

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. Κυριαζή Νικολάου

Την Παρασκευή, 13 Ιουνίου 2014 και ώρα 16:00 στην αίθουσα Τηλεδιάσκεψης Κ206 του Πανεπιστημίου Κρήτης στο Ηράκλειο, θα γίνει η δημόσια παρουσίαση και υποστήριξη της Διδακτορικής Διατριβής του υποψηφίου διδάκτορος του Τμήματος Επιστήμης Υπολογιστών κ. Κυριαζή με θέμα:

“Ένα υπολογιστικό πλαίσιο για την παρατήρηση και κατανόηση της αλληλεπίδρασης ανθρώπων με αντικείμενα του περιβάλλοντός τους”

“A computational framework for observing and understanding the interaction of humans with objects of their environment”

ΠΕΡΙΛΗΨΗ

Εστιάζουμε στο πρόβλημα της κατανόησης μιας δυναμικής σκηνής με βάση οπτική πληροφορία, δηλαδή στον μετασχηματισμό μιας τέτοιας σκηνής σε μια συμβολική αναπαράσταση, την οποία να μπορεί να επεξεργαστεί ένα υπολογιστικό σύστημα. Ενδιαφερόμαστε για σκηνές εσωτερικού χώρου, στις οποίες ένας άνθρωπος αλληλεπιδρά σκόπιμα με το περιβάλλον.

Παρατηρούμε ότι οι έως τώρα σχετικές προσεγγίσεις πραγματοποιούσαν κατανόηση σκηνής μέσω κυρίως χονδρικής μοντελοποίησης της παρατηρούμενης διαδικασίας, καθώς λεπτομερέστερες μοντελοποιήσεις είναι πολύ απαιτητικές σε υπολογιστικούς πόρους και δυσκολεύουν την απαιτούμενη ενοποίηση με υπολογιστικές μεθόδους όρασης.

Υποστηρίζουμε πως αυτήν τη στιγμή είναι όντως δυνατό να εκμεταλλευτούμε λεπτομερείς μοντελοποιήσεις, που να ενοποιούνται εύκολα με υπολογιστικές τεχνικές όρασης και να ανταπεξέρχονται στις σχετικές υπολογιστικές απαιτήσεις. Σε

ότι αφορά την κατανόηση σκηνής, είμαστε σε θέση να μοντελοποιήσουμε και να προσομοιώσουμε τόσο την διαδικασία ανάκτησης εικόνων μέσω 3D rendering (παρουσιαστικό), όσο και την δυναμική των παρατηρούμενων διεργασιών μέσω προσομοίωσης φυσικής (συμπεριφορά). Έτσι, προσδιορίζουμε το 3D rendering και την προσομοίωση φυσικής σαν δύο σημαντικές διεργασίες για την κατανόηση σκηνής και προτείνουμε τον συνδυασμό της προσομοιωτικής δυνατότητας σχετικών υπολογιστικών μεθόδων με ισχυρές μεθόδους βελτιστοποίησης προς την ανάδειξη αποτελεσματικών εργαλείων συμπερασμού.

Ειδικότερα, θεωρούμε την διαδικασία κατανόησης μιας δυναμικής σκηνής ως ένα πρόβλημα βελτιστοποίησης. Σχεδιάζουμε παραμετρικά μοντέλα που περιγράφουν το τι είναι δυνατόν να διαδραματιστεί σε μια σκηνή και πως αυτό μπορεί να παρατηρηθεί από τα διαθέσιμα οπτικά μέσα. Ορίζουμε σαν πεδίο ορισμού του προβλήματος βελτιστοποίησης τις παραμέτρους καθαυτές. Η βελτιστοποίηση πραγματοποιείται σε ξεχωριστή διαδικασία από αυτή της μοντελοποίησης, με υπόθεση-και-δοκιμή, μέσω μεθόδων βελτιστοποίησης black-box. Το αποτέλεσμα της βελτιστοποίησης είναι εκείνη η παραμετροποίηση των μοντέλων που «εξηγούν» με βέλτιστο τρόπο τις παρατηρήσεις. Οι υποθέσεις που δοκιμάζονται είναι σε συμφωνία με κανόνες φυσικής αφού πηγάζουν από προσομοιωτές φυσικής. Για κάθε υπόθεση αποτιμάται η συμβατότητά της με πραγματικές παρατηρήσεις μέσω 3D rendering. Έτσι, η πρότασή μας εστιάζει σε τρία σημεία: (α) μοντελοποίηση σκηνής, (β) ενσωμάτωση προσομοίωσης φυσικής και (γ) εκμετάλλευση των μεθόδων βελτιστοποίησης black-box.

Έχουμε αναπτύξει ένα υπολογιστικό πλαίσιο που βασίζεται στα παραπάνω για να επιτυγχάνει επίλυση επιμέρους προβλημάτων κατανόησης μιας τρισδιάστατης σκηνής. Παρουσιάζουμε αυτό το πλαίσιο και τις εφαρμογές του σε προβλήματα τρισδιάστατης παρακολούθησης και εκτίμησης κίνησης σε σκηνές εσωτερικού χώρου. Δίνουμε έμφαση στην αναγκαιότητα για ένταξη φυσικής. Πιο ειδικά, δείχνουμε ότι με το να αναγνωρίζουμε ότι οι οπτικές παρατηρήσεις αφορούν φυσικά φαινόμενα που εξηγούνται από κανόνες φυσικής, μπορούμε να εφαρμόσουμε συμπερασμό ακόμα και σε αρχικά «κρυφές» παραμέτρους. Επομένως, μπορούμε να εφαρμόσουμε λογισμό σε παραμέτρους που πριν την ένταξη φυσικής δεν ήταν άμεσα παρατηρήσιμες και τις οποίες μπορούμε να ανακτήσουμε μόνο με τη θεώρηση φυσικών φαινομένων και των συνεπειών τους.

