Virtually Enriched NYU Depth V2 Dataset for Monocular Depth Estimation: Do We Need Artificial Augmentation?

Supplementary Material

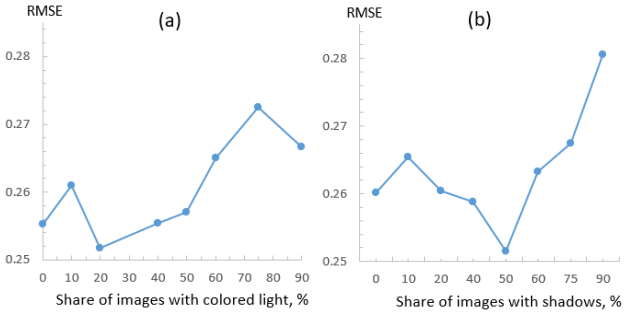

Figure 8. RMSE dependence on the share of images with colored lights (a) or with shadows (b).

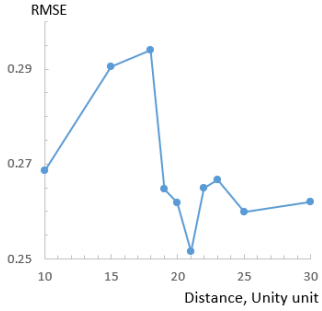

Figure 9. RMSE dependence on the distance between virtual camera and background NYU-v2 image.

Here we provide additional details that allow for a better understanding of the processes described in the main paper.

8. Selecting Image Generation Parameters

While most of the parameters of the image virtualization process are automatically randomly selected by our Unity program over the full range of their values, some of them significantly affect the accuracy of models trained for monocular depth estimation and have to be manually chosen. These parameters include the percentage of images generated with colored light or shadows from virtual objects as illustrated in Fig. 8.

Another parameter that can be seen in Fig. 10 is the distance between virtual camera – a small white object located near the left edge of the screenshots – and a background image presented on the rectangular plane on the right side of the screenshots. This distance is used in the initial stage of image virtualization which is described in detail in Sec. 9. Based of the results of preliminary experiments with the VPD model presented in Fig. 8 – 9, we set the values of the mentioned distance to 21 Unity units, and the fraction of generated images with colored lights and with shadows to 20% and 50%, respectively, and used these values in all subsequent experiments described in the article.

9. Initial Stage of Image Virtualization

To obtain a correct color at the edges of the virtual objects, they are rendered in the Unity 3D platform against a background of images selected from the NYU-v2 database. Fig. 11 – 16 contain screenshots of the virtualization process of images that meet the criteria described in our paper (Sec. 3, *Augmentation and Culling*), and therefore selected for subsequent virtualization steps. The remaining part of the screenshots represents this process in the Unity 3D space: the virtual camera on the left, virtual 3D objects in front of it, and the training image from the original NYU-v2 dataset in the background rectangular plane, which moves synchronously with the camera. On the right side of the screenshots, one can see the generated image.

10. Incorporating Virtual Objects into RGB-D Image Pairs

It is important to emphasize that we described above only the first step of image virtualization, where we set NYU-v2 data as the background of the image to obtain a smooth color change at the edges of the virtual objects. Afterwards, virtual objects from these images and incorporated into the original NYU-v2 RGB images, taking into account the depth maps and following the procedure described in the main paper in Sec. 3. After virtualization, the color distribution of new images is normalized according to the mean and standard deviation of the RGB color brightness distribution in the original NYU-v2 training set (Tab. 3).

Table 3. Color normalization parameters.

	Red	Green	Blue
Mean	123.675	116.28	103.53
σ	58.395	57.12	57.375

Examples of RGB-D NYU-v2 image enrichment with virtual objects are presented in the Fig. 17 – 24.

Figure 10. Screenshots of the Unity 3D at the initial stage of image virtualization: the distance between the virtual camera (left) and the NYU-v2 image plane (right) determines how many virtual objects can be effectively placed between the camera and the image plane.

