A Tensor-Based Deep Learning Framework

Konstantinos Charalampous, Antonios Gasteratos

PII: S0262-8856(14)00129-2
DOI: doi: 10.1016/j.imavis.2014.08.003
Reference: IMAVIS 3352

To appear in: *Image and Vision Computing*

Received date: 22 October 2013
Revised date: 9 May 2014
Accepted date: 4 August 2014

Please cite this article as: Konstantinos Charalampous, Antonios Gasteratos, A Tensor-Based Deep Learning Framework, *Image and Vision Computing* (2014), doi: 10.1016/j.imavis.2014.08.003

A Tensor-Based Deep Learning Framework

Konstantinos Charalampous, Antonios Gasteratos

The authors are with the Department of Production and Management Engineering, Democritus University of Thrace, Xanthi, Greece, 67100.
E-mail: {kchara, agaster}@pme.duth.gr

Abstract

This paper presents an unsupervised deep learning framework that derives spatio-temporal features for human-robot interaction. The respective models extract high-level features from low-level ones through a hierarchical network, viz. the Hierarchical Temporal Memory (HTM), providing at the same time a solution to the curse of dimensionality in shallow techniques. The presented work incorporates the tensor-based framework within the operation of the nodes and, thus, enhances the feature derivation procedure. This is due to the fact that tensors allow the preservation of the initial data format and their respective correlation and, moreover, attain more compact representations. The computational nodes form spatial and temporal groups by exploiting the multilinear algebra, subsequently express the samples according to those groups in terms of proximity. This generic framework may be applied in a diverse of visual data, whilst it has been examined on sequences of color and depth images, exhibiting remarkable performance.

Keywords: Deep Learning, Hierarchical Temporal Memory (HTM), Tensor Algebra, L_1-norm, Support Vector Clustering, Spatio-temporal Features.

1. Introduction

A prerequisite of the Human-Robot interaction is the ability of the robot to comprehend the respective inputs given by humans. Similarly to humans, the most intensive sensory input of a robot is the visual one. The latter suggests that should the robots collaborate with humans they ought to be endowed with subsystems suitable for action, gesture or even emotion recognition. An interest point for such desired recognition tasks is the fact that the corresponding data encapsulate both spatial and temporal information.