Το προτεινόμενο υπολογιστικό πλαίσιο έχει χρησιμοποιηθεί για τη λύση προβλημάτων που ποικίλουν από την παρακολούθηση ενός αντικειμένου έως την παρακολούθηση δύο χεριών καθώς αυτά αλληλεπιδρούν με πολλά αντικείμενα, στις τρεις διαστάσεις και με βάση παρατηρήσεις που προέρχονται από διάφορα οπτικά μέσα. Μέσα από μια σειρά πειραμάτων δείχνουμε τη σημασία της ενσωμάτωσης γραφικών υπολογιστών και προσομοίωσης φυσικής στην τρισδιάστατη κατανόηση σκηνής. Οι ανωτέρω διαδικασίες χρησιμοποιήθηκαν επιτυχώς σαν προσομοιωτές black-box, χωρίς η εγγενής πολυπλοκότητά τους να εμποδίσει την ενοποίηση με υπολογιστικές μεθόδους όρασης, χάρη στη σχεδιαστική επιλογή της εμπλοκής μεθόδων βελτιστοποίησης black-box. Δείχνουμε επίσης ότι το προτεινόμενο πλαίσιο επιδεικνύει καλά χαρακτηριστικά ως προς την κλιμακώσιμη αντιμετώπιση

προβλημάτων μεγάλης πολυπλοκότητας. Μέσω προσεκτικής σχεδίασης, η επίκληση έως τώρα υπολογιστικά ακριβών προσομοιώσεων μπορεί να επιτευχθεί τόσο αποδοτικά ώστε να επιτυγχάνεται επεξεργασία σε γρήγορους ρυθμούς. Τα παραπάνω συνηγορούν υπέρ μιας αρθρωτής υπολογιστικής λύσης σε προβλήματα τρισδιάστατης παρακολούθησης σκηνής, με ξεκάθαρη δυνατότητα για βελτίωση ή γενίκευση: αντικαθιστώντας μέρη με καλύτερες ή γενικότερες υλοποιήσεις βελτιώνεται αυτόματα το σύνολο.

Επόπτης Διδακτορικής Διατριβής: Καθηγητής, Αντώνης Αργυρός

ABSTRACT

We focus on the problem of vision-based scene understanding, i.e. “lifting” a scene which is observed by visual means across time, to a symbolic representation that can be processed by a computational system. We are interested in dynamic indoor scenes, in which humans purposefully interact with their environment.

We observe that existing approaches have been performing scene understanding mainly through coarse modelling of the observed processes, as more detailed modelling is very demanding in terms of computational resources and exhibits difficulties with respect to the required integration of computer vision methods.

We suggest that currently, it is indeed feasible to incorporate detailed scene modelling, which can be easily integrated with computer vision techniques and can efficiently cope with the associated computational requirements. With respect to scene understanding, we are in position to model and simulate the process of image acquisition through 3D rendering (appearance), and the dynamics of the observed processes through physics simulation (behavior). Thus, we identify 3D rendering and physics simulation as two significant processes towards scene understanding. We propose the combination of the simulation power of these tools with powerful optimization methods, in order to yield powerful inference tools towards scene understanding.

More specifically, we consider the process of scene understanding as an optimization problem. We design parametric models that describe what can take place in a dynamic scene and how this can be observed by visual means. We define these parameters to constitute the domain of the optimization problem. Optimization is decoupled from modelling and is performed in a hypothesize-and-test framework which is implemented based on black box optimization techniques. The outcome of the optimization is the instance of the parametric models which best “explain” the observations. Ultimately, in the context of this work, the tested hypotheses are in agreement with laws of physics as they originate from physics simulators. For every hypothesis, its compatibility with actual observations of the scene is evaluated through 3D rendering. Thus, our proposal focuses on three points: (a) forward

modelling of the scene, (b) incorporation of physics simulation and (c) exploitation of black-box optimization methods.

We have developed a computational framework which, based on the above, performs aspects of 3D scene understanding. We present this framework and its application to the problems of 3D tracking and motion estimation. We emphasize the necessity for the incorporation of physics. More specifically, we show that by acknowledging that visual observations regard physical phenomena governed by laws of physics, we can even apply inference on initially "hidden" parameters. More specifically, we can estimate parameters that prior to incorporating physics were not directly observable, and which can be recovered only by attributing observations to side-effects of physical processes.

The proposed computational framework has been employed to solve problems that vary from tracking a single object to tracking two hands while interacting with many objects, in 3D and from different visual modalities and camera arrangements. Through a series of experiments we show how important it is to incorporate computer graphics and physics processes in 3D scene understanding. These processes were successfully used as black box simulation tools and their inherent complexity has not hindered the integration with computer vision processes, thanks to the design choice of employing black-box optimization. We were also able to show that the proposed framework exhibits a favorable scalability profile when applied to domains of increasing complexity. Through careful design, the invocation of otherwise expensive simulations can be performed so efficiently that interactive processing frame rates are achieved. All the above advocate a modular computational solution to 3D scene understanding problems with a clear potential for improvement or generalization: substituting parts with better or more general modules automatically improves the entire framework.

Supervisor: Professor Antonis Argyros