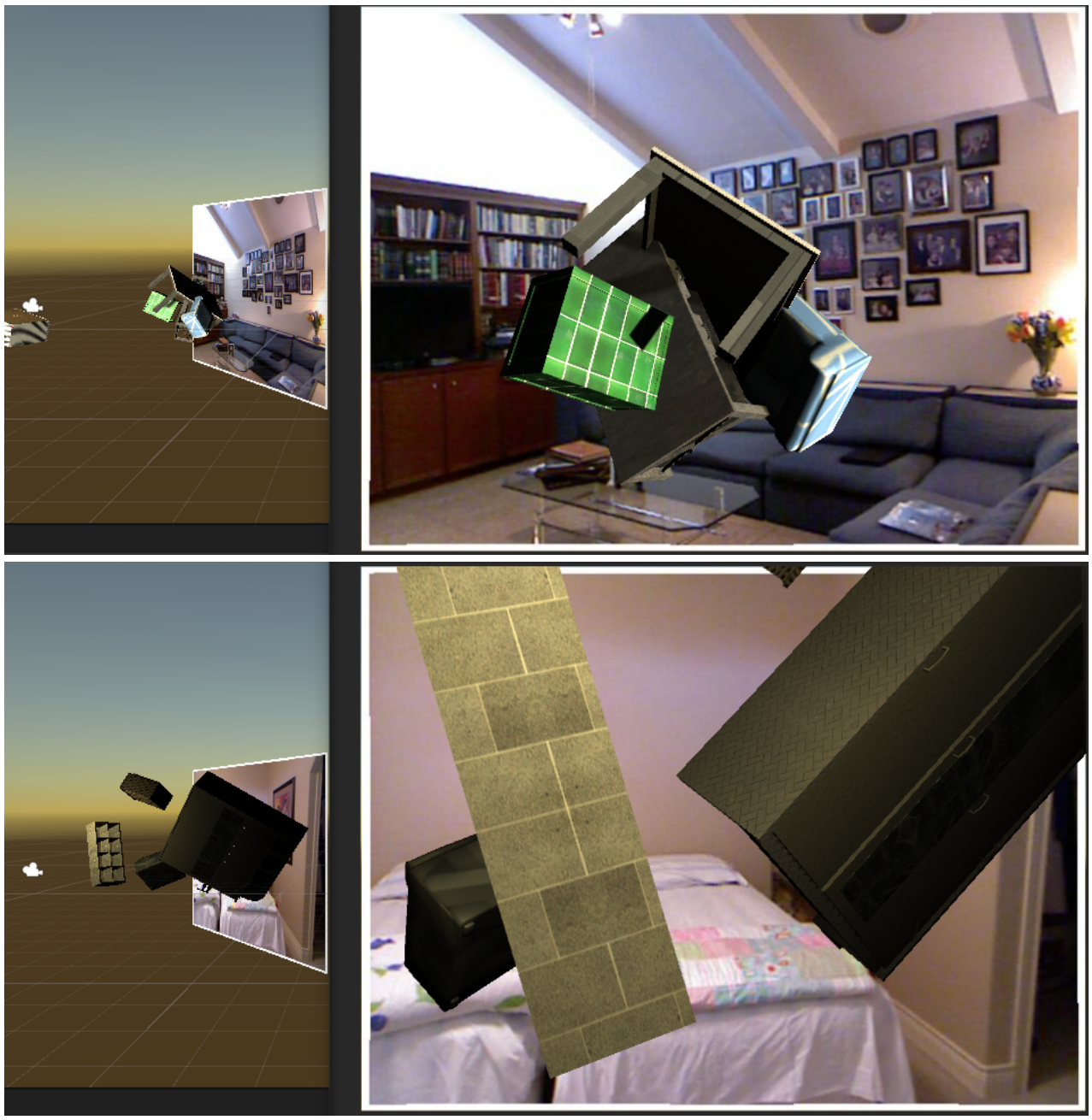

Figure 11. The initial stage of image virtualization in Unity 3D.

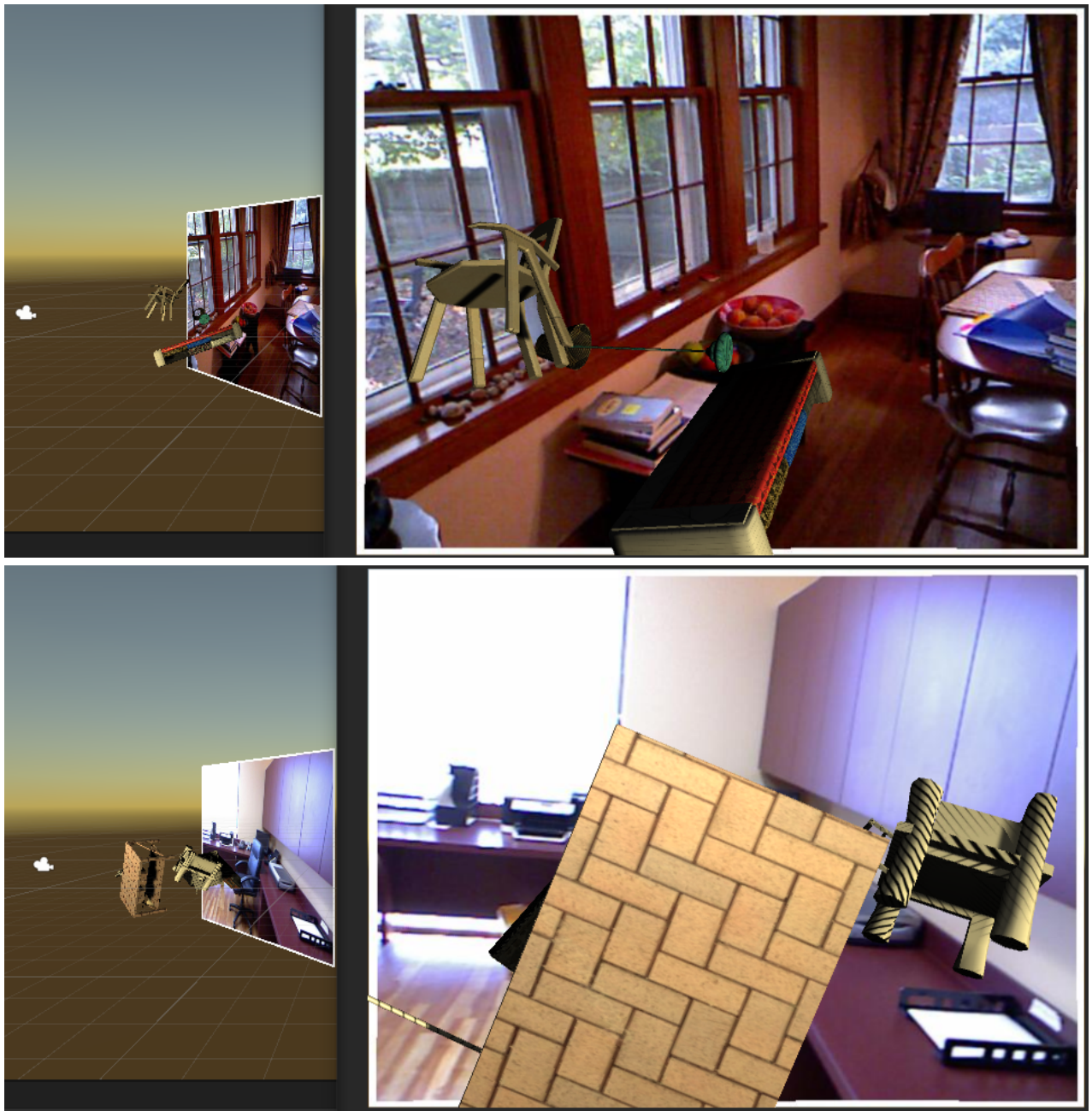

Figure 12. The initial stage of image virtualization in Unity 3D.

