HAL
open science

VISION PAR ORDINATEUR ET APPRENTISSAGE STATISTIQUE : VERS UN INSTRUMENT DE MUSIQUE IMMATERIEL

Sotiris Manitsaris, Tsagaris Apostolos, Vassilios Matsoukas, Athanasios Manitsaris

To cite this version:

Sotiris Manitsaris, Tsagaris Apostolos, Vassilios Matsoukas, Athanasios Manitsaris. VISION PAR ORDINATEUR ET APPRENTISSAGE STATISTIQUE : VERS UN INSTRUMENT DE MUSIQUE IMMATERIEL. Journées d'Informatique Musicale, 2012, Mons, France. hal-03041751

HAL Id: hal-03041751

https://hal.science/hal-03041751v1

Submitted on 5 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

VISION PAR ORDINATEUR ET APPRENTISSAGE STATISTIQUE : VERS UN INSTRUMENT DE MUSIQUE IMMATERIEL

[1]*Sotiris Manitsaris* [2]*Tsagaris Apostolos* [2]*Vassilios Matsoukas* [2]*Athanasios Manitsaris*
[1]Centre de Robotique (CAOR)
Ecole Nationale Supérieure des Mines de Paris (Mines ParisTech)
France
[2]Laboratoire des Technologies Multimédias et de l'Infographie
Département d'Informatique Appliquée
Université de Macédoine
Grèce
[1]sotiris.manitsaris@mines-paristech.fr [2]{tsagaris, vmats, manits}@uom.gr

RÉSUMÉ

Le présent article propose une méthodologie de vision par ordinateur pour la reconnaissance simultanée des gestes musicaux et complexes des doigts qui sont effectués dans l'espace sans instrument de musique matériel. Les techniques d'analyse d'images sont appliquées pour segmenter la main ainsi que pour détecter et identifier les doigts dans une vidéo. La reconnaissance et la prédiction des gestes des doigts se basent sur la modélisation stochastique des caractéristiques de haut niveau à l'aide des Modèles de Markov Cachés (MMC) et des Modèles de Mélanges Gaussiens (MMG). Des techniques d'invariance d'échelle et de rotation ont été appliquées. Cet article propose également une méthodologie de mise en correspondance des gestes des doigts avec des paramètres sonores pour le contrôle gestuel du son. Les applications concernent directement la composition de la musique contemporaine et en général les arts du spectacle.

Keywords : reconnaissance des gestes, modélisation stochastique, vision par ordinateur, contrôle gestuel du son

1. INTRODUCTION

Il y a quelques années le synthétiseur électronique présentait un concept révolutionnaire d'un nouvel instrument musical. Le synthétiseur produit des sons par génération de signaux électriques de différentes fréquences. Cet instrument musical a donné des nouvelles perspectives aux musiciens qui avaient déjà des connaissances au jeu pianistique. Pourtant, le clavier d'un piano, ou même celui d'un synthétiseur électronique, est constitué d'un mécanisme intermédiaire entre le musicien et la musique. Aujourd'hui, la nécessité de contrôler la musique en effectuant des gestes naturels dans un environnement réel devient de plus en plus importante. L'ordinateur peut désormais être considéré comme un instrument musical digital. Une fois que le geste est reconnu, plusieurs paramètres sonores issus de l'instrument peuvent être contrôlés de manière dynamique.

L'objectif de la recherche présentée ici consiste à proposer une méthodologie de vision par ordinateur pour la reconnaissance des gestes musicaux des doigts en temps réel sans instrument de musique matériel (*système PianOrasis:max*). La recherche actuelle est basée sur la méthodologie existante pour des gestes effectués sur un clavier de piano en temps différé (*système PianOrasis*). La méthode proposée permet la reconnaissance et la prédiction des gestes musicaux des doigts en utilisant des techniques d'apprentissage. La méthodologie est : (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. Deux différentes mises en œuvre de cette méthodologie sont faites. PianOrasis a été développé sur Matlab tandis que PianOrasis:max sur Max/MSP.

