Orientation-conditioned Facial Texture Mapping for Video-based Facial Remote Photoplethysmography Estimation

Supplementary Material

7. Supplementary Material

7.1. Data Visualization

We provide select video frames approved for publication from the PURE [4] dataset after different video processing steps but excluding the normalized frame-difference, pixel outlier clipping and standardization steps (F_D) for visualization and comparison.

Figure 6. Frame of *Subject 1* in scenario *Medium Rotation* from PURE [4] with static cropping (×1.5-scale Box) applied. Subsequent frames will use the same bounding box, hence the in-plane position of the face will vary due to subject motion.

Figure 7. Frame of *Subject 1* in scenario *Medium Rotation* from PURE [4] with static cropping (×1.5-scale Box) and facial segmentation applied. Subsequent frames will use the same bounding box, hence the in-plane position of the face will vary due to subject motion.

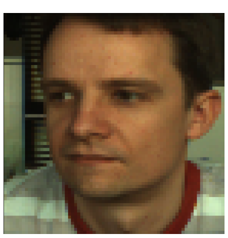

Figure 8. Frame of *Subject 1* in scenario *Medium Rotation* from PURE [4] with dynamic cropping (×1.5-scale Box) and square padding applied. However, subsequent frames will remain centered on the square padded and scaled facial region allowing for larger in-plane subject motion within a video sequence.

Figure 9. Frame of *Subject 1* in scenario *Medium Rotation* from PURE [4] with UV transformation and masking ($\Theta \geq 45°$) applied - the UV transformation process inherently dynamically localizes and segments the facial region. Subsequent frames will have the same structure with varying texture.

7.2. Intra-dataset Testing

In Table 5 we report the full set of performance metrics referenced in Section 4.3 using the evaluation pipeline and metric implementations provided in [15] for intra-dataset testing on the PURE [33] dataset using subject-independent cross-validation. We obtain these results using the protocol described in Section 4.4. We report both the results including and excluding samples from *Subject 7 - Talking* (S7-T) to provide insight into the evaluation variability. We denote the sequence first-order normalized frame difference, pixel outlier clipping, and standardization operations as F_D for brevity. We also denote the UV transformation operation as T_{UV}.

7.3. Cross-dataset Testing

In Table 6 we report the full set of performance metrics referenced in Section 4.3 using the evaluation pipeline and metric implementations provided in [15]. We obtained these results using the protocol described in Section 4.5, we perform cross-dataset testing on the MMPD [34] dataset using PhysNet models trained on the PURE [4] dataset. We report additional ablations for the operations applied after T_{UV} to demonstrate the impact of the sequence of operations, and provide additional internally consistent comparisons. We denote the sequence first-order normalized frame difference, pixel outlier clipping, and standardization operations as F_D for brevity. We also denote the UV transformation operation as T_{UV}.

Video Processing Pipeline	MAE ± SE (BPM)	RMSE ± SE (BPM)	r ± SE	SNR ± SE (dB)
F_D + Crop$_{Static}$ (×1.5-Box) + Resize	0.492 ± 0.172	1.408 ± 0.946	0.998 ± 0.008	10.721 ± 1.044
Crop$_{Static}$ (×1.5-Box) + Resize + F_D (**PhysNet-XY**)	1.318 ± 0.979	7.632 ± 56.531	0.945 ± 0.043	11.061 ± 1.025
Crop$_{Static}$ (×1.5-Box) + Resize + F_D (Excl. S7-T)	0.341 ± 0.138	1.108 ± 0.727	0.999 ± 0.007	11.457 ± 0.963
T_{UV} + F_D + Resize	2.734 ± 1.510	11.918 ± 98.699	0.862 ± 0.067	11.546 ± 1.135
T_{UV} + F_D + Resize (Excl. S7-T)	0.594 ± 0.217	1.739 ± 1.435	0.996 ± 0.011	12.228 ± 1.063
T_{UV} + Mask ($\Theta_{UV} \geq 90°$) + F_D + Resize	1.393 ± 0.938	7.338 ± 51.500	0.949 ± 0.042	12.011 ± 1.140
T_{UV} + Mask ($\Theta_{UV} \geq 90°$) + F_D + Resize (Excl. S7-T)	0.462 ± 0.171	1.381 ± 0.959	0.998 ± 0.008	12.470 ± 1.063
T_{UV} + Mask ($\Theta_{UV} \geq 60°$) + F_D + Resize	1.676 ± 1.316	10.243 ± 102.807	0.899 ± 0.058	12.211 ± 1.084
T_{UV} + Mask ($\Theta_{UV} \geq 60°$) + F_D + Resize (Excl. S7-T)	0.356 ± 0.139	1.114 ± 0.727	0.999 ± 0.007	12.617 ± 1.023
T_{UV} + Mask ($\Theta_{UV} \geq 45°$) + F_D + Resize (**PhysNet-XY**)	1.639 ± 1.141	8.919 ± 76.940	0.924 ± 0.051	11.842 ± 1.106
T_{UV} + Mask ($\Theta_{UV} \geq 45°$) + F_D + Resize (Excl. S7-T)	0.500 ± 0.171	1.397 ± 0.958	0.998 ± 0.008	12.159 ± 1.079
T_{UV} + Mask ($\Theta_{UV} \geq 30°$) + F_D + Resize	1.594 ± 1.113	8.693 ± 72.994	0.928 ± 0.049	11.486 ± 1.108
T_{UV} + Mask ($\Theta_{UV} \geq 30°$) + F_D + Resize (Excl. S7-T)	0.485 ± 0.172	1.397 ± 0.958	0.998 ± 0.008	11.817 ± 1.077