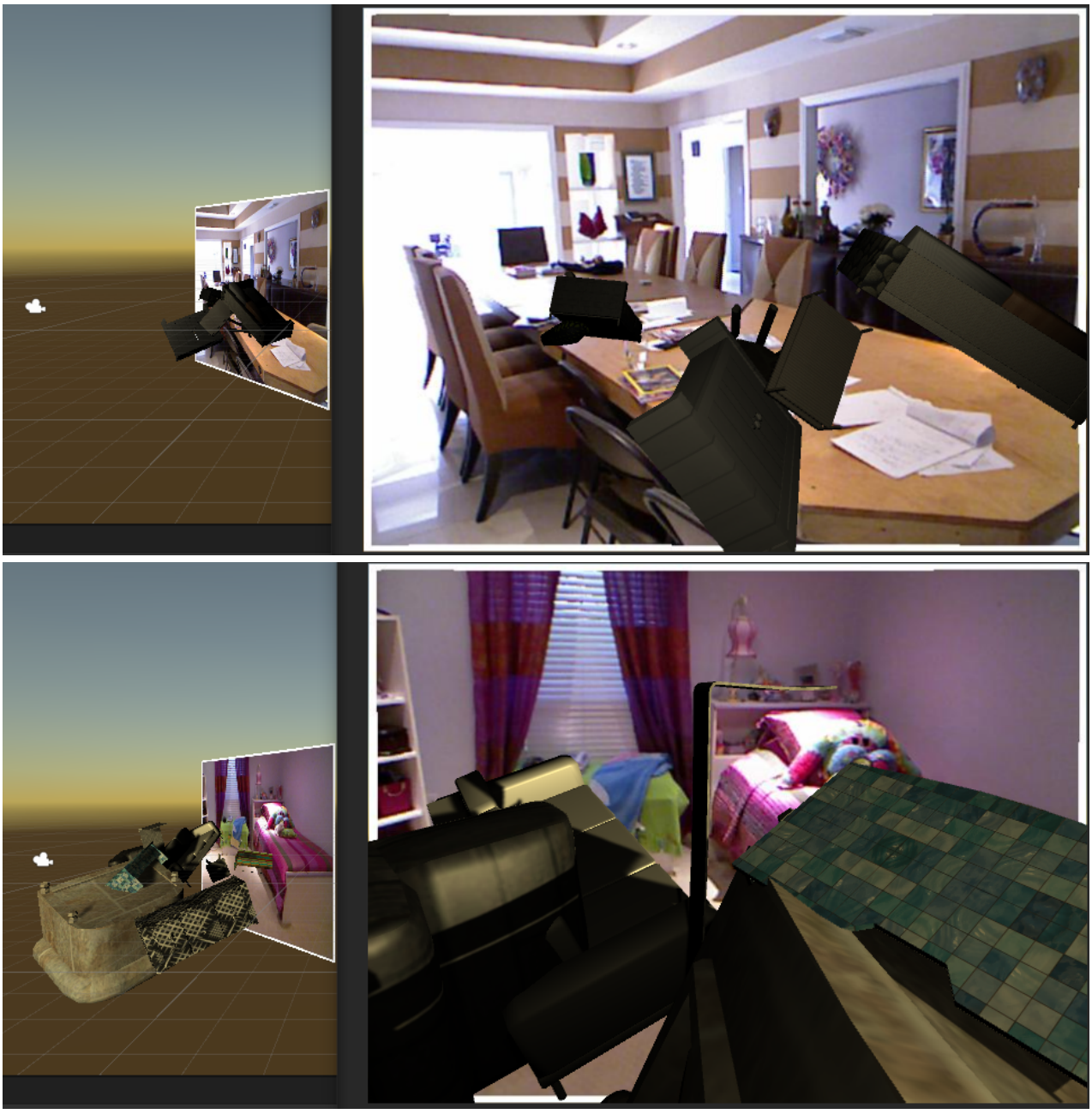

Figure 13. The initial stage of image virtualization in Unity 3D.

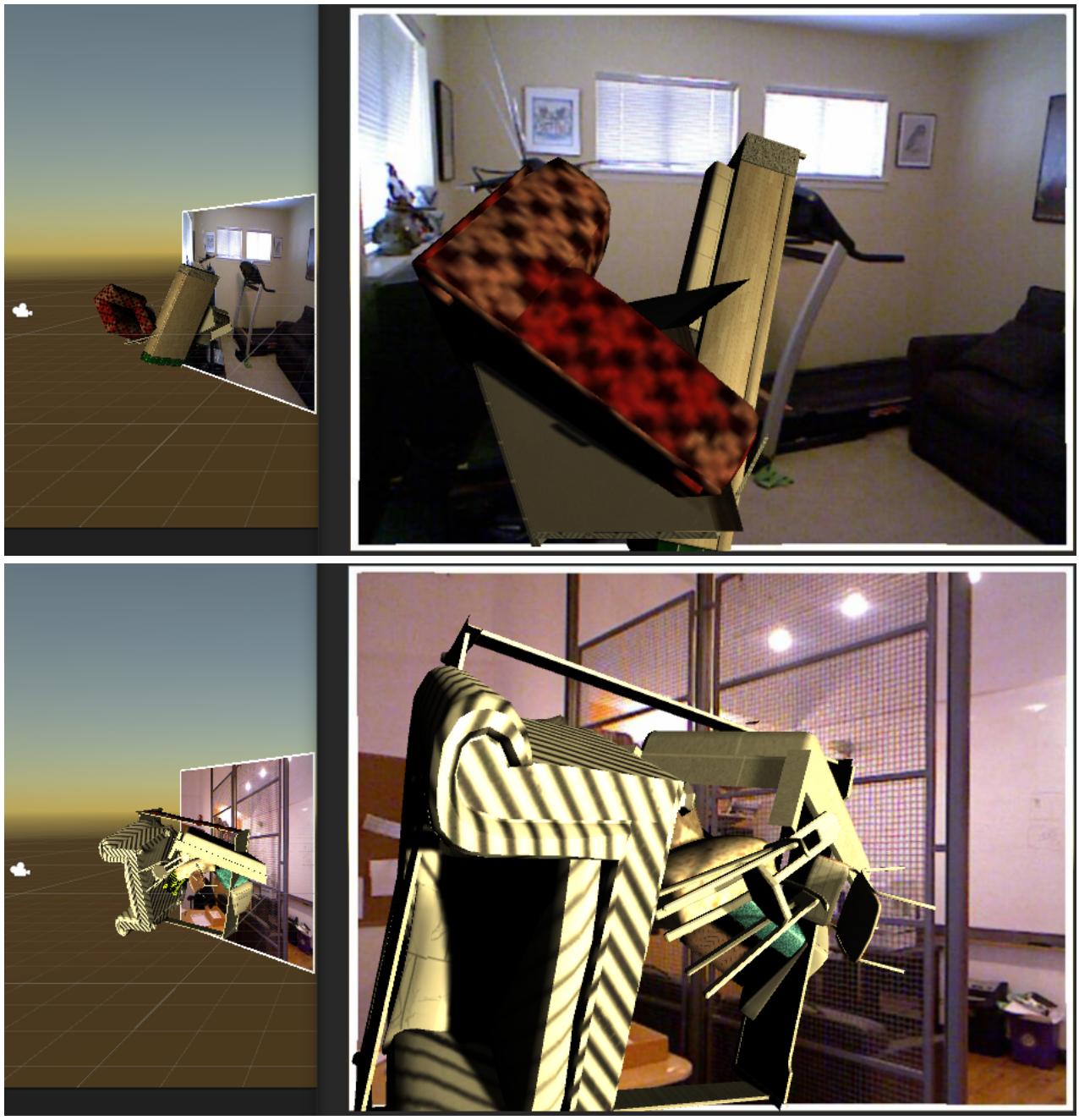

Figure 14. The initial stage of image virtualization in Unity 3D.

