

ΠΡΟΣ

- 1) **Όλα τα μέλη ΔΕΠ του Τμήματος Επιστήμης Υπολογιστών**
- 2) **Τους εκπροσώπους των Μεταπτυχιακών φοιτητών του Τμήματος Επιστήμης Υπολογιστών**
- 3) **Την Επταμελή Εξεταστική Επιτροπή**
- 4) **Όλα τα μέλη της Πανεπιστημιακής Κοινότητας**

Πρόσκληση σε Δημόσια Παρουσίαση της Διδακτορικής Διατριβής του

κ. Οικονομίδη Ιάσων

Την Δευτέρα, 12 Ιανουαρίου 2015 και ώρα 10:00 στην αίθουσα Κ206 τηλεδιάσκεψης του Τμήματος Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης στο Ηράκλειο, θα γίνει η δημόσια παρουσίαση και υποστήριξη της Διδακτορικής Διατριβής του υποψηφίου διδάκτορος του Τμήματος Επιστήμης Υπολογιστών κ. Οικονομίδη Ιάσων με θέμα:

"Αποδοτική Παρακολούθηση της 3D Αρθρωτής Κίνησης του Ανθρώπινου Χεριού"

"Efficient Tracking of the 3D Articulated Motion of Human Hands "

ΠΕΡΙΛΗΨΗ

Το πρόβλημα της τρισδιάστατης παρακολούθησης του ανθρώπινου χεριού έχει τόσο θεωρητικό όσο και πρακτικό ενδιαφέρον. Είναι ένα απαιτητικό πρόβλημα που δεν έχει λυθεί στην πλήρη γενικότητά του, παρά τη σημαντική ερευνητική προσπάθεια που έχει αφιερωθεί σε αυτό. Αυτή η διατριβή αντιμετωπίζει αυτό το πρόβλημα και παρουσιάζει μεθόδους για την παρακολούθηση της 3D θέσης της παλάμης του χεριού και των δακτύλων σε ένα ευρύ φάσμα από ενδιαφέροντα σενάρια.

Τέτοια σενάρια περιλαμβάνουν την παρακολούθηση ενός ή δύο χεριών, καθώς και την παρακολούθηση του χεριού(-ών) μεμονωμένα ή σε αλληλεπίδραση με το περιβάλλον. Επιλογές σχετικές με τη σχεδίαση των διάφορων παρουσιαζόμενων

μεθόδων αφορούν στην επιλογή κατάλληλων χαρακτηριστικών εικόνας συμπεριλαμβάνοντας τον τρόπο με τον οποίο αυτά μπορούν να συντεθούν και να αποτιμηθούν, καθώς και αλγόριθμους για την επίλυση των προβλημάτων βελτιστοποίησης που προκύπτουν. Όλα τα σενάρια προβλέπουν σαν είσοδο οπτική παρατήρηση της σκηνής χωρίς χρήση υποβοηθητικών σημαδιών. Τα χαρακτηριστικά εικόνας που χρησιμοποιούμε είναι οι ακμές, οι περιοχές χρώματος δέρματος, η απόσταση από τον αισθητήρα και το τρισδιάστατο οπτικό περίγραμμα (visual hull). Οι παρατηρήσεις μπορούν να προέρχονται είτε από ένα δίκτυο συμβατικών καμερών, είτε από μία κάμερα που επιπρόσθετα με το χρώμα καταγράφει και την απόσταση του κάθε σημείου της σκηνής από τον αισθητήρα (RGB-D sensor). Η επιλογή του τύπου εισόδου καθορίζει μερικώς και τα χρησιμοποιούμενα χαρακτηριστικά εικόνας.

Ακολουθούμε την προσέγγιση μεθόδων που βασίζονται σε μοντέλο, διατυπώνοντας το πρόβλημα της εκτίμησης πόζας σε κάθε εικόνα εισόδου σαν ένα πρόβλημα βελτιστοποίησης. Ο χώρος αναζήτησης αυτού του προβλήματος βασίζεται στη χρησιμοποιούμενη παραμετροποίηση της κινηματικής του χεριού. Για την περίπτωση του ενός χεριού, ο χώρος αναζήτησης ταυτίζεται με αυτή την παραμετροποίηση, ενώ για τις περιπτώσεις αλληλεπίδρασης χεριού-χεριού ή χεριού-αντικειμένου, αυτός ο χώρος προσαυξάνεται κατάλληλα ώστε να συμπεριλάβει όλες τις παρακολουθούμενες οντότητες. Αυτή η από κοινού θεώρηση, παρότι οδηγεί σε προβλήματα βελτιστοποίησης με δεκάδες παραμέτρων, έχει το πλεονέκτημα ότι επιτρέπει την μοντελοποίηση της αλληλεπίδρασης των παρακολουθούμενων οντοτήτων με άμεσο τρόπο. Η υπόθεση της χρονικής συνέχειας χρησιμοποιείται μέσω της αρχικοποίησης της αναζήτησης σχετικά με κάποια εικόνα στην περιοχή της εκτίμησης λύσης για την προηγούμενη χρονικά εικόνα.

