Accurate and robust face recognition from RGB-D images with a deep learning approach

Yuancheng Lee
http://cv.cs.nthu.edu.tw/php/people/profile.php?uid=150
Jiancong Chen
http://cv.cs.nthu.edu.tw/php/people/profile.php?uid=153
Ching-Wei Tseng
http://cv.cs.nthu.edu.tw/php/people/profile.php?uid=156
Shang-Hong Lai
http://www.cs.nthu.edu.tw/~lai/

Computer Vision Lab,
Department of Computer Science,
National Tsing Hua University,
Hsinchu, Taiwan

In this paper, we propose a face recognition system based on deep learning, which can be used to verify and identify a subject from the colour and depth face images captured with a consumer-level RGB-D camera. To recognize faces with colour and depth information, our system contains 3 parts: depth image recovery, deep learning for feature extraction, and joint classification.

In the depth image recovery and enhancement stage, a pipeline is applied to improve quality of depth face images from a consumer-level depth camera. First, re-meshing and coarse-to-fine depth fusion are used to alleviate the random depth loss and noise of a depth map and project the facial surface point clouds into 3-D space. Second, a template facial landmark set is computed and frontalized, so that we can align the other faces (point clouds) onto the template using landmark-based transformation for frontalization and compute their head poses (vertical and horizontal rotation). Third, by re-projecting the fused and frontalized 3-D point cloud onto a canvas which is 2 times of the original resolution, we can obtain a super-resolved depth image by resampling. If there are still any holes on the depth image. we can fill the holes by Poisson Blending [1], with super-resolution depth image as background, pre-computed mean depth face image as foreground, and detected hole pixel map as mask. Last, to get high quality 3-D face mesh model, we just mesh and project the super-resolved depth map into 3-D space again. To synthesize depth maps from different view angles for a single 3-D face model, we rotate the model horizontally and then vertically before rendering.

To learn discriminative feature transformation by deep network, we first train our network on CASIA-WebFace [2] dataset for colour (RGB and greyscale) face images. The model for greyscale images is further fine-tuned on the merged depth dataset for transfer learning.

Database similarity standard deviation is proved to be highly correlated to reliability of similarity, and can be viewed as an estimation of image quality. A support vector classifier [3] with probability output and pairwise information as input is used to estimate the confidence score that a pair of images are come from the same subject. The SVM is trained with the following feature: group-wise colour/depth similarity, Average database colour/depth similarity standard deviation of 2 images, and estimated capture-time head pose difference. A binary label indicates whether the 2 images of each pair are from the same subject.

Our experiments show that higher accuracy can be achieved by using the proposed bi-model confidence estimation, especially under harsh illumination environment or large head pose variation.

[1] P. Perez, M. Gangnet, and A. Blake, “Poisson Image Editing” *ACM Siggraph*, 2003.

[2] D. Yi, Z. Lei, S. Liao and S. Z. Li, “Learning Face Representation from Scratch” *Computer Vision and Pattern Recognition*, 2015.

[3] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. “A training algorithm for optimal margin classifiers.” Proceedings of the fifth annual workshop on Computational learning theory. ACM, 1992.