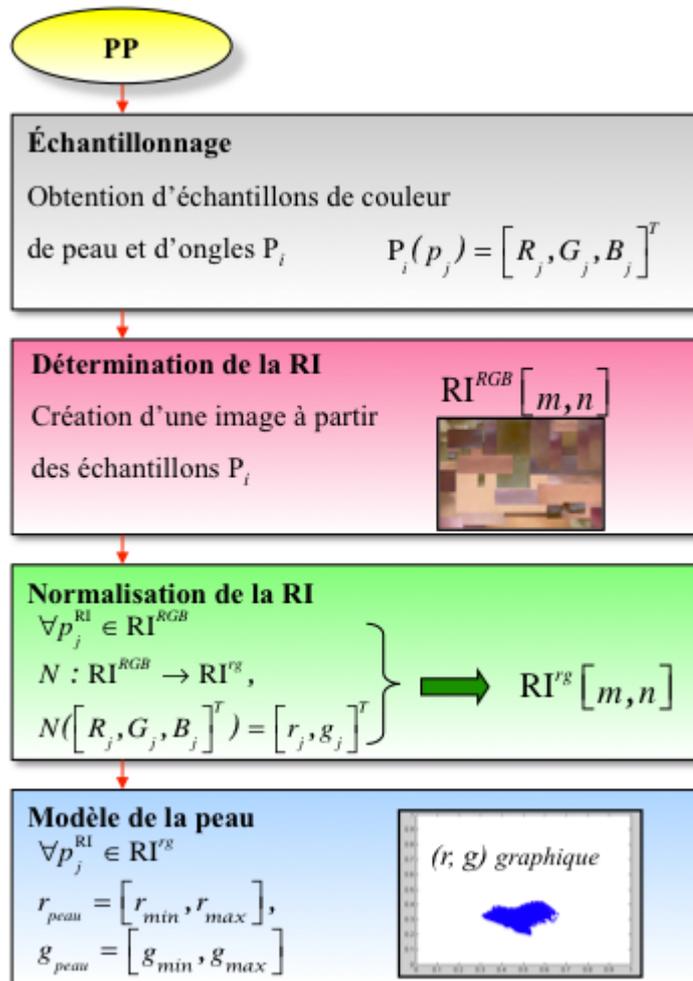


Figure 4. 1. 1. Création du modèle de la peau

Par obtention d'échantillons des pixels de couleur de la peau et d'ongles extraits de la photothèque du pianiste (PP), la région d'intérêt (RI) a été déterminée. La normalisation de la RI, autrement dit la conversion de l'espace RGB vers l'espace normalisé  $rg$ , rend PianOrasis moins dépendant des variations de luminosité et permet d'identifier le MP en tant qu'un ensemble de valeurs. Le résultat exporté est une image binaire contenant soit la valeur 1 pour les pixels de peau (avant-plan), soit la valeur 0 pour le reste des couleurs (arrière-plan).

Par la suite, une séquence d'images binaires a été créée à partir de la vidéo importée, déterminant ainsi les régions contenant de l'information de peau et d'ongles dans l'image. Parfois, bien que le MP n'étant pas parfait, de petites zones de l'arrière-plan sont considérées par le système comme si elles appartenaient à l'avant-plan et vice versa. Ce problème peut être résolu en appliquant des méthodes de morphologie mathématique pour la réduction du bruit.

## 4. 2. Segmentation de la main

La main du pianiste prend une posture semi-étendue durant son interprétation, augmentant ainsi le niveau de difficulté dans la reconnaissance. Vue de face (vue de devant pour la caméra), la zone intérieure de la main étant également une région de peau, dans plusieurs cas la silhouette de la main est extraite en masse avec du bruit. En conséquence, la distinction des doigts dans l'image devient extrêmement difficile, surtout si la distance entre les bouts des doigts est très faible (Figure 4. 2. 1.).