2. ETAT DE L'ART

Plusieurs recherches ont été menées pour la reconnaissance des gestes appliquée à l'interaction musicale en utilisant des techniques différentes. Ces techniques sont classées en trois catégories : a) l'approche basée sur la vision avec marqueurs, b) les interfaces tangibles et les capteurs embarqués et c) l'approche basée sur la vision sans marqueurs.

2.1. Approche basée sur la vision avec marqueurs

Les systèmes basés sur la vision à l'aide des marqueurs, tels que ViconPeak ou Optitrack, ont déjà été utilisé pour l'analyse de la marche, la rééducation de personnes handicapées ainsi que la réalisation d'effets spéciaux pour le cinéma d'animation en 3D [2]. Palmer (2000) a attaché des marqueurs réflectifs sur le vêtement

d'un pianiste afin de capter ses mouvements expressifs par mesure de déplacement des marqueurs [3]. Dans un autre cas, une recherche a été menée sur la mesure optique de capture des mouvements des violonistes en 3D, en utilisant le système Vicon 460 [4]. 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 œuvre d'une analyse du geste dans un temps différé par rapport à celui de l'interprétation musicale.

2.2. Interfaces tangibles et capteurs embarqués

L'information gestuelle délivrée en temps réel par les interfaces tangibles ou les capteurs embarqués [5], comme dans le cas de la manette Wii, est 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'elles soient appliquées dans la reconnaissance des gestes des doigts sur une surface ou un objet, puisque les musiciens se sentiraient extrêmement contraints.

L'IRCAM a développé un Réseau de Capteurs Sans Fil (RCSF) pour le suivi continu et la reconnaissance des gestes dansés et musicaux en temps réel [6]. Dans l'application du «violon augmenté», l'architecture tangible comprenait des capteurs d'accélération, des gyroscopes et un capteur de pression monté sur l'archet du violon, tandis qu'un bracelet autour du poignet du violoniste intégrait l'alimentation et l'émetteur sans fil ZigBee. 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.

Des objets simples, tels que le ballon ou les échecs, ont été utilisés pour la création des interfaces sonores en utilisant un module central sans fil, qui comprend accéléromètre et gyroscope. C'est une approche à bas coût de type « do it yourself » plutôt, qu'une « prête à l'emploi » et l'utilisateur peut effectuer ses propres gestes pendant une partie d'échecs en y associant des sons de son choix [7].

2.3. Approche basée sur la vision sans marqueurs

Dans le cadre du projet iMuse (integrated Multimodal Score-following Environment), des techniques de reconnaissance des gestes par la vision sans marqueurs ont été utilisées. Le système reconnait les gestes d'un pianiste en temps réel et les synchronise avec la partition de la pièce. Une camera a été installée sur le clavier du piano, les gestes sont analysés durant l'interprétation et un graphique représentant les mouvements des doigts forme l'alignement avec la partition. Pendant les interprétations vivantes, les gestes sont suivis et sont comparés aux graphiques des gestes déjà enregistrés. La main du musicien est analysée d'une manière globale. Ceci dit, le système reconnaît plutôt les postures globales de la main que les gestes des doigts du pianiste.

Une autre approche est celle de type bio-musical, où il y a une source d'information, de type meta-musique, qui est placée sur le processus interne et externe et résulte en la création d'un contenu musical dynamique. Ce concept est transféré à l'expression musicale du cerveau et du corps, en utilisant le système EPOC EMOTIV (*électroencéphalogrammes*) pour l'acquisition des signaux du cerveau qui pourrait être utilisés pour la détection des émotions de l'artiste ainsi que la caméra Microsoft KINECT pour la reconnaissance des postures du corps de l'artiste [8].

3. METHODOLOGIE

La méthodologie se base sur celle proposée par Manitsaris (2010) sur la reconnaissance des gestes musicaux sur un clavier de piano [1] et vise à reconnaître les gestes musicaux des doigts qui sont effectués dans l'espace. Afin d'obtenir cet objectif, les bouts des doigts doivent être identifiés et localisés à l'aide d'une caméra bas coût. L'obtention des échantillons des pixels de l'espace RGB contenant de l'information de la peau et des ongles permet la création d'un modèle de peau (MP) la détermination de la Région d'Intérêt (RI) (Figure 1). L'application du MP permet la détection des régions de la peau sur n'importe quelle séquence d'images optiques.