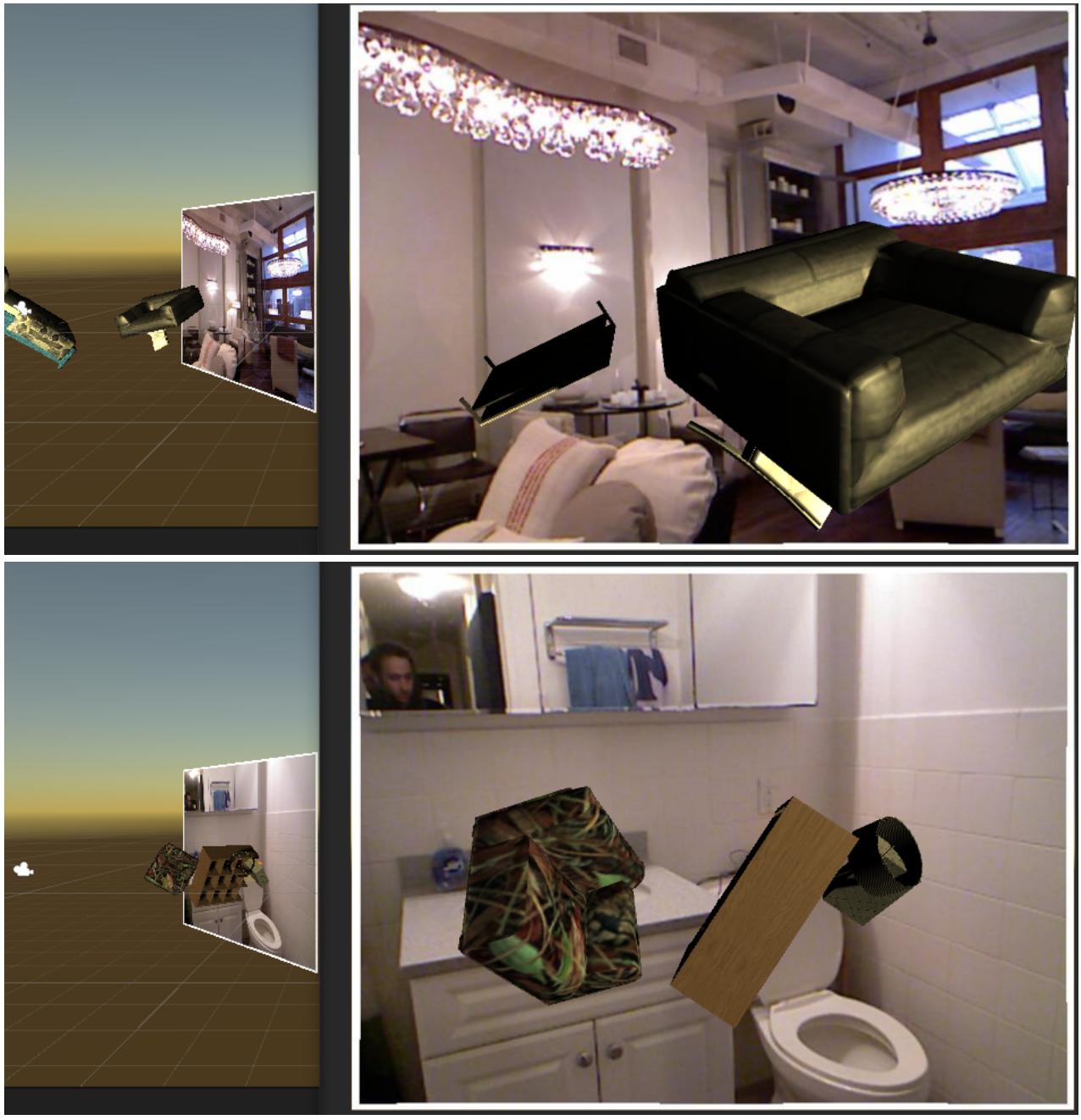

Figure 15. The initial stage of image virtualization in Unity 3D.