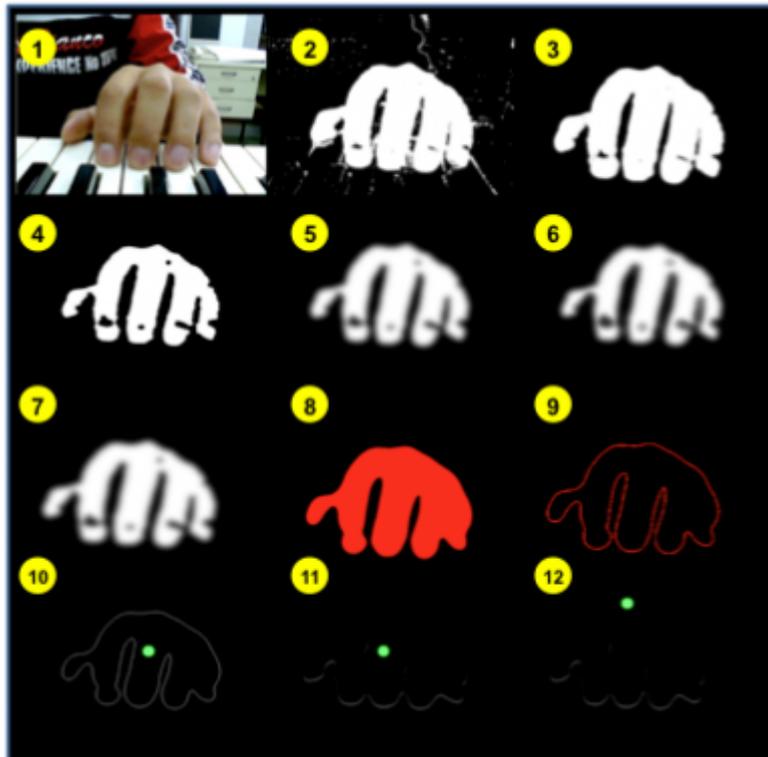


Figure 4. 2. 1. Segmentation de la main et détection des doigts  
(1) image initiale de la vidéo ; (2) application du modèle de la peau ; (3) filtre alternatif séquentiel ; (4) 1<sup>re</sup> dilatation ; (5) 1<sup>er</sup> filtre Gauss ; (6) 2<sup>e</sup> dilatation ; (7) 2<sup>e</sup> filtre Gauss ; (8) seuillage ; (9) extraction du contour ; (10) calcul du centroïde ; (11) localisation des bouts des doigts ; (12) mise à zéro de l'ordonnée du centroïde

Pour cela, l'image binaire est importée dans l'algorithme de la segmentation de la main afin qu'un ensemble de méthodes de traitement d'image lui soit appliqué, comprenant (a) la simplification de l'image binaire par réduction de bruit et extraction de la silhouette de la main et (b) la décomposition de l'image par extraction du contour de la main et des bouts des doigts (Papamarkos, 2000).

### 4. 3. Localisation des doigts

Plusieurs algorithmes de localisation/identification individuelle des doigts dans l'image, utilisant diverses techniques de détection telles que la projection des signatures, la transformée de Hough, les marqueurs colorés, aussi bien que d'autres basés sur des propriétés géométriques, ont été développés.

Le nouvel algorithme, développé dans le cadre de la recherche doctorale et mis en oeuvre dans PianOrasis, suit les critères de détection et de localisation définis par Canny, tout en exploitant les propriétés géométriques de la posture de la main en paume semi-étendue. Les autres techniques ne sont pas forcément satisfaisantes car elles ne localisent qu'indirectement les bouts des doigts, augmentant ainsi la puissance de calcul nécessaire.

La localisation des doigts s'effectue en calculant les distances Euclidiennes entre le centroïde et les coordonnées des pixels appartenant au contour des doigts. Le calcul des maxima locaux des distances Euclidiennes contribue à l'identification des doigts, suivant le pseudo-code ci-dessous :

```
[xc,yc] : =centroïde(ci)
correspondre (xc,yc) sur di[mc,nc]
pour chaque pixel di[mc,nc]
si di[mc,nc]<>0 alors
```

```
distance[s] :=deucl[(xc,0),(xm^2,yn^2)]
position[s,1] :=xm
position[t,2] :=yn
fin si
fin pour
l :=longueur(A)
pour s de 2 à l
si distance[s] = distance[s-1] alors
distance[s] := distance[s]+1/10^4
fin si
fin pour
[MaxDis,MaxPos] :=max(distance,position)
tant que longueur(MaxPos)>5
trouve indice(min(MaxDis))
supprime MaxDis[indice(min(MaxDis))]
supprime MaxPos[indice(min(MasPos))]
fin tant que
pour s de 1 à 5
pour t de 1 à 5
imprime MaxPos[s,t]
fin pour
fin pour
```

Dans le cas d'un « doigt caché », PianOrasis prévoit la position du doigt dans l'image suivante à l'aide des classificateurs, en tenant compte de la « mémoire du geste », calculée en continu par les positions des doigts dans les trois images précédentes.

