Rotating Your Face Using Multi-task Deep Neural Network

Junho Yim[1], Heechul Jung[1], ByungIn Yoo[1,2], Changkyu Choi[2], Dusik Park[2], Junmo Kim[1]
[1]School of Electrical Engineering, KAIST, South Korea. [2]Samsung Advanced Institute of Technology.

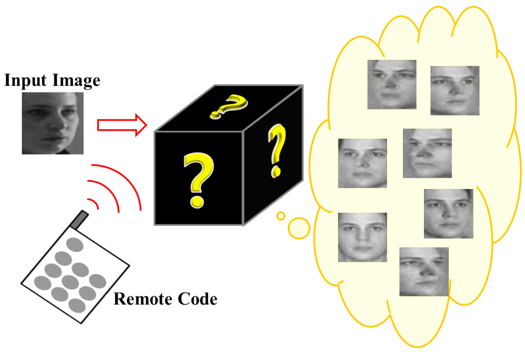

Figure 1: Conceptual diagram of our proposed model. The Input image under an arbitrary pose and illumination is transformed into another pose image. The Remote Code represents the target pose code corresponding to the output image. By interacting between the input image and the Remote Code, our model produces desired pose image.

Face recognition under viewpoint and illumination changes is a difficult problem, so many researchers have tried to solve this problem by producing the pose- and illumination- invariant feature. Zhu et al. [2] changed all arbitrary pose and illumination images to the frontal view image to use for the invariant feature. In this scheme, preserving identity while rotating pose image is a crucial issue. This paper proposes a new deep architecture based on a novel type of multitask learning, which can achieve superior performance in rotating to a target-pose face image from an arbitrary pose and illumination image while preserving identity. The target pose can be controlled by the user's intention.

The concept is illustrated in Figure 1. We train a deep neural network (DNN) that takes a face image and a binary code encoding a target pose, which we call Remote Code, and generates a face image with the same identity viewed at the target pose indicated by the Remote Code. It is as if the user has a remote control and a black-box rotator, which can rotate a given face image according to the user's Remote Code. The quality of this rotator can be measured by the degree to which the output face image accords with the desired pose and the degree to which the identity of the face is preserved. Figure 2 shows the final results of our model. From the input images under various illuminations and poses with the same identity, our model can produce almost the same images for each controlled pose under frontal illumination.

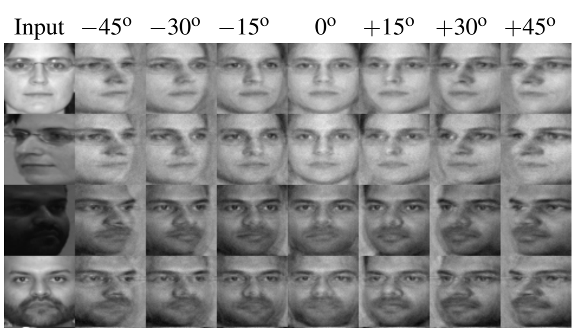

Figure 2: The first column represents the input test images of two individuals from the MultiPIE dataset. The remaining columns are the outputs from the input images with different Remote Codes. For example, the third column represents the -30^o pose images resulting from the first column images and the Remote Code that represents -30^o. The top two rows have the same identity, and the bottom two rows are the same identity under different illuminations and poses.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

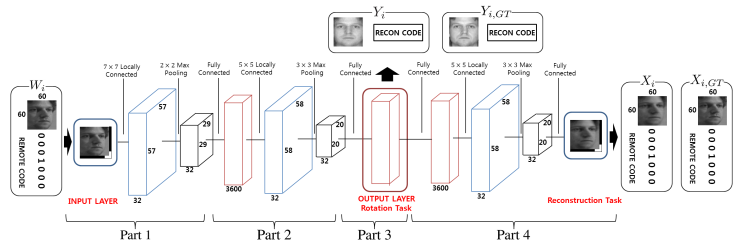

Figure 3: Complete architecture of our DNN model containing four main parts: the feature extraction part, the feature rotation part, the imaging part, and the reconstruction task part which is the auxiliary task. Each part consists of the locally connected layer, the max-pooling layer, and the fully connected layer.

To improve the identity-preserving ability of the deep neural network, we introduce an auxiliary DNN and an auxiliary task that requires that the series interconnection of the main DNN, which generates the desired pose image, and the auxiliary DNN reconstructs the original input image, i.e. the auxiliary DNN reconstructs the original input image back from the output image of the main DNN. The idea is that if the series interconnection of the main DNN and auxiliary DNN can reconstruct the original input image, the output of the main DNN should be identity-preserving and contain sufficient information about the identity of the input image. If the identity is not preserved by the main DNN, the output image of the main DNN already takes a different identity and the result of the next auxiliary DNN would deviate even further from the set of valid face images of the original identity. Figure 3 represents the final design of the network. In the third part, the red box represents the output layer where the target pose images are generated. Furthermore, the novel element of the additional task part is attached after the third part.

To demonstrate how our model maintains the identity of input images, we take the face recognition task by using the MultiPIE dataset [1]. We used 200 subjects (ID 001 $\sim$ 200) under 9 poses (-60^o to $+60^o$) with 20 illuminations for training. For the testing, we used remaining 137 subjects under 9 poses with 20 illuminations, $137 \times 9 \times 20$ images in total. We compared our results with the state-of-the-art results [3] and Zhu et al. [2]. The results of recognition rates for different poses are shown in Table 1. As with human perception, our model found it difficult to imagine the face identity from the greatly rotated images, -60^o and $+60^o$ cases. However, Table 1 shows that proposed model outperformed the state-of-the-art for most poses.

	-60^o	-45^o	-30^o	-15^o	0^o	$+15^o$	$+30^o$	$+45^o$	$+60^o$	*Avg*
FIP[2]	49.3	66.1	78.9	91.4	94.3	90.0	82.5	62.0	42.5	72.9
RL[2]	44.6	63.6	77.5	90.5	94.3	89.8	80.0	59.5	38.9	70.8
Z.Zhu[3]	60.2	75.2	83.4	93.3	95.7	92.2	83.9	70.6	60.0	79.3
Proposed	**63.2**	**80.4**	**88.1**	**94.5**	**99.5**	**95.4**	**88.9**	**79.4**	**60.6**	**83.3**

Table 1: Recognition rates (%) for the various poses, compared with proposed model and state-of-the-art model. Best results are written in bold.

[1] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[2] Zhenyao Zhu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning identity-preserving face space. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 113–120. IEEE, 2013.

[3] Zhenyao Zhu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Multi-view perceptron: a deep model for learning face identity and view representations. In *Advances in Neural Information Processing Systems*, pages 217–225, 2014.