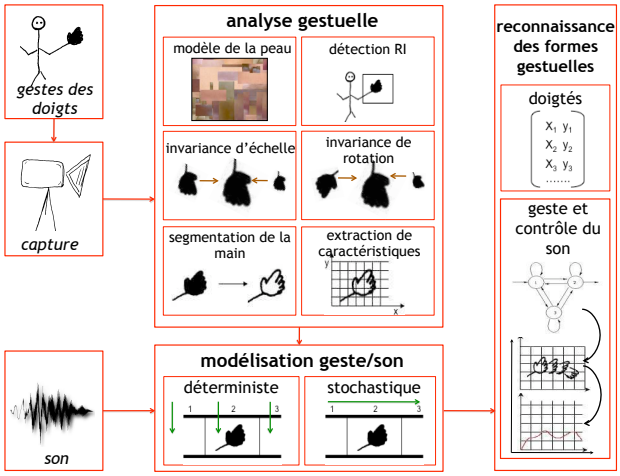

Figure 1. Méthodologie

L'étape suivante consiste à segmenter la main du musicien à l'aide des techniques de traitement d'images. Un ensemble de filtres sont appliqués à la séquence d'images afin de distinguer la zone de la silhouette de la main (*premier plan*) et le bruit (*arrière plan*).

Cependant, la problématique la plus intéressante reste l'identification des doigts. Plusieurs algorithmes pour l'identification individuelle des doigts sur l'image ont déjà été proposés. L'algorithme propose la détection des bouts des doigts en tenant compte des particularités de la posture de la main lorsque la paume est semi-étendue. Plus spécifiquement, la détection des bouts des doigts est effectuée en calculant les distances euclidiennes

entre le centroïde de la silhouette de la main et les coordonnées des pixels de son contour. Pour augmenter la performance de l'identification des doigts, des techniques d'invariance d'échelle et de rotation sont appliquées à chaque image de la séquence.

L'étape finale consiste à reconnaître les gestes. Un alphabet et un dictionnaire des gestes des doigts ont été créés. Les gestes étant des mouvements musicaux (*gamme ascendante/descendante, arpège, etc*) ont été analysés harmoniquement et mélodiquement afin d'extraire leurs états structuraux. Quatorze caractéristiques décrivant les gestes des cinq doigts sont extraits de chaque image. Ces caractéristiques des valeurs continues, autrement dites des vecteurs d'observation, ont été modélisées à l'aide des MMG, tandis qu'un modèle Markov différent est attribué à chaque geste musical [9, 10].

4. VISION PAR ORDINATEUR

La caméra est positionnée devant la main du musicien afin que les gestes des doigts puissent être enregistrés sans « doigt caché » en position de repos. Une vidéo de résolution maximale de 320x240 pixels est enregistrée.

4.1. Analyse et segmentation de l'image

Afin de rendre le système capable de détecter la peau dans une vidéo, un modèle de la peau (MP) a été développé. 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, 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 traitement d'images (*morphologie mathématique et filtrage*) pour la réduction du bruit. Plus précisément, ces méthodes comprennent (a) la simplification de l'image binaire par réduction de bruit, extraction de la silhouette de la main et découpage de l'image binaire aux limites de la main et (b) la décomposition de l'image par extraction du contour de la main et des bouts des doigts [11].

Plus précisément, juste après le découpage de la RI aux limites de la silhouette de la main, des filtres Open, Min et Gauss s'appliquent de manière répétitive. Les paramètres des filtres déterminent le taux de leur implication à l'image. Suite à une recherche antérieure [12], le choix de ces paramètres se fait automatiquement en fonction du quotient du nombre des pixels de la peau sur le nombre des pixels de l'image.

On en déduit ainsi que le système est relativement indépendant des variations de la distance entre la main et la caméra (*invariance d'échelle*). En outre, le découpage de la RI aux limites de la silhouette de la main permet une analyse de l'image bien plus rapide [13].

4.2. Exportation des caractéristiques

La main du musicien 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.