## 4. 4. Extraction de vecteurs d'observations

À partir du moment où le centroïde est calculé et les bouts des doigts sont identifiés et localisés dans l'image, PianOrasis peut extraire les vecteurs d'observations, en fonction desquels la reconnaissance des gestes sera effectuée.

Les vecteurs d'observation enregistrés par PianOrasis sont : (a) les différences entre l'ordonné de chaque doigt et celle du centroïde ; (b) les abscisses des doigts et (c) les différences entre les abscisses des doigts adjacents. La reconnaissance statique relie uniquement les vecteurs du premier cas, tandis que la reconnaissance dynamique tient compte des trois cas.

## 4. 5. Reconnaissance des doigtés

L'extraction (reconnaissance) des doigtés, autrement dit reconnaissance statique, se met en oeuvre en déterminant le seuil d'appui effectué sur une touche pour chaque doigt. Même dans le cas d'un « doigt caché », le doigté sera extrait sans délai dans les images suivantes.

## 4. 6. Reconnaissance des gestes

La combinaison des doigtés forme un geste dit « pianistique ». Pour cela un dictionnaire des gestes ainsi qu'un alphabet des doigtés ont été créés. Les gestes, se projetant en mouvements musicaux, sont analysés à la fois harmoniquement et mélodiquement afin d'extraire leurs états structurels.

Les valeurs continues des vecteurs d'observation, extraits par les séquences d'images, sont modélisées à l'aide des modèles de mélanges gaussiens (MMG), tandis que chaque geste est modélisé par les modèles de Markov cachés (MMC) (Bakis, 1976 ; Baum, 1972), offrant ainsi une certaine flexibilité à l'entraînement des modèles et permettant l'importation de vidéos de longueurs différentes ou de données manquantes (Alani *et al.*, 1994).

Plus précisément, les MMC continus ont été choisis du fait (a) de la précision fournie dans la classification ; (b) qu'ils ne nécessitent pas de quantification des données ; (c) du petit nombre de données d'entraînement pour les modèles (Rabiner, 1989). Le modèle du geste est évalué en estimant le maximum de vraisemblance (similarité entre le geste effectué et les gestes modélisés).

## 4. 7. Le système PianOrasis

PianOrasis met en oeuvre la méthodologie développée pour la reconnaissance, statique et dynamique, des gestes musicaux des doigts. Le système, ainsi que son interface, ont été entièrement développés sous Matlab. Plusieurs boîtes à outils ont été utilisées, telles que « Image Acquisition » pour la capture des vidéos, « DIPimage » pour le traitement statistique de l'image et le « Kevin Murphy » pour la modélisation stochastique à l'aide des MMC et des MMG.

## 5. Évaluation

### 5. 1. Reconnaissance statique

La première étape d'évaluation se réfère à un extrait de la « Sonate pour piano no 16 en do majeur » de Wolfgang Amadeus Mozart (1756-1791) (Figures 5. 1. 1. et 5. 1. 2.).

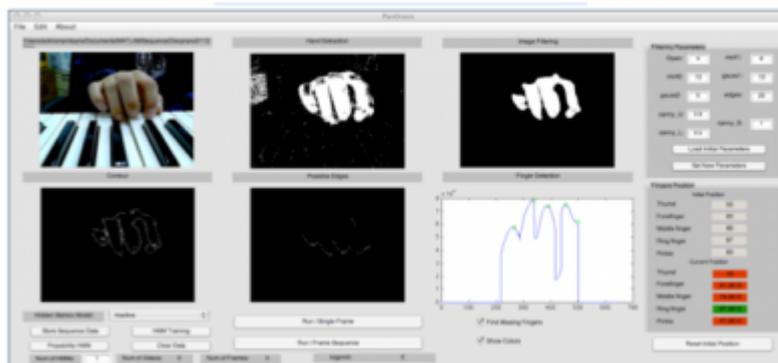


Figure 5. 1. 1. Vue de PianOrasis pour la reconnaissance des gestes statiques : L'annulaire effectue un geste et son indicateur devient vert après avoir analysé et traité l'image



Figure 5. 1. 2. Extrait musical : sonate ?? 16 de Wolfgang Amadeus Mozart

Une vidéo en 19 fps a été prise afin d'évaluer le système PianOrasis en utilisant peu d'échantillons par seconde, à un tempo assez lent et dans de bonnes conditions d'éclairage.

