

AVIS DE SOUTENANCE DE THESE

Monsieur Qi WANG est autorisé à présenter ses travaux en vue de l'obtention du diplôme national de DOCTORAT délivré par l'école CENTRALE MARSEILLE

le 9 juillet 2018, à 14H00

à : Amphi Massiani, Campus Saint Charles, 13003, Marseille

Titre: **Statistical Models for Human Motion Synthesis**

Ecole doctorale : ED 184 Mathématiques et Informatique de Marseille

Spécialité : Informatique

Rapporteurs :

Monsieur Thierry PAQUET, Professeur, LITIS, Université de Rouen, France.

Monsieur Frédéric BEVILACQUA, DR IRCAM, Paris, France.

Membres du Jury :

Monsieur Thierry PAQUET, Professeur, LITIS, Université de Rouen, France.

Monsieur Frédéric BEVILACQUA, DR IRCAM, Paris, France.

Madame Catherine PÉLACHAUD, Directeur de Recherche CNRS, Institut des Systèmes Intelligents et de Robotique, UPMC Sorbonne Université, France.

Monsieur Ludovic DENOYER, Professeur, LIP6, UPMC Sorbonne Université, France.

Monsieur Benoît FAVRE, Maître de conférences, LIS, Aix-Marseille Université, France.

Monsieur Christophe BOURDIN, Professeur, Institut des Sciences du Mouvement, AMU, France.

Monsieur Thierry ARTIERES, Professeur, Ecole Centrale de Marseille, France.

Résumé :

Cette thèse porte sur la synthèse de séquences de motion capture avec des modèles statistiques. La synthèse de ce type de séquences est une tâche pertinente pour des domaines d'application divers tels que le divertissement, l'interaction homme-machine, la robotique, etc. Du point de vue de l'apprentissage machine, la conception de modèles de synthèse consiste à apprendre des modèles génératifs, ici pour des données séquentielles. Notre point de départ réside dans deux problèmes principaux rencontrés lors de la synthèse de données de motion capture, assurer le réalisme des positions et des mouvements, et la gestion de la grande variabilité dans ces données. La variabilité vient d'abord des caractéristiques individuelles, nous ne bougeons pas tous de la même manière mais d'une façon qui dépend de notre personnalité, de notre sexe, de notre âge, de notre morphologie, et de facteurs de variation plus court terme tels que notre état émotionnel, que nous soyons fatigués, etc.

Divers modèles statistiques ont été étudiés pour la synthèse de ce type de données. Les modèles sont appris sur des jeux de données étiquetés où les séquences de motion capture sont réalisées par des acteurs exécutant diverses activités comme marcher, danser, courir, etc. Notre travail porte sur ce domaine de recherche et concerne en particulier la conception de modèles générateurs de séquences capables de prendre en compte certaines informations contextuelles pour permettre une génération contrôlée de séquences, et capables de générer des séquences réalistes avec une certaine diversité.

Une première partie présente des travaux préliminaires que nous avons réalisés en étendant des approches de l'état de l'art basées sur des modèles de Markov cachés et des processus gaussiens pour aborder les deux problèmes principaux liés au réalisme et à la variabilité. Nous décrivons d'abord une variante de modèles de Markov cachés contextuels pour gérer la variabilité dans les données en conditionnant les paramètres des modèles à une information contextuelle supplémentaire telle que l'émotion avec laquelle un mouvement a été effectué. Nous proposons ensuite une variante d'une méthode de l'état de l'art utilisée pour réaliser une tâche de synthèse de mouvement spécifique appelée Inverse Kinematics, où nous exploitons les processus gaussiens pour encourager le réalisme de chacune des postures d'un mouvement généré. Nos résultats montrent un certain potentiel de ces modèles statistiques pour la conception de systèmes de synthèse de mouvement humain. Pourtant, aucune de ces technologies n'offre la flexibilité apportée par les réseaux de neurones et la récente révolution de l'apprentissage profond et de l'apprentissage Adversarial que nous abordons dans la deuxième partie.

La deuxième partie de la thèse décrit les travaux que nous avons réalisés avec des réseaux de neurones et des architectures profondes. Nos travaux s'appuient sur la capacité des réseaux neuronaux récurrents à traiter des séquences complexes et sur l'apprentissage Adversarial qui a été introduit très récemment dans la communauté du Deep Learning pour la conception de modèles génératifs performants pour des données complexes, notamment images. Nous proposons une première architecture simple qui combine l'apprentissage Adversarial et des autoencodeurs de séquences, qui permet de mettre au point des systèmes performants de génération aléatoire de séquences réalistes de motion capture. A partir de cette architecture de base, nous proposons plusieurs variantes d'architectures neuronales conditionnelles qui permettent de concevoir des systèmes de synthèse que l'on peut contrôler dans une certaine mesure en fournissant une information de haut niveau à laquelle la séquence générée doit correspondre, par exemple l'émotion avec laquelle une activité est réalisée. Pour terminer nous décrivons une dernière variante qui permet de réaliser de l'édition de séquences de motion capture, où le système construit permet de générer une séquence dans le style d'une autre séquence, réelle.