La localisation des doigts s'effectue en calculant les distances euclidiennes entre le centroïde et les coordonnées des pixels appartenant au contour des doigts. 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.

À partir du moment où le centroïde est calculé et les bouts des doigts sont identifiés et localisés dans l'image, le système peut extraire les vecteurs d'observations, en fonction desquels la reconnaissance des gestes sera effectuée. Les vecteurs d'observation enregistrés 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 se base uniquement sur les vecteurs du premier cas, tandis que la reconnaissance dynamique tient compte des trois cas.

La main du musicien prend souvent une posture inclinée, rendant la reconnaissance plus difficile. Suite à une recherche antérieure, des techniques d'invariance de rotation peuvent être appliquées à l'image. Etant donné que les doigtés sont repérés et identifiés, l'angle de la ligne qui résulte du centroïde et les coordonnées du majeur et de l'axe horizontal est calculé. Si cet angle dépasse un certain seuil, l'image pivote automatiquement à l'orientation inverse de l'inclinaison [13]. Cela permet une reconnaissance robuste même si la main est inclinée et assure l'indépendance des caractéristiques gestuelles, élément très important pour la partie entraînement des MMC. Pour des valeurs d'angle d'inclinaison inférieures au seuil, le système est capable de reconnaître le geste sans correction de la rotation.

5. MODELISATION, APPRENTISSAGE ET MAPPING

5.1. Modélisation stochastique et reconnaissance

L'extraction (*reconnaissance*) des doigtés, autrement dit reconnaissance statique, se met en œuvre 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 dans les images suivantes.

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.

Pour l'instant, deux versions du système existent : (a) PianOrasis développé sur Matlab étant capable de reconnaître les gestes à la fois statiques et dynamiques des doigts en temps différé, et (b) PianOrasis:max, développé sur Max/MSP, étant capable de reconnaître les gestes statiques des doigts en temps réel.

Les valeurs continues des vecteurs d'observation, extraits par les séquences d'images, sont modélisées à l'aide des MMG, tandis que chaque geste est modélisé par les MMC [9, 10], offrant ainsi une certaine flexibilité à l'entraînement des modèles et permettant l'importation de vidéos de longueurs différentes ou de données manquantes.

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'apprentissage pour les modèles. Le modèle du geste est évalué en estimant le maximum de vraisemblance (*similarité entre le geste effectué et les gestes modélisés*).

5.2. Correspondance de la musique au geste

Afin de faire correspondre le geste à la musique, on doit avant tout répondre à des questions comme « à quoi peut-on associer tel ou tel paramètre gestuel ou sonore ? » et « comment le faire ? ».

Dans le cadre de la méthodologie proposée, il s'agit de faire correspondre les caractéristiques des gestes des doigts de haut niveau aux paramètres acoustiques de bas niveau. Les caractéristiques extraites décrivant le geste du musicien sont automatiquement enregistrées dans un tableau. Dans l'hypothèse où les données gestuelles sont valables, les listes des données importées sont interprétées en tant que valeurs de contrôle. L'importation d'une nouvelle liste des données gestuelles alors correspond à un nouvel ensemble de valeurs dans l'espace sonore.

La méthodologie présentée ici repose sur une correspondance au niveau du temps (*temporal mapping*) au lieu d'une correspondance au niveau de l'espace (*spatial mapping*). Cela est lié au besoin de développer un système de mapping en temps réel. Ceci dit, une procédure temporelle de correspondance entre le geste et le profil temporel du son doit être déterminée [14]. La correspondance temporelle peut être considérée en tant qu'une procédure de synchronisation entre les paramètres gestuels d'entrée et les paramètres sonores de sortie. Le rythme d'exécution de geste peut être synchronisé à l'aide des procédures temporelles spécifiques. La correspondance temporelle peut ainsi être prolongée au contrôle d'une combinaison des profils temporels continus et d'événements discrets dans l'espace, en les synchronisant avec les caractéristiques du geste d'entrée.