PianOrasis présente un dysfonctionnement lorsqu'un mouvement de la main est provoqué par un saut mélodique, accompagné d'un silence dans la partition. Dans ce cas, de faux doigtés sont extraits, du fait de l'augmentation de la distance entre les bouts des doigts et le centroïde, dépassant ainsi le seuil prédéfini.

Un autre cas intéressant à citer concerne la contribution des classificateurs pour la prévision de la localisation des doigts. Dans les images 287 à 289 le pouce n'est pas détecté (rectangle en vert, figure 5. 1. 3.), probablement parce qu'il est caché derrière un autre doigt. Pourtant, grâce à l'opération de prévision de la localisation des doigts pour la triade précédente des images 284 à 286, le doigté est extrait correctement (rectangle en jaune).

| Image | Localisation par doigt |                  |                  |                  |                  | Localisation totale | Reconnaissance des doigtés |                  |                  |                  |                  | Note | Appui |
|-------|------------------------|------------------|------------------|------------------|------------------|---------------------|----------------------------|------------------|------------------|------------------|------------------|------|-------|
|       | 1 <sup>er</sup>        | 2 <sup>ème</sup> | 3 <sup>ème</sup> | 4 <sup>ème</sup> | 5 <sup>ème</sup> |                     | 1 <sup>er</sup>            | 2 <sup>ème</sup> | 3 <sup>ème</sup> | 4 <sup>ème</sup> | 5 <sup>ème</sup> |      |       |
| 284   | 1                      | 1                | 1                | 1                | 1                | 1                   | 1                          | 0                | 0                | 0                | 0                | SOL  |       |
| 285   | 1                      | 1                | 1                | 1                | 1                | 1                   | 1                          | 0                | 0                | 0                | 0                |      |       |
| 286   | 1                      | 1                | 1                | 1                | 1                | 1                   | 1                          | 0                | 0                | 0                | 0                |      |       |
| 287   | 0                      | 1                | 1                | 1                | 0                | 1                   | 1                          | 0                | 0                | 0                | 0                |      |       |
| 288   | 0                      | 1                | 1                | 1                | 0                | 1                   | 1                          | 0                | 0                | 0                | 0                |      |       |
| 289   | 0                      | 1                | 1                | 1                | 0                | 1                   | 1                          | 0                | 0                | 0                | 0                |      |       |
| 290   | 0                      | 1                | 1                | 1                | 1                | 0                   | 0                          | 1                | 0                | 0                | 0                | LA   |       |
| 291   | 0                      | 1                | 1                | 1                | 1                | 0                   | 0                          | 1                | 0                | 0                | 0                |      |       |

Figure 5. 1. 3. Reconnaissance statique des gestes par PianOrasis pour les images 284 à 291

## 5. 2. Reconnaissance dynamique

Au cours de la seconde expérimentation, le système PianOrasis a été entraîné à reconnaître les gestes pianistiques présentés autour de la gamme de DO : (a) la gamme ascendante (GA) ; (b) la gamme descendante (GD) ; (c) l'arpège ascendant (AA) ; (d) l'arpège descendant (AD) ; (e) les tierces ascendantes (TA) et (f) les tierces descendantes (TD) (Figure 5. 2. 1.).

| gestes         | GA  | GD  | AA   | AD   | TA  | TD  | non classifiés |
|----------------|-----|-----|------|------|-----|-----|----------------|
| GA             | 7   | 0   | 0    | 0    | 2   | 0   | 1              |
| GD             | 0   | 8   | 0    | 0    | 0   | 2   | 0              |
| AA             | 0   | 0   | 10   | 0    | 0   | 0   | 0              |
| AD             | 0   | 0   | 0    | 10   | 0   | 0   | 0              |
| TA             | 2   | 0   | 0    | 0    | 6   | 0   | 2              |
| TD             | 0   | 2   | 0    | 0    | 0   | 5   | 3              |
| reconnaissance | 70% | 80% | 100% | 100% | 60% | 50% | 10%            |