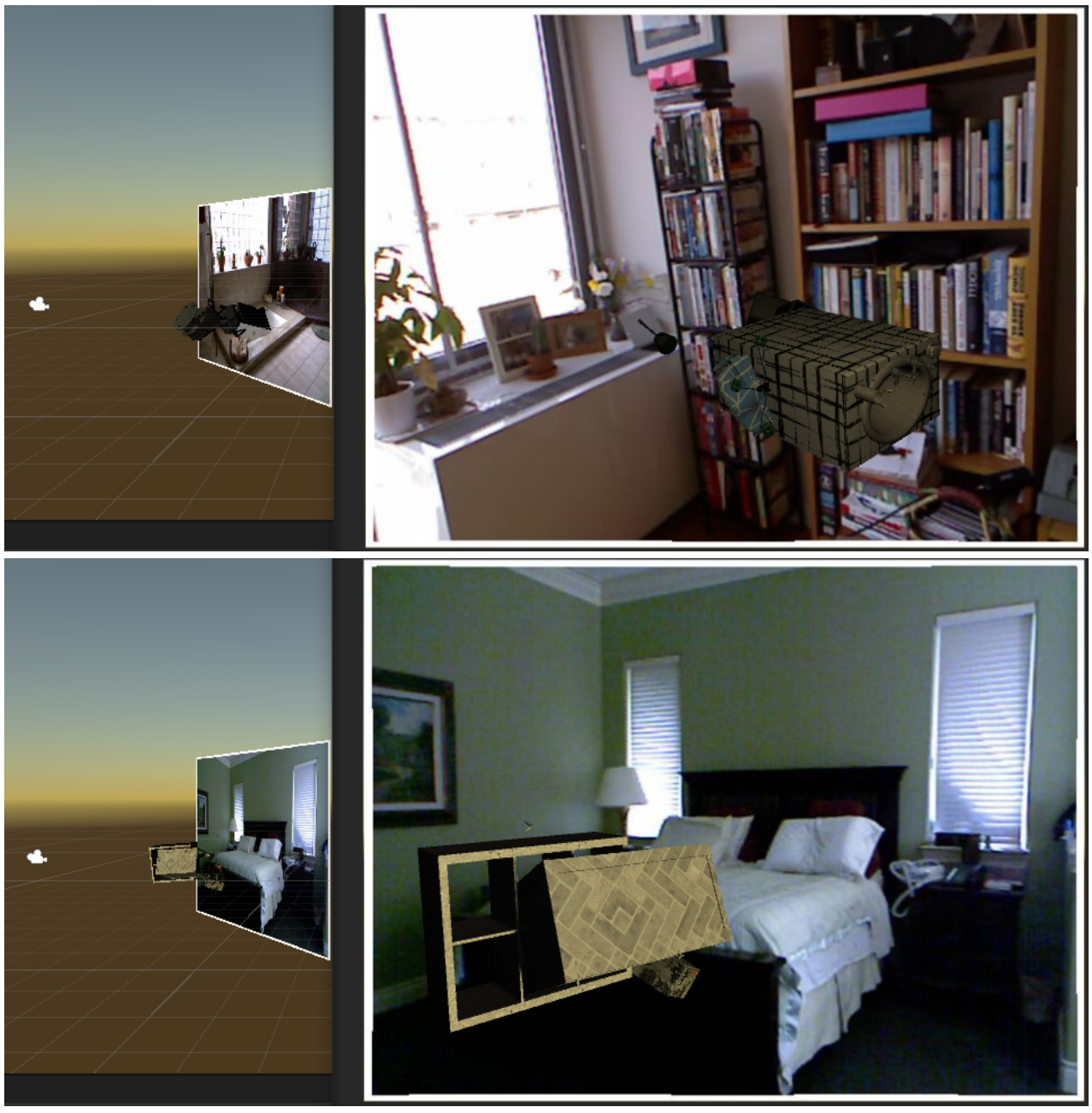

Figure 16. The initial stage of image virtualization in Unity 3D.

Figure 17. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).

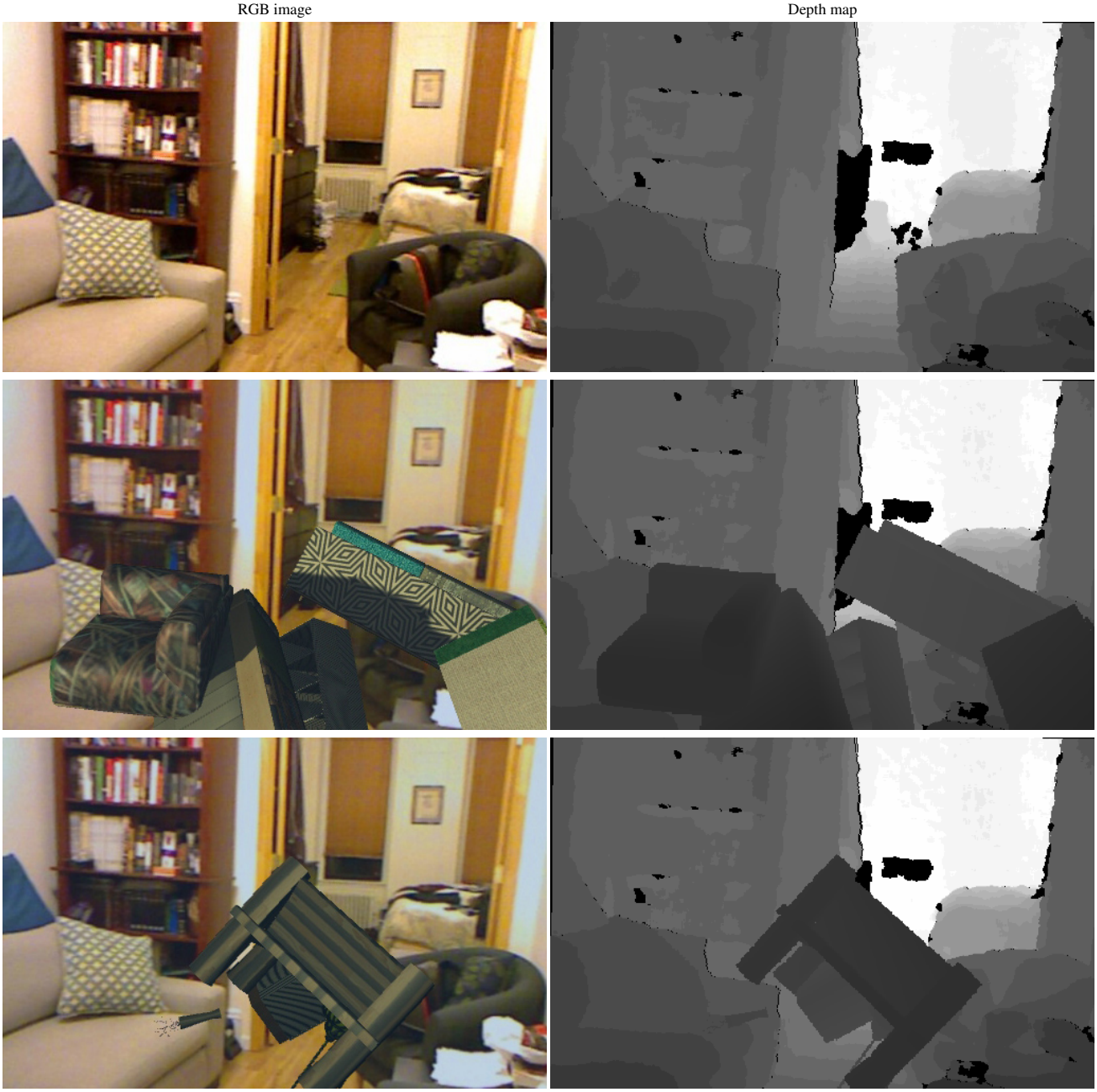

Figure 18. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).

