Spherical View Synthesis for Self-Supervised 360° Depth Estimation - Supplementary Material

Nikolaos Zioulis[1,2], Antonis Karakottas[1], Dimitrios Zarpalas[1], Federico Alvarez[2], and Petros Daras[1]

[1]Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Greece
[2]Signals, Systems and Radiocommunications Department (SSRD), Universidad Politécnica de Madrid (UPM), Madrid, Spain

1. Introduction

In this supplementary material, we complement our original manuscript with additional qualitative results.

Figure 1 shows qualitative results for the fully supervised (SV) model, the horizontal stereo (LR) model, the vertical stereo (UD) model and the best performing trinocular (TC6) model, along with the corresponding color image used as input to the network, and the ground truth depth map. An interesting observation is that while SV offers superior performance, it nonetheless produces predictions with high frequency noise. On the other hand, the inferior quantitatively, unsupervised methods produce smoother predictions but with a bias towards farther predictions. This is the reason for the bigger supervised-unsupervised deviation in quantitative performance of RMSE against that of RMSLE, which penalizes larger depth prediction errors and gives higher precedence to smaller depth prediction errors. Additionally, we observe that LR predictions are smoother, removing finer structures, while the UD and TC6 predictions offer more detailed predictions.

Next, we offer additional qualitative results for the unsupervised methods, including the all TC models to showcase the gradual progression from UD to LR as the ratio λ_{ratio} balancing their loss contributions increases. Figure 2 presents these results and showcases that the TC variants preserve details similar to UD, while LR only suffers from blurry predictions. Furthermore, there is no easily discernible visible qualtitative differences between the TC variants and UD. In addition, the vertical singularity effect is more prevalent in LR, and it gradually fades out as the ratio (and consequently, UD's effect) increases. Interestingly, the spherical attention mask of both UD and LR includes the vertical distortion attenuation factor $|\sin(\theta)|$. However, only LR is supervised by horizontal disparity which is unstable at the north and south poles, as any disparity value is practically sampling from the same pixels.

Finally, Figure 3 presents the horizontal stereo (right) re-constructions for the samples presented in Figure 2. Results are offered for the supervised and the subset of unsupervised methods presented in Figure 1 as well, with the addition of the TC4 variant. Overall, we find that LR produces heavier singularity artifacts that manifest as empty (true black) pixels. In addition, the high frequency noise existing in the SV predictions produces more noisy reconstructions compared to the unsupervised models that, nonetheless, have been trained with a reconstruction loss.
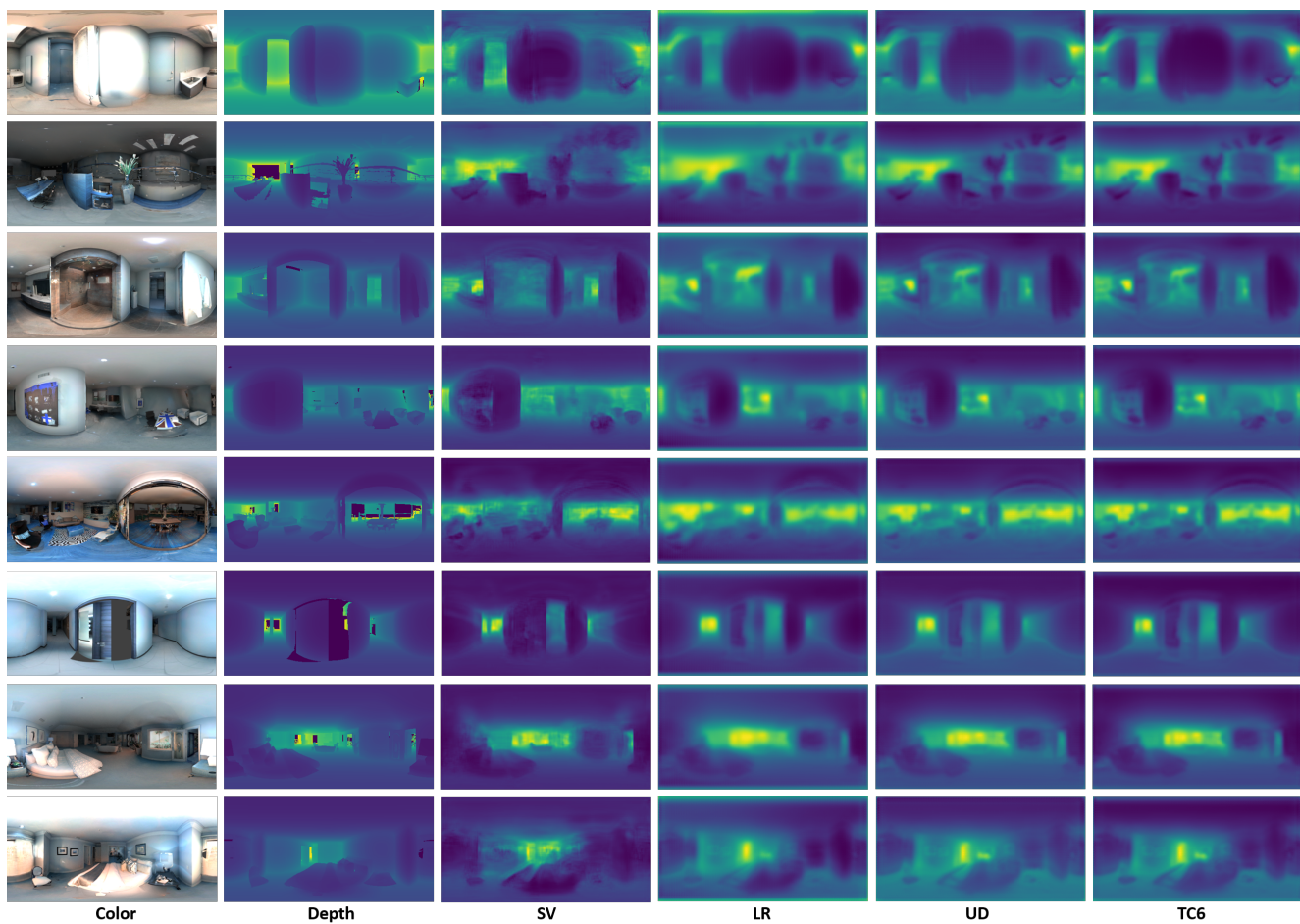
| Color | Depth | SV | LR | UD | TC6 |