Figure 5. 2. 1.Reconnaissance dynamique des gestes par PianOrasis  
Croisement rang/colonne : nombre de fois où le modèle du geste (colonne) a eu la vraisemblance maximale pour le geste importé (rang) dans PianOrasis ;  
Gestes non classifiés : les vraisemblances ont été inférieures à un seuil pour tous les modèles

PianOrasis a été entraîné sur les 14 différents vecteurs d'observation, à l'aide de 120 vidéos en 19 fps (20 pour chaque geste modélisé). Le nombre des états du modèle du geste a été défini en fonction du nombre total des appuis (doigtés) effectués mélodiquement sur les touches. Par exemple, le modèle de l'arpège ascendant a été modélisé suivant 4 états. L'évaluation de PianOrasis dans ce scénario consiste à reconnaître des « gestes isolés » dans une séquence d'images. Il a été demandé au système de reconnaître chaque geste dans 10 vidéos différentes.

Les taux de reconnaissance pour les gestes TA et TD sont les moins élevés parmi tous les gestes. Ces deux gestes sont assez complexes ayant un niveau stochastique très élevé. Par conséquent, plus le geste est stochastique, plus le nombre nécessaire de vidéos d'entraînement augmente et plus la reconnaissance devient difficile. Dans huit vidéos différentes, les modèles du GA et du GD ont eu la vraisemblance maximale pour les gestes TA et TD et vice versa. Il existe une forte similarité entre les paires de gestes TA/GA et TD/GD, car tous les quatre gestes ont été modélisés suivant le même nombre d'états et les positions de départ de la main/centroïde, ainsi que celles d'arrivée, sont très proches l'une de l'autre pour les deux paires des gestes.

Dans le cas des gestes AA et AD, les taux de reconnaissance sont très élevés étant donné qu'il s'agit de gestes « simples » ayant un niveau stochastique bas. Un autre élément qui prouve la simplicité de ces deux gestes est le petit nombre d'appuis effectués mélodiquement sur les touches, provoquant un nombre d'états assez réduit par rapport aux autres gestes modélisés, ainsi que des vidéos d'entraînement très courtes.

### 5. 3. Reconnaissance des gestes effectués dans l'espace

La troisième étape d'évaluation consiste en une reconnaissance des gestes musicaux sans instrument musical (Figure 5. 3. 1.).

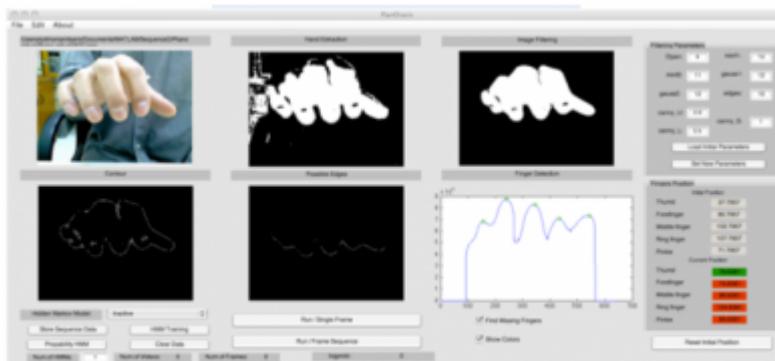


Figure 5. 3. 1. Vue de PianOrasis pour la reconnaissance des gestes effectués dans l'espace : Le pouce effectue un geste et son indicateur devient vert après avoir analysé et traité l'image

Il a été demandé au pianiste d'interpréter l'arpège ascendant de la gamme de DO en bougeant ses doigts dans l'espace. Dans les 130 images qui ont été évaluées, la localisation des doigts, par détection et classificateurs, a été réussie à 100 %, tandis que les doigtés ont été extraits correctement à 97 %. Il est facile de constater que même dans le cas où la couleur de certains objets de l'arrière plan est très proche de celle de la peau, comme par exemple dans la première image de la figure 5. 3. 1., l'application des techniques de la morphologie mathématique sur la sortie du MP (seconde image) est très efficace car cette région est éliminée et dans l'image suivante la silhouette de la main est clairement extraite. Dans tous les cas, puisque la caméra est positionnée devant le musicien, le nombre de pixels de l'arrière plan ayant une couleur proche de celle de la peau est réduit car le corps du musicien occupe la plus grande partie de l'image.