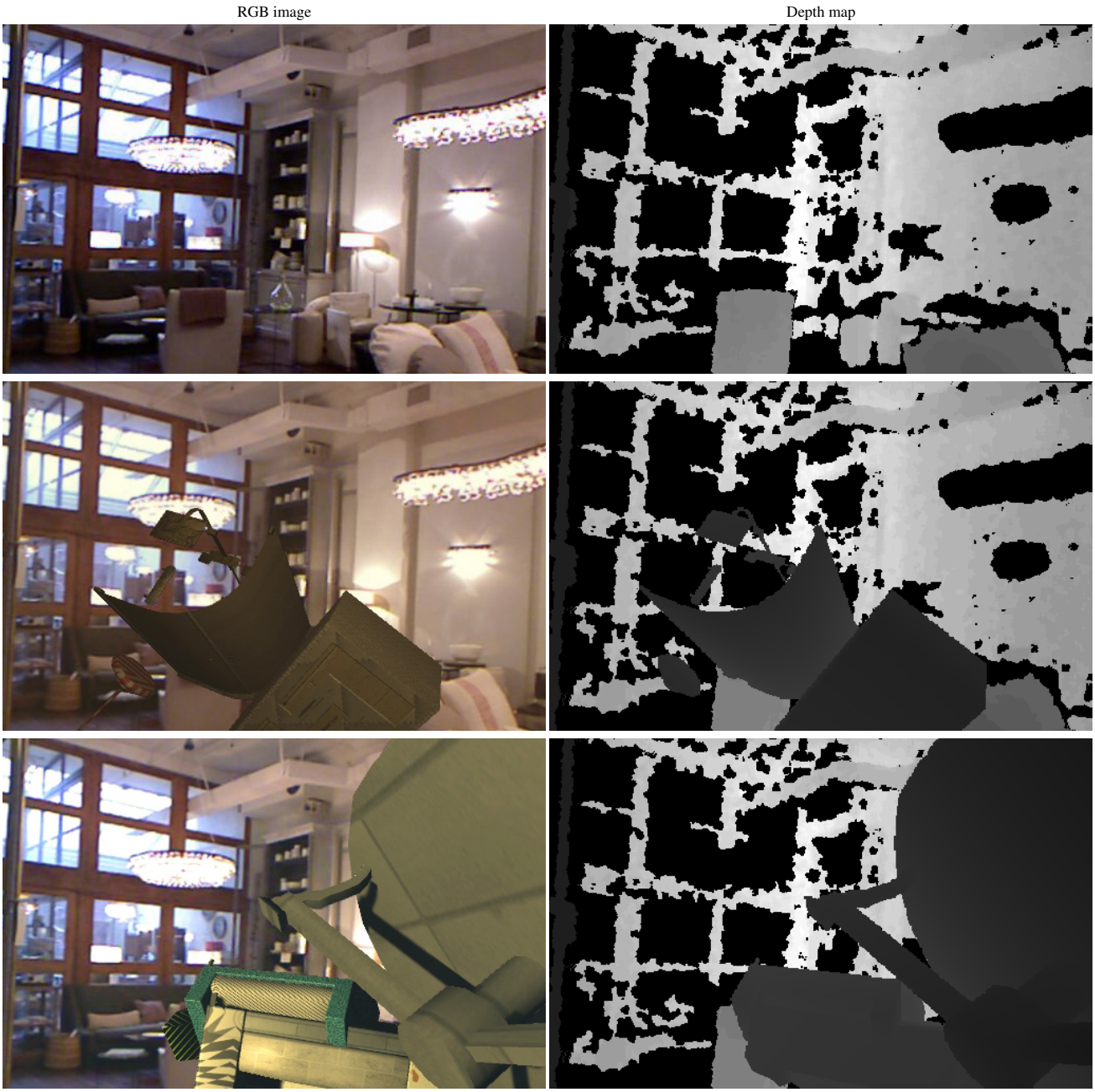

Figure 19. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).