Η από κοινού θεώρηση των παρατηρούμενων οντοτήτων της σκηνής επιτρέπει την αντιμετώπιση σεναρίων που περιλαμβάνουν πολύπλοκη αλληλεπίδραση ανάμεσα σε αυτές τις οντότητες. Για την περίπτωση της αλληλεπίδρασης χεριού με αντικείμενο, δείχνουμε πώς οι προκύπτουσες αλληλεπικαλύψεις μπορούν να παράσχουν χρήσιμη πληροφορία αντί να αντιμετωπίζονται ως πρόβλημα. Για την περίπτωση των δύο χεριών σε ισχυρή αλληλεπίδραση, οι αλγόριθμοι που προτείνουμε αντιμετωπίζουν την πιο περίπλοκη αλληλεπίδραση χεριών που έχει ως τώρα αναφερθεί στη σχετική βιβλιογραφία.

Για τη βελτιστοποίηση των αντικειμενικών συναρτήσεων, όπως προκύπτουν από την υιοθετούμενη διατύπωση του προβλήματος, χρησιμοποιούμε αλγόριθμους βελτιστοποίησης που δεν απαιτούν γνώση της παραγώγου της αντικειμενικής συνάρτησης. Συγκεκριμένα, στις περισσότερες περιπτώσεις, χρησιμοποιούνται παραλλαγές του αλγορίθμου Βελτιστοποίησης Σμήνους Σωματιδίων (BSS) (Particle Swarm Optimization). Ο BSS είναι ένας γενετικός αλγόριθμος που δεν απαιτεί γνώση της παραγώγου της αντικειμενικής συνάρτησης που βελτιστοποιεί, και παραλληλοποιείται εύκολα. Είναι κατάλληλος για το πρόβλημα διότι μπορεί να αντιμετωπίσει μη παραγωγίσιμες συναρτήσεις με πολλά τοπικά βέλτιστα. Παρουσιάζεται επίσης ένας νέος εξελικτικός αλγόριθμος βελτιστοποίησης, και δοκιμάζεται σε δύο από τα εξεταζόμενα σενάρια παρακολούθησης της κίνησης

χειρών. Αυτός ο αλγόριθμος εκμεταλλεύεται τις χρήσιμες ιδιότητες της ημι-τυχαίας δειγματοληψίας, συνδυάζοντάς τις με την δύναμη των εξελικτικών υπολογισμών.

Τα διάφορα υπολογιστικά βήματα όλων των παρουσιαζόμενων μεθόδων είναι προσεκτικά σχεδιασμένα ώστε να περιλαμβάνουν υπολογισμούς που επιδέχονται παραλληλοποίηση. Γίνεται έτσι εφικτή η εκμετάλλευση σύγχρονων αρχιτεκτονικών όπως οι κάρτες γραφικών, έτσι ώστε τα συστήματα που προκύπτουν να επιτυγχάνουν επιδόσεις οι οποίες, ανάλογα με το πρόβλημα, είναι πραγματικού χρόνου ή κοντά σε αυτές.

Επόπτης Διδακτορικής Διατριβής: Καθηγητής, Αντώνιος Αργυρός

ABSTRACT

The problem of hand pose estimation and tracking is both theoretically and practically interesting. It is a challenging problem that hasn't been solved in its full generality despite the significant amount of effort that has been devoted to it. This thesis presents methods to track the position, orientation and full articulation of human hands in various everyday scenarios.

Investigated scenarios include tracking one or two hands and tracking the hand(s) in isolation or in interaction with the environment. Design choices for the various presented methods regard the type of input, the selection of appropriate visual cues and furthermore the way they are synthesized and evaluated, as well as the optimization algorithms used to solve the formulated optimization problems. All scenarios use markerless visual observations of the scene as input. We explore the visual cues of skin color, edges, depth map, and visual hull. These observations can come either from a network of cameras or from an RGB-D sensor. The choice of input type partially mandates the visual cues that are employed.

We follow a model-based approach to the problem, formulating the pose estimation task for each frame as an optimization problem. The search space of this problem uses the adopted representation for the hand kinematics. For the case of single hand, the search space is this set of kinematics parameters, whereas for hand-object or hand-hand interaction, this search space is appropriately augmented to include all the tracked entities. This joint consideration, while resulting in optimization problems with tens of parameters, has the advantage that the interaction between the tracked objects can be effortlessly modeled and evaluated. The temporal continuity

assumption is used by initializing the search for a frame near the solution for the previous frame.

Joint modeling of the observed entities in the scene allows for effortlessly treating scenarios of complex interaction between these entities. For the case of hand-object interaction, we show how the observed occlusions can provide useful information instead of being an obstacle. For the case of two hands in strong interaction, to the best of our knowledge, the presented results involve the most complex hand-hand interaction attempted so far in the relevant literature.

For the task of optimizing the objective functions that result from the adopted formulation of the problem, we use black-box optimization algorithms. Specifically, variants of Particle Swarm Optimization (PSO) are employed in most scenarios. PSO is an evolutionary optimization algorithm that is derivative-free and easily parallelizable. It is suitable for our task, since it is well-suited to multi-modal, non-differentiable objective functions. A novel evolutionary optimization algorithm is also presented in this thesis, and applied to two of the examined scenarios. This algorithm exploits the useful properties of quasi-random sampling, as well as the power of evolutionary computing.

The various computational steps of all presented methods are carefully designed so that they include parallelizable computations. It is then possible to make use of modern hardware such as the GPU architecture, resulting in practical systems that achieve real-time or interactive frame-rates.

Supervisor: Professor, Antonis Argyros

Panagiotis Tsakalides

Chairman

Department of Computer Science