Le principe de la méthodologie repose sur le fait que les gestes sont des procédures temporelles qui se caractérisent par des profils temporels. La notion de la correspondance temporelle est approchée en examinant les côtés temporels de la relation entre geste, son et structures musicales. C'est l'évolution temporelle des données et non pas leurs valeurs absolues qui devient un élément fondamental des systèmes musicaux d'interaction. Cette approche se base sur la modélisation des caractéristiques temporelles très spécifiques, lesquelles sont soit acquises lors de la procédure de l'apprentissage statistique, soit réglées manuellement.

6. SYSTEMES ET INTERFACES

La méthodologie a été mise en œuvre à l'aide de deux environnements de programmation différents. La première version du système a été faite sur Matlab (Figure 2) et la deuxième sur max/MSP (Figure 3).

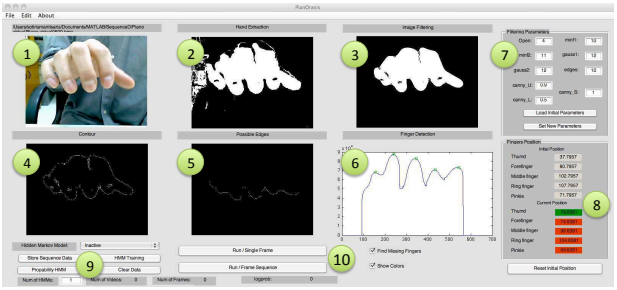

Figure 2. Interface du PianOrasis
(1) image initiale en RGB ; (2) application du modèle de la peau ; (3) application de la morphologie mathématique ; (4) extraction du contour de la main ; (5) arêtes probables correspondant aux bouts des doigts ; (6) localisation des bouts des doigts dans l'image et prédiction de la position des doigts cachés ; (7) paramétrage manuel des filtres et de la détection des contours ; (8) initialisation de la position de repos pour chaque doigt et détection des doigtés ; (9) entraînement des modèles gestuels à partir des vidéos ; (10) option de reconnaissance statique ou dynamique.

PianOrasis propose des fonctionnalités à la fois pour la reconnaissance statique (*reconnaissance des doigtés*) et la reconnaissance dynamique des gestes musicaux des doigts en temps différé (Figure 2). Le système, ainsi que son interface, ont été entièrement développés sous Matlab. Plusieurs toolbox 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 MGM. PianOrasis assure différentes fonctions concernant l'importation et le suivi de traitement de la vidéo, le filtrage, l'entraînement et la reconnaissance.

PianOrasis:max est entièrement développé sous Max/MSP et propose actuellement des fonctionnalités de reconnaissance statique en temps réel (Figure 3). Max/MSP propose tout un ensemble de routines qui existent sous la forme des toolbox. La bibliothèque « Computer Vision for Jitter » de Jean-Marc Pelletier a été utilisé pour la segmentation de l'image mais en même temps d'autres routines appelées « objets extérieurs » ont été créées.

La mise en œuvre du module de mapping geste/son sera faite en utilisant le patch MnM. Il s'agit d'un toolbox de Max/MSP qui se base sur le patch FTM. Il est composé d'un groupe des unités qui procurent des algorithmes de l'algèbre linéaire de base, algorithmes de correspondance et des algorithmes de modélisation statique comme le PCA, GMMs et HMMs.

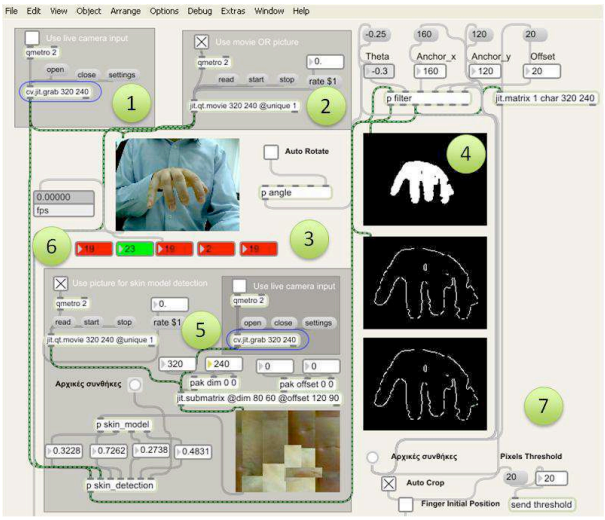

Figure 3. Interface du PianOrasis:max lorsque un doigté de l'annulaire est détecté