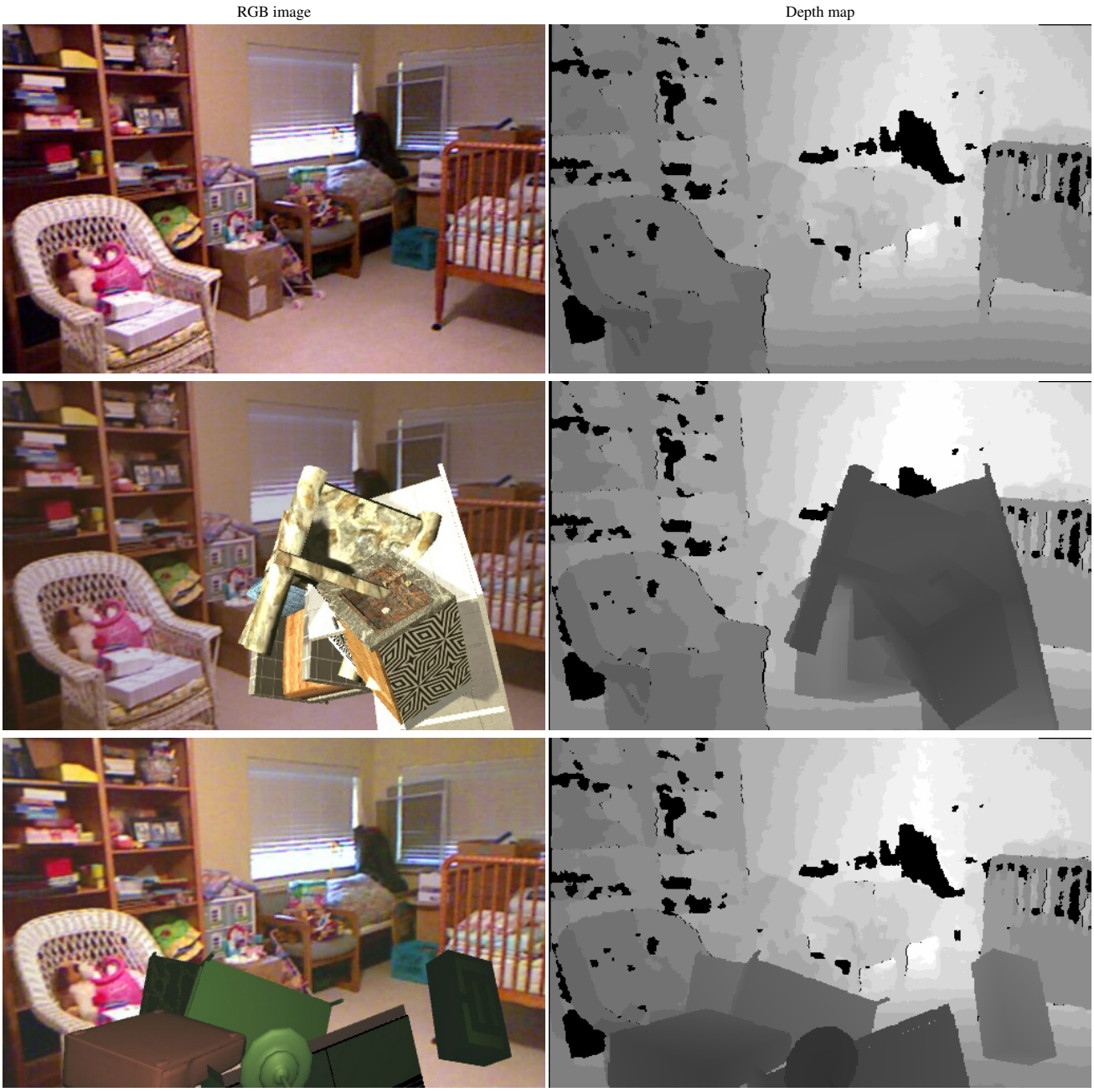

Figure 20. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).

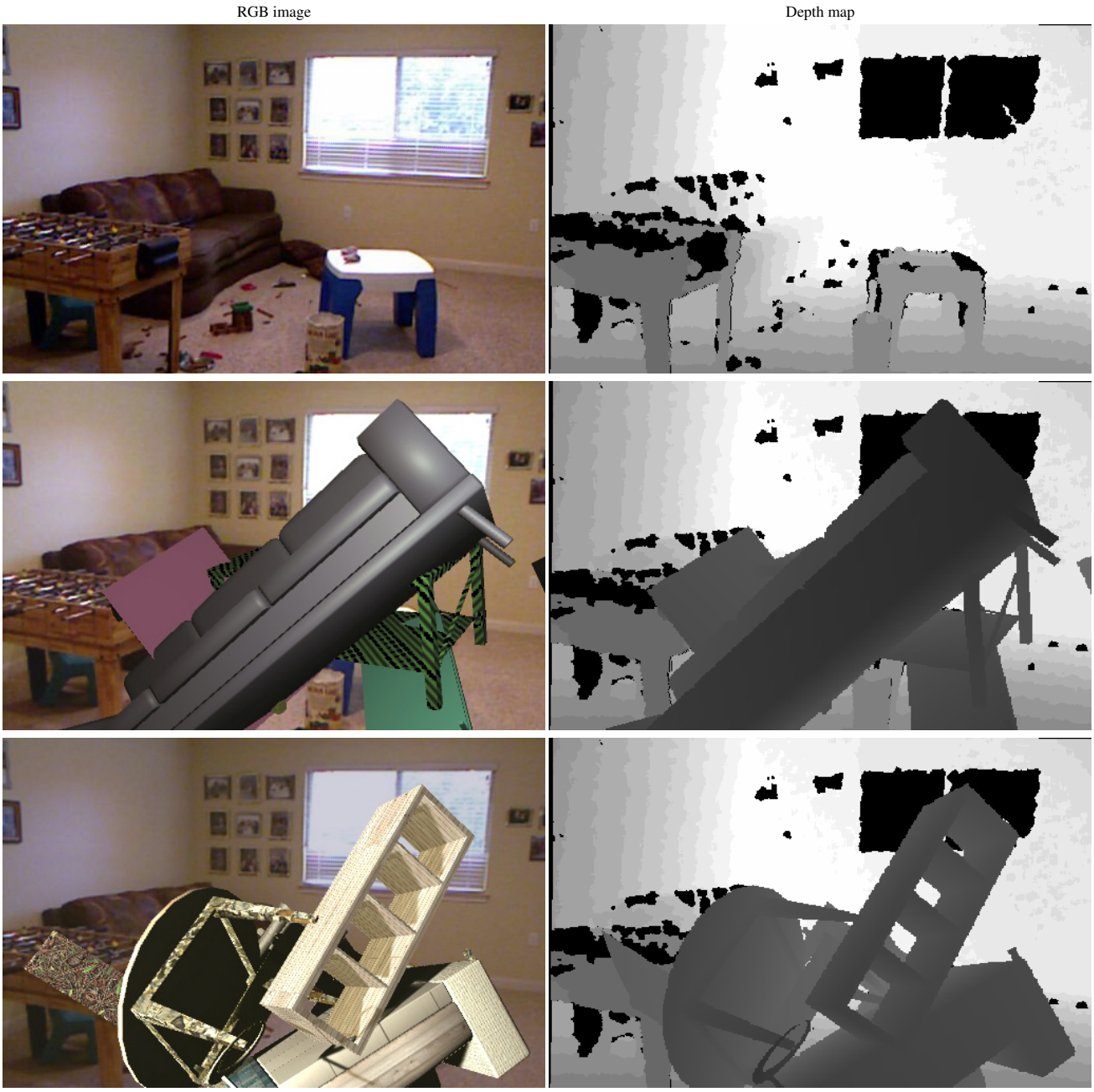

Figure 21. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).