Table 5. Intra-dataset subject-independent performance of PhysNet across different video processing pipelines on the PURE [33] dataset using averaged results across all folds from subject-independent cross-validation training on the PURE [4] dataset.

Video Processing Pipeline	MAE ± SE (BPM)	RMSE ± SE (BPM)	r ± SE	SNR ± SE (dB)
F_D + Crop$_{Static}$ (×1.5-Box) + Resize	17.492 ± 0.307	24.827 ± 16.908	0.047 ± 0.017	-6.225 ± 0.074
Crop$_{Static}$ (×1.5-Box) + Resize + F_D (**PhysNet-XY**)	14.905 ± 0.295	22.542 ± 15.837	0.155 ± 0.017	-6.882 ± 0.080
Crop$_{Static}$ (×1.5-Box) + Segment + Resize + F_D	15.237 ± 0.312	23.524 ± 17.217	0.120 ± 0.017	-6.053 ± 0.088
Crop$_{Dynamic}$ (×1.5-Box) + Pad$_{Square}$ + Resize + F_D	17.988 ± 0.307	25.183 ± 16.488	0.033 ± 0.017	-6.263 ± 0.072
Crop$_{Dynamic}$ (×1.5-Box) + Pad$_{Square}$ + Segment + Resize + F_D	14.683 ± 0.298	22.563 ± 15.526	0.138 ± 0.017	-6.553 ± 0.082
T_{UV} + Resize + F_D	13.168 ± 0.285	21.014 ± 14.394	0.227 ± 0.017	-6.606 ± 0.084
T_{UV} + Mask ($\Theta_{UV} \geq 90°$) + Resize + F_D	13.547 ± 0.288	21.391 ± 14.610	0.210 ± 0.017	-6.644 ± 0.085
T_{UV} + Mask ($\Theta_{UV} \geq 60°$) + Resize + F_D	12.949 ± 0.284	20.840 ± 14.129	0.243 ± 0.017	-6.305 ± 0.087
T_{UV} + Mask ($\Theta_{UV} \geq 45°$) + Resize + F_D	15.222 ± 0.302	23.072 ± 15.474	0.156 ± 0.017	-6.105 ± 0.083
T_{UV} + Mask ($\Theta_{UV} \geq 30°$) + Resize + F_D	15.771 ± 0.298	23.285 ± 15.670	0.133 ± 0.017	-6.643 ± 0.082
T_{UV} + F_D + Resize	12.687 ± 0.280	20.454 ± 13.843	0.248 ± 0.017	-6.679 ± 0.088
T_{UV} + Mask ($\Theta_{UV} \geq 90°$) + F_D + Resize	13.038 ± 0.285	20.900 ± 14.249	0.216 ± 0.017	-6.473 ± 0.086
T_{UV} + Mask ($\Theta_{UV} \geq 60°$) + F_D + Resize	12.890 ± 0.280	20.629 ± 13.794	0.256 ± 0.017	-6.284 ± 0.088
T_{UV} + Mask ($\Theta_{UV} \geq 45°$) + F_D + Resize (**PhysNet-UV**)	12.187 ± 0.273	19.849 ± 13.102	0.294 ± 0.017	-6.265 ± 0.092
T_{UV} + Mask ($\Theta_{UV} \geq 30°$) + F_D + Resize	13.300 ± 0.279	20.834 ± 13.611	0.277 ± 0.017	-6.496 ± 0.087

Table 6. Cross-dataset performance of PhysNet across different video processing pipelines on the MMPD [33] dataset using averaged results across all folds from subject-independent cross-validation training on the PURE [33] dataset.