(1) Option de reconnaissance en temps réel ; (2) Option de reconnaissance en temps différé ; (3) vidéo initiale en RGB ; (4) application du modèle de la peau et de la morphologie mathématique ; (5) Option de paramétrage du modèle de la peau ; (6) Reconnaissance des doigtés ; (7) Option automatique du découpage de la RI et de la rotation de l'image

7. EVALUATION

Un premier scénario d'évaluation pour PianOrasis:max a été basé sur un geste d'arpège complet pour la gamme de Do ; un geste assez simple car les distances entre les doigts sont grandes et aucun doigt ne se cache. Nous avons demandé à un musicien d'effectuer ce geste en tempo 92 BMP. Une vidéo en 19 fps a été enregistrée pour environ 15 sec, soit 285 images au total. La main du musicien était pivotée par angle $\vartheta \in [15,20]$ et le taux de couverture de la peau sur l'image très basse. Le système a identifié avec succès tous les doigts pour la totalité des images de la vidéo ainsi que tous les appuis du musicien.

PianOrasis:max a été ensuite évalué sur huit vidéos (19 fps) d'une durée entre 15 et 20 sec et d'une couverture de la silhouette de la main sur l'image de l'ordre de 14% (Figure 4). Il s'agit des gestes de a) gamme ascendante, b) gamme descendante, c) arpège ascendant et d) arpège descendant toujours pour la gamme de Do. En comparant le temps moyen du traitement entre PianOrasis et PianOrasis:max pour les mêmes vidéos nous constatons une diminution de 31,7%. Cela est dû à la procédure du découpage de la RI aux limites de la silhouette de la main sur l'image binaire, procédure n'étant pas disponible à la version antérieure. Pourtant le taux de reconnaissance reste relativement constant entre les deux versions (*83,64% pour PianOrasis et 81,64% pour PianOrasis:max*)

Vidéo	Durée (sec)	Temps de traitement sans découpage (sec)	Temps de traitement avec découpage (sec)	Reconnaissance PianOrasis (%)	Reconnaissance PianOrasis:max (%)	Réduction du temps (%)
A	9	98.51	66.73	87.5	100	32.3
B	9	94.89	66.40	75	87.5	30
C	11	120.15	82.95	77.7	89	30.9
D	11	114.92	79.82	89	77.7	30.5
E	9	99	66.67	85.7	85.7	32.5
F	15	157.39	105.36	90.9	81.8	33.1
G	7	81.78	55.68	80	80	31.9
H	7	76.75	51.86	83.3	66.6	32.4

Figure 4. Trajectoires des bouts des doigts

Les trajectoires des cinq bouts des doigts du geste de la vidéo E apparaissent dans la figure 5. L'axe vertical représente la distance entre les coordonnées du bout du doigt et celles du centroïde tandis que l'axe horizontal représente les images de la séquence.

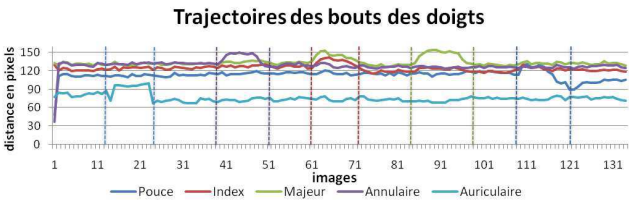

Figure 5. Trajectoires des bouts des doigts

Lorsque la distance entre le bout d'un doigt et le centroïde dépasse de 20 pixels le seuil, un doigté est enregistré. Les doigtés du geste de la figure 4 sont

indiqués par les lignes pointillées verticales de même couleur que la trajectoire du doigt. Il est facile de constater que le geste de cette vidéo correspond à une inversion majeur/index. Par ailleurs, le rythme de l'interprétation n'est pas constant et la position d'un doigt est souvent influencée par le geste des doigts adjacents.