Figure 1. Additional qualitative results for the supervised (SV) and all the unsupervised (vertical stereo - UD, horizontal stereo - LR, and the best performing trinocular - TC6) models, along with the input color image and the ground truth depth map. The SV predictions contain slight high frequency noise, while the unsupervised predictions (and especially LR) are more blurry. Moreover, as UD increasingly influences supervision, more details are added to the unsupervised models.
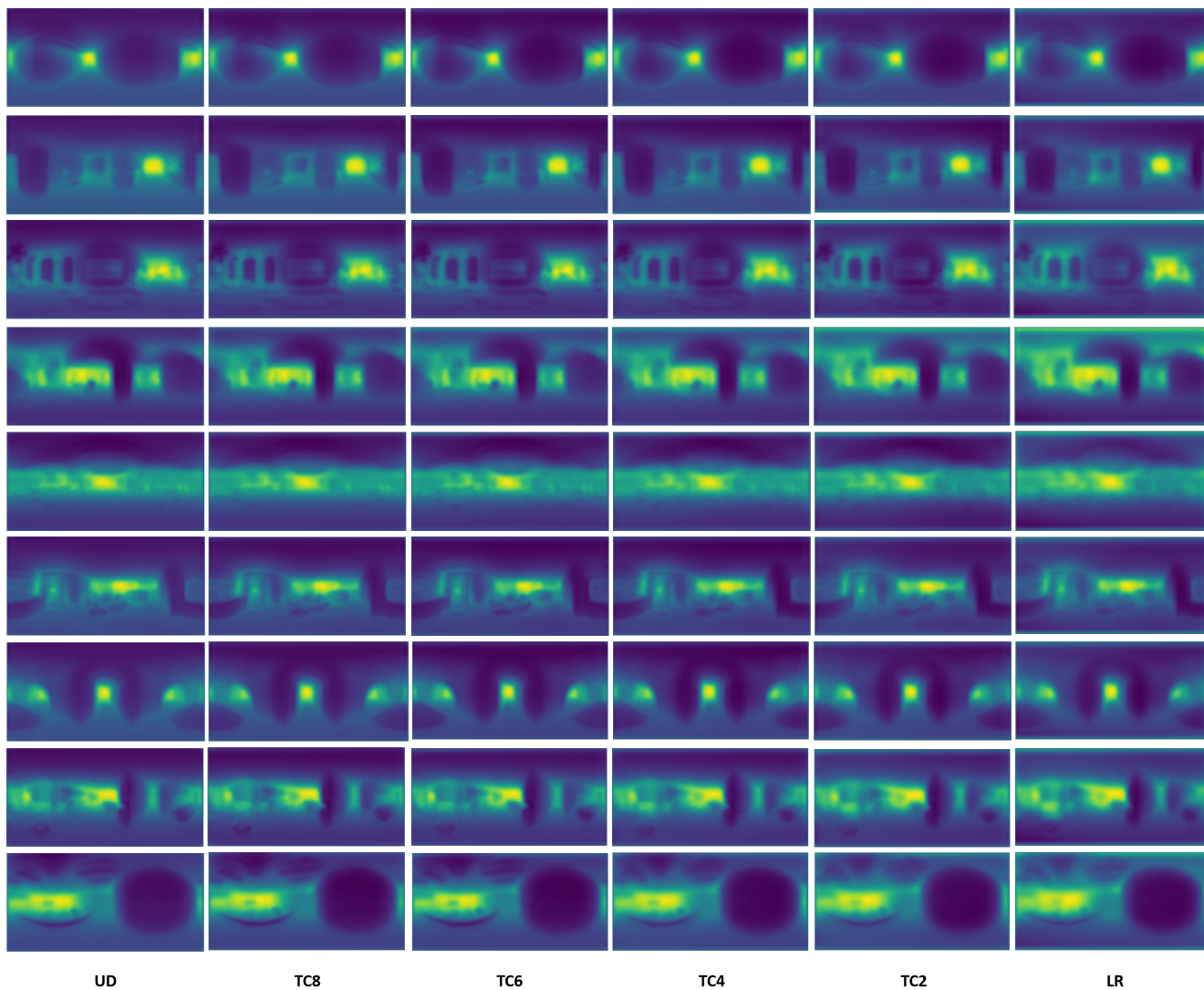
Figure 2. Qualitative results for trinocular interpolation of the UD/LR balancing factor λ_{ratio} from 1.0 (UD) to 0.0 (LR) with a step of 0.2.
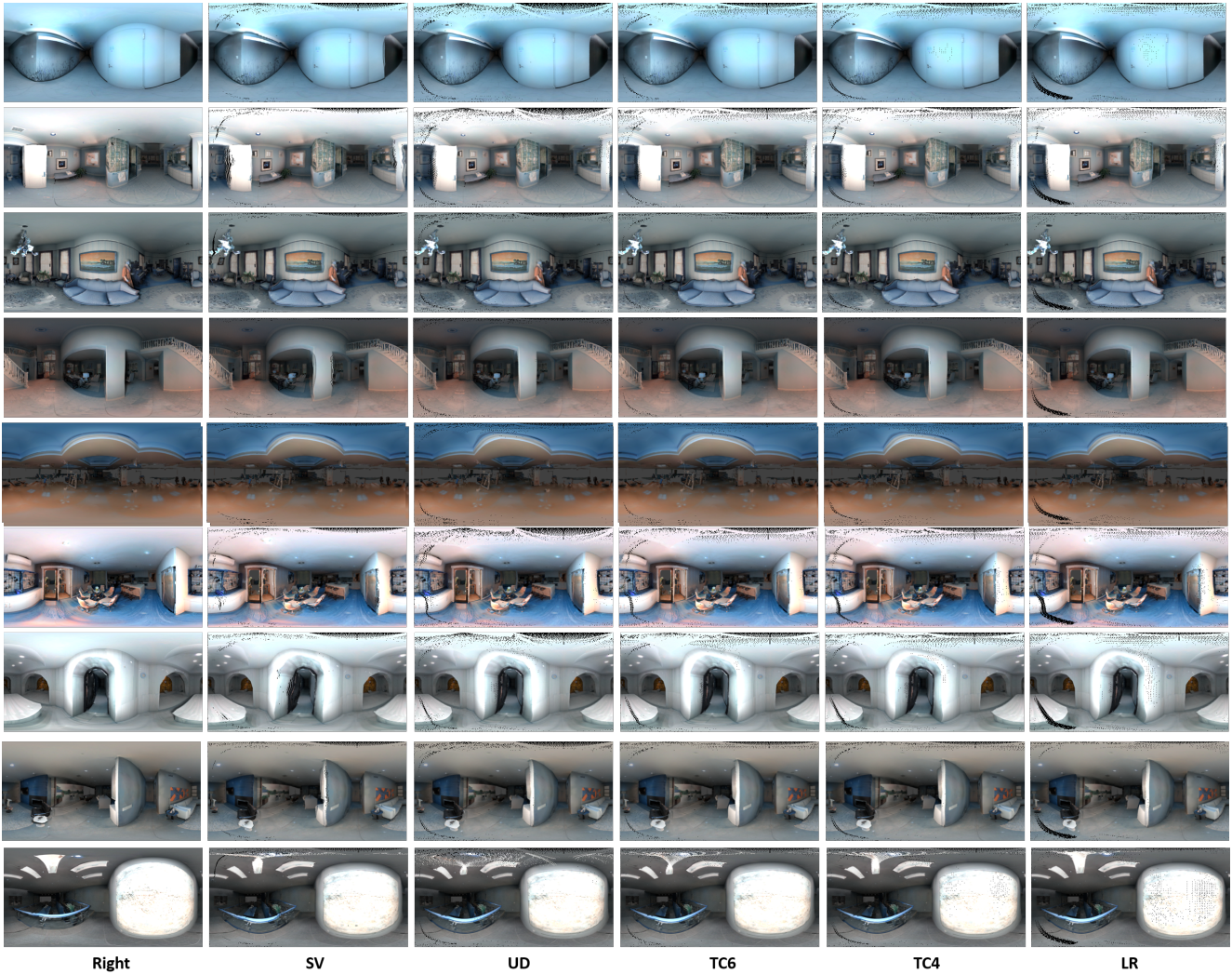
Figure 3. Qualitative horizontal (right stereo image) reconstruction results. From left to right: right ground truth image, the reconstructions of the supervised, vertical stereo, trinocular with $\lambda_{ratio} = 6$, trinocular with $\lambda ratio = 4$ and horizontal stereo models. Areas of focus from top to bottom: first row – a low detail image that is largely accurately reconstructed with a small deviation for SV at the right corridor, second row – left door and right entryway, third row – chandelier, fourth row – center wall, firth row – center ceiling, sixth row – wall texture and center chairs, seventh row – center doorway, eighth row – center right wall, ninth row – ceiling lights.