

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Μανουσάκη Βικτωρία
Μεταπτυχιακή Φοιτήτρια**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Καθηγητής Α. Αργυρός**

Δευτέρα, 2/10/2017, 10:00

Αίθουσα Κ206, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

“Αξιολόγηση σχεδιαστικών επιλογών μεθόδων βασιζόμενων σε συλλογή οπτικών λέξεων για την κατηγοριοποίηση ανθρώπινων δραστηριοτήτων”

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια, τα προβλήματα της ανάλυσης και της κατηγοριοποίησης / αναγνώρισης των ανθρώπινων δράσεων με βάση οπτική πληροφορία έχουν προσελκύσει μεγάλο ερευνητικό ενδιαφέρον εξαιτίας της σημασίας της επίλυσής τους σε τομείς όπως η υποβοηθούμενη διαβίωση, η επιτήρηση χώρων, η αλληλεπίδραση ανθρώπου-υπολογιστή/ρομπότ, κλπ. Παρά τη σημαντική πρόοδο που έχει επιτευχθεί, η κατηγοριοποίηση και η αναγνώριση των ανθρώπινων δράσεων παραμένει ένα δύσκολο πρόβλημα που δεν έχει επιλυθεί στη γενικότητά του.

Σε αυτή την εργασία, μας ενδιαφέρει η κατηγοριοποίηση ανθρώπινων δράσεων με βάση 3D σκελετικά δεδομένα και βασιζόμαστε στη μέθοδο της Συλλογής Οπτικών Λέξεων - ΣΟΛ (Bags of Visual Words - BoVWs) που αποτελεί ένα αρκετά επιτυχημένο πλαίσιο κωδικοποίησης χαρακτηριστικών για την επίλυση αυτού του προβλήματος. Ακολουθούμε ένα σχήμα κατηγοριοποίησης δράσεων που αποτελείται από τρία βασικά

βήματα: (α) εξαγωγή χαρακτηριστικών, (β) αναπαράσταση / κωδικοποίηση βασισμένη σε συλλογή κωδικοποιημένων οπτικών χαρακτηριστικών και (γ) κατηγοριοποίηση των παραστάσεων δράσης που προκύπτουν. Σε αυτή τη μελέτη, ο στόχος μας είναι να παρέχουμε μια πειραματική αξιολόγηση διαφόρων σχεδιαστικών επιλογών για τα παραπάνω βήματα, οι οποίες, εφόσον προσδιοριστούν, οδηγούν σε συγκεκριμένες και διαφορετικές μεθόδους κατηγοριοποίησης δράσεων.

Σε αυτή την κατεύθυνση, οι συνεισφορές αυτής της εργασίας είναι πολλές. Αρχικά, ερευνούμε την απόδοση τριών υφιστάμενων τύπων 3D σκελετικών χαρακτηριστικών. Προτείνουμε επίσης ένα νέο χαρακτηριστικό για την αναπαράσταση της 3D στάσης του ανθρώπινου σώματος που εμπνέεται από υφιστάμενη έρευνα σε Ιστογράμματα Ορίου Κίνησης - ΙΟΚ (Motion Boundary Histograms – MBH). Η χρήση του προτεινόμενου χαρακτηριστικού φαίνεται να παράγει αποτελέσματα που είναι ανταγωνιστικά με αυτά των κορυφαίων υφιστάμενων ανταγωνιστικών μεθόδων.