## 6. Contribution

La contribution de la méthodologie présentée dans la musique repose sur deux axes principaux : (a) la pédagogie musicale et (b) la composition de musique contemporaine. PianOrasis peut contribuer à la pédagogie musicale en tant que support informatisé pour l'apprentissage du piano. La technique des doigts est pour un pianiste l'alphabet de son interprétation musicale. Dans ce cadre, PianOrasis peut être utilisé en tant que système d'optimisation de la technique des doigts et du choix des doigtés au piano.

Les résultats obtenus, lors de l'expérimentation de la reconnaissance des gestes musicaux sans instrument musical, indiquent que PianOrasis peut être utilisé, avec peu de modifications, comme un nouvel instrument numérique pour la composition de musique contemporaine. Cela peut être réalisé en mettant en relation deux domaines distincts, que sont le geste et le son ou bien la parole par synthèse concaténative (Aperghis *et al.*, 2011). PianOrasis pourrait servir à une large gamme de scénarii de composition de musique contemporaine, en tant qu'une interface tangible pour la reconnaissance des gestes musicaux effectués sur un objet ou une surface dans un environnement réel.

Jusqu'à ce jour, les gestes des artisans, constituant du patrimoine culturel immatériel, ne se sont jamais laissés « mettre en boîte », enregistrer, classifier, codifier de manière à pouvoir être transmis, même après leur extinction, par quelque moyen que ce soit. Par conséquence, la méthodologie proposée, ainsi que sa mise en oeuvre pour la reconnaissance et la modélisation des interactions gestuelles entre les artisans et leur matière, consisterait une innovation dans le domaine de la sauvegarde des savoir-faire rares.

La similarité entre les gestes effectués sur un clavier de piano et ceux effectués sur un clavier d'ordinateur est forte sans aucun doute. Une telle méthodologie de VpO peut favoriser le remplacement des périphériques restrictifs d'un ordinateur, tels que le clavier ou la souris, par une caméra embarquée à l'écran de l'ordinateur et un algorithme de reconnaissance des mouvements articulés des doigts, en paume semi-étendu pour chaque doigt.

## 7. Conclusion

Motivés par l'absence de liens opérationnels entre les domaines de la vision par ordinateur et de l'interaction musicale, nous avons étudié dans ce travail de recherche la conception et la mise en oeuvre d'une méthodologie et d'un système de vision par ordinateur au service de l'expression musicale libre, en obtenant ainsi l'objectif de la reconnaissance des gestes musicaux des doigts sans interférence entre le musicien et l'ordinateur. Le piano a été un excellent cas d'étude pour l'analyse et la modélisation des gestes musicaux des doigts. Cependant, ce type de modélisation ouvre divers perspectives d'application de la méthodologie, allant au delà d'interfaces ou de claviers spécifiques.

PianOrasis a certaines restrictions techniques pour la reconnaissance des gestes effectués sur le clavier d'un piano. La reconnaissance ne s'effectue que pour les touches blanches et le plan de la capture se limite à deux octaves du piano. Les données dans la séquence d'images concernent uniquement la main droite du pianiste. Afin de résoudre les restrictions concernant les touches et le plan de capture, des techniques d'invariance de rotation et d'échelle seront intégrées à la version du système pour la reconnaissance des gestes des doigts en temps réel. Le système sera ainsi capable de détecter la RI de manière beaucoup plus efficace en analysant des images qui contiennent uniquement la silhouette de la main. Cela peut aussi contribuer à la reconnaissance en temps réel, laquelle sera mise en oeuvre en utilisant l'environnement de programmation Max/MSP et ses patches MuBu et FTM.

Néanmoins, les restrictions actuelles ne constituent pas de contraintes pour le contrôle gestuel du son ou bien de la parole par synthèse concaténative à partir des gestes des doigts effectués dans l'espace. Le développement d'un système de VpO pour la reconnaissance des gestes musicaux des doigts dans un environnement réel (espace, surface ou objet), ayant comme objectif le contrôle gestuel du son et la composition de la musique contemporaine, est envisagé.

---

**Pour citer ce document:**

Sotiris Manitsaris, « Vision par ordinateur pour la reconnaissance des gestes musicaux des doigts », *RFIM* [En ligne], Numéros, n° 1 - automne 2011, Mis à jour le 23/09/2011

URL: <http://revues.mshparisnord.org/rfim/index.php?id=107>

Cet article est mis à disposition sous [contrat Creative Commons](#)