KEPLER: Simultaneous Estimation of Keypoints and 3D Pose of Unconstrained Faces in a Unified Framework by Learning Efficient H-CNN Regressors

Amit Kumar, Azadeh Alavi, Rama Chellappa

PII: S0262-8856(18)30154-9
DOI: doi:10.1016/j.imavis.2018.09.009
Reference: IMAVIS 3718

To appear in: *Image and Vision Computing*

Received date: 2 October 2017
Revised date: 17 May 2018
Accepted date: 12 September 2018

Please cite this article as: Amit Kumar, Azadeh Alavi, Rama Chellappa , KEPLER: Simultaneous Estimation of Keypoints and 3D Pose of Unconstrained Faces in a Unified Framework by Learning Efficient H-CNN Regressors. Imavis (2018), doi:10.1016/j.imavis.2018.09.009

KEPLER: Simultaneous Estimation of Keypoints and 3D Pose of Unconstrained Faces in a Unified Framework by Learning Efficient H-CNN Regressors

Amit Kumar, Azadeh Alavi, and Rama Chellappa[1]

Center for Automation Research, UMIACS
University of Maryland, College Park, USA

Abstract

Keypoint detection is one of the most important pre-processing steps in tasks such as face modeling, recognition and verification. In this paper, we present an iterative method for Keypoint Estimation and Pose prediction of unconstrained faces by Learning Efficient H-CNN Regressors (KEPLER) for addressing the unconstrained face alignment problem. Recent state of the art methods have shown improvements in facial keypoint detection by employing Convolution Neural Networks (CNNs). Although a simple feed forward neural network can learn the mapping between input and output spaces, it does not learn the inherent structural dependencies that well. We present a novel architecture called H-CNN (Heatmap-CNN) acting on an N-dimensional input image which captures informative structured global and local features and thus favors accurate keypoint detecion in in-the wild face images. H-CNN is jointly trained on the visibility, fiducials and 3D-pose of the face. As the iterations proceed, the error decreases making the gradients small and thus requiring efficient training of deep networks to mitigate this. KEPLER performs global corrections in pose and fiducials for the first four iterations followed by local corrections at a later stage. As a by-product, KEPLER also provides robust estimate of 3D pose (pitch, yaw and roll) of the face. We also show that without using any 3D in-

[1]Email: {akumar14,azadeh,rama}@umiacs.umd.edu