Εξετάζουμε τρεις διαφορετικές μεθόδους κατηγοριοποίησης (K-Κοντινότεροι Γείτονες - ΚΚΓ , Μηχανές Διανυσμάτων Υποστήριξης - ΜΔΥ, Νευρωνικό Δίκτυο Συνάρτησης Ακτινικής Βάσης - ΝΔΣΑΒ). Διερευνούμε επίσης το μέγεθος της συλλογής κωδικών που χρησιμοποιείται για την αναπαράσταση δράσεων, το οποίο αποτελεί σημαντική σχεδιαστική επιλογή σε μεθόδους που βασίζονται σε ΣΟΛ. Για να επιτευχθεί αυτό, βασιζόμαστε σε μια εμπειρική, σχεδόν εξαντλητική μελέτη για να καθορίσουμε το καλύτερο μέγεθος συλλογής κωδικών για κάθε τύπο χαρακτηριστικού και μεθόδου κατηγοριοποίησης. Οι περισσότερες υφιστάμενες εργασίες ορίζουν ένα συγκεκριμένο μέγεθος συλλογής κωδικών χωρίς να παρέχουν λεπτομέρειες για το πώς αυτό αποφασίζεται. Αντιθέτως, εμείς διερευνούμε μεθόδους που καθορίζουν αυτόματα το μέγεθος της συλλογής. Τα αποτελέσματα της διερεύνησης δείχνουν ότι η μέθοδος Διάδοσης Συνάφειας (Affinity Propagation), μια τεχνική μη επιτηρούμενης ομαδοποίησης που καθορίζει αυτόματα τον αριθμό των συστάδων σε ένα σύνολο δεδομένων, μπορεί να χρησιμοποιηθεί αποτελεσματικά ως αντικατάσταση του αλγόριθμου K-μέσων (k-Means) που χρησιμοποιείται στις περισσότερες από τις μεθόδους κατηγοριοποίησης που βασίζονται σε ΣΟΛ. Επιπρόσθετα, διερευνάμε εναλλακτικές προσεγγίσεις κωδικοποίησης χαρακτηριστικών όπως οι Συλλογή Χρονικών Λέξεων - ΣΧΛ (Bag of Temporal Words - BoTW) και τα Διανύσματα Τοπικά Συσσωρευμένων Περιγραφών - ΔΤΣΠ (Vector of Locally Aggregated Descriptors - VLAD). Τα αποτελέσματα που προκύπτουν δείχνουν ότι η σχετικά απλούστερη κωδικοποίηση ΣΟΛ υπερέρχει έναντι αυτών των πολυπλοκότερων επιλογών.

Όλες οι επιλογές σχεδιασμού μεθόδων ταξινόμησης δράσεων αξιολογήθηκαν βάσει μιας σειράς πειραμάτων που πραγματοποιήθηκαν στο ευρέως χρησιμοποιούμενο σύνολο

δεδομένων κίνησης και ανθρώπινων δραστηριοτήτων Berkeley MHAD. Το σύνολο δεδομένων MHAD είναι επισημειωμένο με τις ορθές κατηγορίες ανθρώπινων δράσεων, γεγονός που επιτρέπει την ποσοτική αποτίμηση των σχετικών αυτοματοποιημένων μεθόδων κατηγοριοποίησης.

Manousaki Victoria

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Professor Antonis Argyros

Monday, 2/10/2017, 10:00

Room K206 , Computer Science Dept., University of Crete

“Evaluating design options of Bag-of Visual-Words based methods for action classification”

ABSTRACT

In recent years, the problems of vision-based human motion analysis and action classification/recognition have attracted a lot of attention due to the significance of their solution in domains such as assisted living, surveillance, human-computer/robot interaction, etc. Despite several breakthroughs, human action recognition remains a challenging problem that is unsolved in its generality.

In this work, we are interested in action classification based on motion capture/skeletal data and we rely on the Bags of Visual Words (BoVWs) features encoding. We follow an action classification framework consisting of three main steps: (a) feature extraction, (b) representation/encoding based on a BoVWs codebook and (c) classification of the resulting action representations. In this study, our goal is to provide an experimental evaluation of various options regarding the selection of the components of this framework that, when instantiated, give rise to a specific action classification method.

In that direction, the contributions of this work are manyfold. First, we investigate the performance of three existing types of 3D skeletal features. We also propose a new

feature for representing human 3D pose data that is inspired by the work on Motion Boundary Histograms (MBH). The use of the proposed feature is shown to produce results that are competitive to the state of the art. We explore three different classification methods (K-Nearest Neighbors, Support Vector Machines, Radial Basis Function Neural Networks). We also investigate the size of the codebook used to represent actions, which is a major design issue in BoVW-based methods. To achieve this, we perform an empirical, almost exhaustive study to determine the best codebook size for each feature type and classifier. Most of the previous works define a specific codebook size without providing details on how this has been decided. In contrast, we explore methods that determine automatically the codebook size. This investigation shows that Affinity Propagation, an unsupervised clustering technique that determines automatically the number of clusters in a dataset, can be used effectively as a replacement of the k-Means algorithm which is used in most of the BoVW-based recognition methods.

Additionally, we explore feature encoding alternatives to BoWs such as the Bag of Temporal Words (BoTW) and the Vector of Locally Aggregated Descriptors (VLAD). The obtained results show that the simple BoVW encoding outperforms these more complicated choices.

All methods and action classification design choices have been evaluated quantitatively based on a series of experiments that have been carried out on the standard, extensive and ground truth-annotated Berkeley MHAD dataset.