8. CONCLUSION

Une méthodologie de vision par ordinateur pour la reconnaissance des gestes des doigts dans l'espace a été présentée. Actuellement, une nouvelle version du système PianOrasis:max, est disponible, en proposant de la reconnaissance des doigtés en temps réel. Des nouvelles techniques d'invariance d'échelle et de rotation ont été introduites, rendant ainsi le système moins dépendant de la distance entre la main et la caméra et des petites rotations de la main. Une méthodologie pour le contrôle gestuel du son pour chaque doigt indépendamment a été présentée. Elle sera mise en œuvre dans un très proche avenir. Le système a été évalué pour une gamme de gestes pianistiques effectués en espace sans aucune interférence entre les doits du musicien et l'ordinateur. Cela constitue une bonne base de recherche pour le développement d'un instrument de musique immatériel, un « espace interactif » pour la production du son, où le musicien ne sera pas limité par une interface spécifique. Un geste naturel effectué dans un environnement réel pourra déclencher un ensemble d'entités musicales, tels que le son d'un instrument musical ou même le segment d'une voix parlée. La connaissance des notions approfondies en musique ne constituera pas un préalable pour l'utilisation d'un tel instrument musical.

9. REFERENCES

[1] Manitsaris, S. *Vision par ordinateur pour la reconnaissance des gestes : analyse et modélisation stochastique du geste dans l'interaction musicale*, Thèse de Doctorat, Université de Macédoine, Thessalonique, Grèce, 2010.

[2] ViconPeak. *Vicon Motion Capture System*, Lake Forest, Ca, 2005.

[3] Palmer, C. & Pfordresher, P. Q. Frommy hand to yourear: the faces of meter in performance and perception. In C. Woods, G. Luck, R. Brochard, F. Seddon& J. A. Sloboda (Eds.) In *Proceedings of the 6th International Conference on Music Perception and Cognition*. Keele, UK: Keele University, 2000.

[4] Rasamimanana, N. & Bevilacqua., F. Effort-based analysis of bowing movements: evidence of anticipation effects. *The Journal of New Music Research*, 37(4):339 – 351, 2009.

[5] Coduys, T., Henry, C. and Cont, A. TOASTER and KROONDE: High-Resolution and High-Speed Real-time Sensor Interfaces, In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME-04)*, Hamamatsu, Japan, 2004.

[6] Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., Rasamimanana, N. Continuous real time gesture following and recognition, *LNAI 5934*, pp. 73–84, 2010.

[7] N. Rasamimanana, F. Bevilacqua, N. Schnell, F. Guedy,, E. Come Maestracci, B. Zamborlin, JL. Frechin,U. Petrevski, « Modular Musical Objects Towards Embodied Control Of Digital Music », In *Proceedings of Tangible Embedded and Embodied Interaction*, 2011.

[8] S. Hadjidimitriou, A. Zacharakis, P. Doulgeris, K. Panoulas, L. Hadjileontiadis, and S. Panas, "Revealing action representation processes in audio perception using Fractal EEG Analysis: A Mirror Neuron System-Based Approach", *IEEE Transactions on Biomedical Engineering* , VOL. 58, NO. 4, pp. 1120-1129, 2011.

[9] Baum, L. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In *Proceedings of the Third Symposium on Inequalities*, New York, USA, 1972.

[10] Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition». In *Proceedings of the IEEE*, 77(2), 257-285, 1989.

[11] Papamarkos, N., Strouthopoulos, C., &Andreadis, I. *Multithresholding of color and gray level images through a neural network technique", Image and Vision Computing*, vol. 18, 213-222, 2000.

[12] Manitsaris S. «Vision par ordinateur pour la reconnaissance des gestes musicaux des doigts : le système PianOrasis», *Revue Francophone de l'Informatique Musicale*, 1(1), MSH Paris Nord, 2011.

[13] Tsagaris, A., Manitsaris S., Dimitropoulos, K., Manitsaris, A. «Scale and rotation invariance for the recognition of finger musical gestures performed in space», In Proc. *Of the Third European Workshop on Visual Information Processing (EUVIP)*, Paris, France, 2011.

[14] Matsoukas, V., Manitsaris S., Manitsaris, A. «Finger gesture control of sound», In *Proc. Of the Third European Workshop on Visual Information Processing (EUVIP)*, Paris, France, 2011.