Supplementary Material

Multi-Dim: A Multi-Dimensional Face Database Towards the Application of 3D Technology in Real-World Scenarios

Feng Liu, Jun Hu, Jianwei Sun, Yang Wang and Qijun Zhao
College of Computer Science, Sichuan University, Chengdu, China

To enable comprehensive evaluation of 3D-based unconstrained 2D face recognition and related techniques, we provide ground truth 3D landmarks, 2D landmarks, and face bounding boxes for the face images / video clips in Multi-Dim. In this supplementary material, we provide the annotation details.

1. 3D Landmarks

We manually picked 51 vertices on each 3D face model as 3D landmarks (see Fig. 1). These landmarks could be used as ground truth to assess 3D/2D face alignment [1, 2, 3, 4, 5] and 3D face segmentation algorithms.

Figure 1. The raw and cropped 3D face model with 51 landmarks.

2. 2D Landmarks

We also manually marked the landmarks on still images and video clips. As shown in Fig. 3, for the face images whose yaw angles are within 30°, 68 landmarks as defined in Multi-PIE [6] are annotated, while only the visible 39 landmarks are annotated for the face images whose yaw angles are beyond 30°.

For the face images cropped from the surveillance video clips, we first applied the automated landmark localization method in [7] to detect five landmarks on them, and then manually refined the landmarks that were not accurately detected. See Fig. 2.

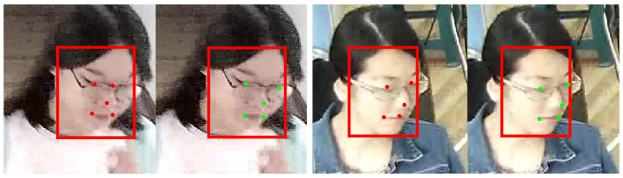

Figure 2. The annotated face bounding boxes and five 2D landmarks on the face images cropped from the surveillance video clips. Red points denote the inaccurate landmarks detected by automatic method, and green points are the corresponding manually refined landmarks.

3. Face Bounding Boxes

For the surveillance video clips, we also applied the method in [7] first to automatically detect the face bounding boxes, and then manually refined them if necessary, resulting in $34,512$ valid face bounding boxes of 72,390 frames in total. See Fig. 2.

Figure 4 shows the distribution of size of the obtained face bounding boxes, from which we can see most are between 50 and 150 pixels. The size of a face bounding box is defined as the maximum of its height and width.

References

[1] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, pages 532–539, 2013.

[2] Amin Jourabloo and Xiaoming Liu. Pose-invariant 3D face alignment. In *ICCV*, pages 3694–3702, 2015.

[3] Xuan Li, Yidan Xu, Qi Lv, and Yong Dou. Affine-transformation parameters regression for face alignment. *IEEE Signal Process. Lett.*, 23(1):55–59, 2016.

[4] László A Jeni, Jeffrey F Cohn, and Takeo Kanade. Dense 3D face alignment from 2D videos in real-time. In *FG*, volume 1, pages 1–8, 2015.

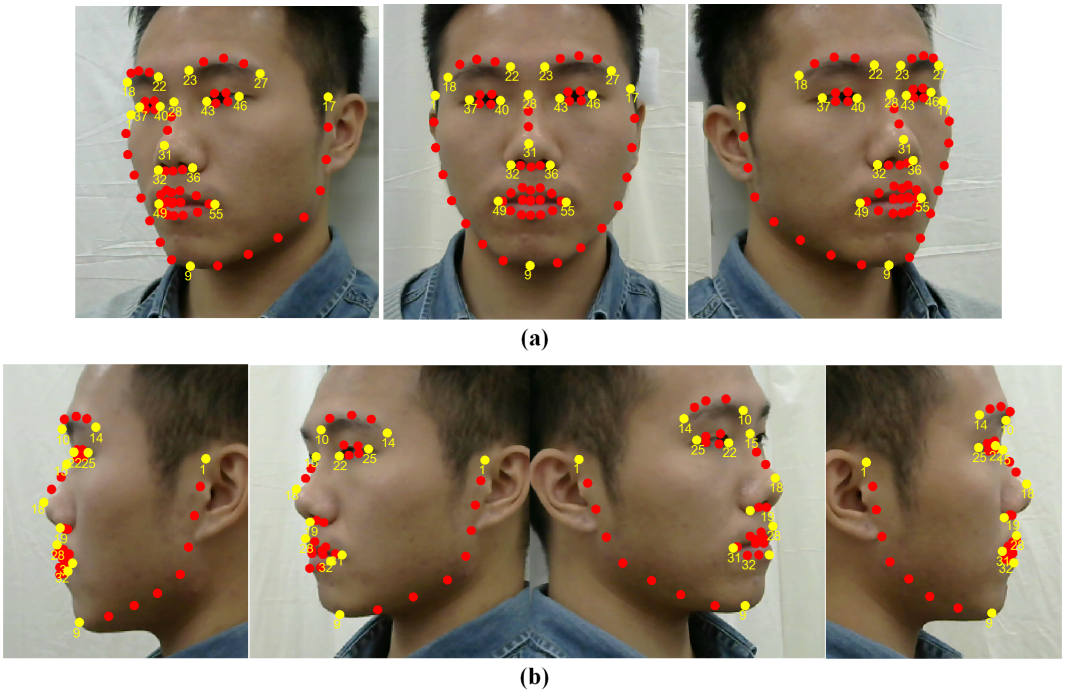

Figure 3. Manually labeled landmarks on 2D still images. (a) 68 landmarks on face images of 0° or $\pm 30^\circ$ yaw rotation and (b) 39 landmarks on face images of $\pm 60^\circ$ or $\pm 90^\circ$ yaw rotation.

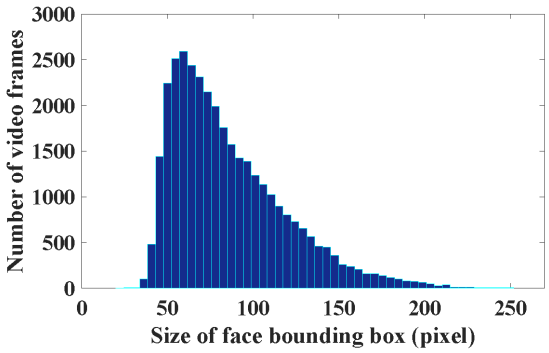

Figure 4. Statistics of the size of the face bounding boxes in the surveillance video clips.

[5] Sergey Tulyakov and Nicu Sebe. Regressing a 3D face shape from a single image. In *ICCV*, pages 3748–3755, 2015.

[6] R Gross, I Matthews, J Cohn, T Kanade, and S Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010.

[7] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.*, 23(10):1499–1503, 2016.