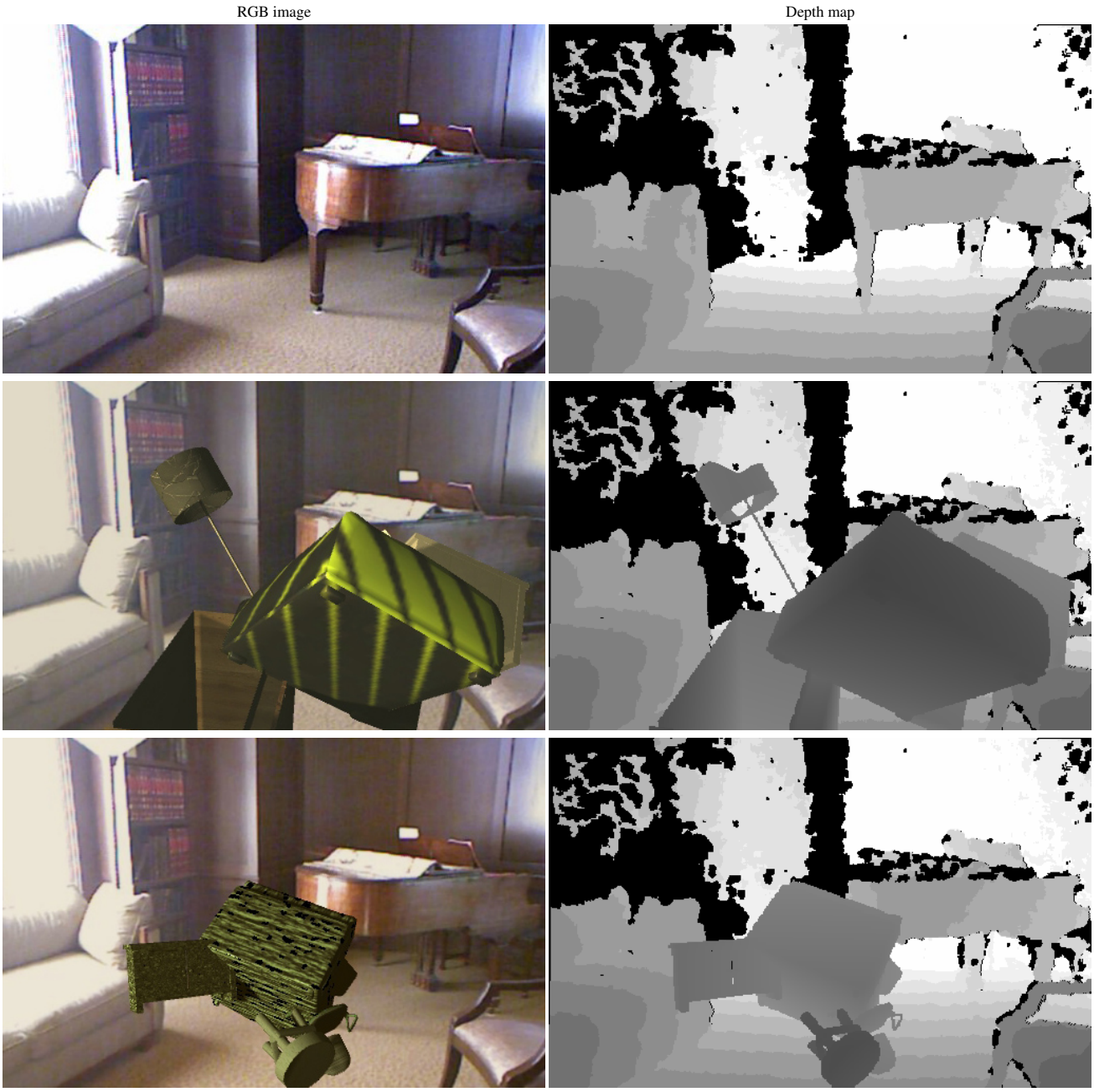

Figure 22. RGB-D image pairs: baseline (first from top) and enriched with virtual objects (the rest).

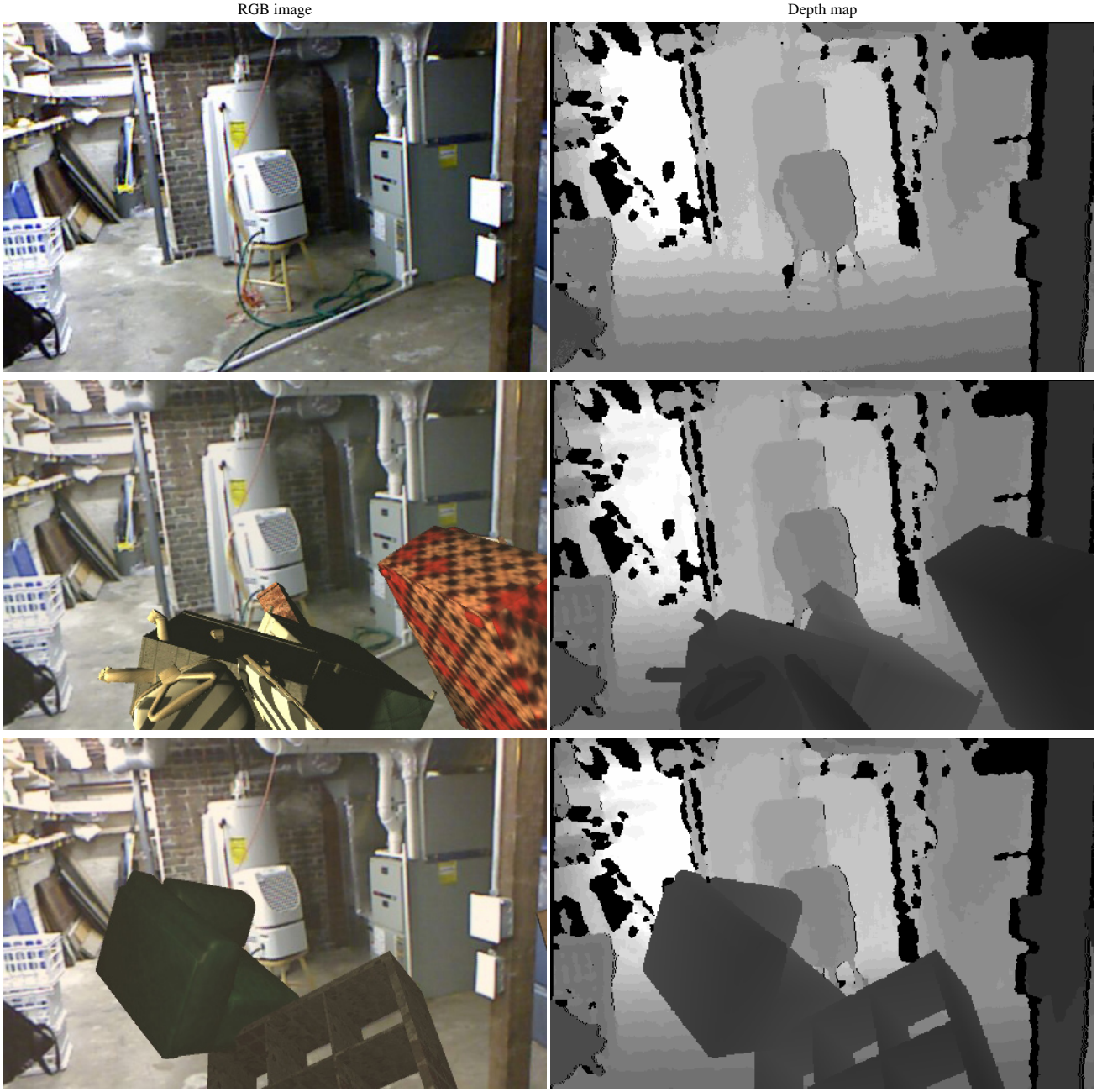

Figure 23. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).

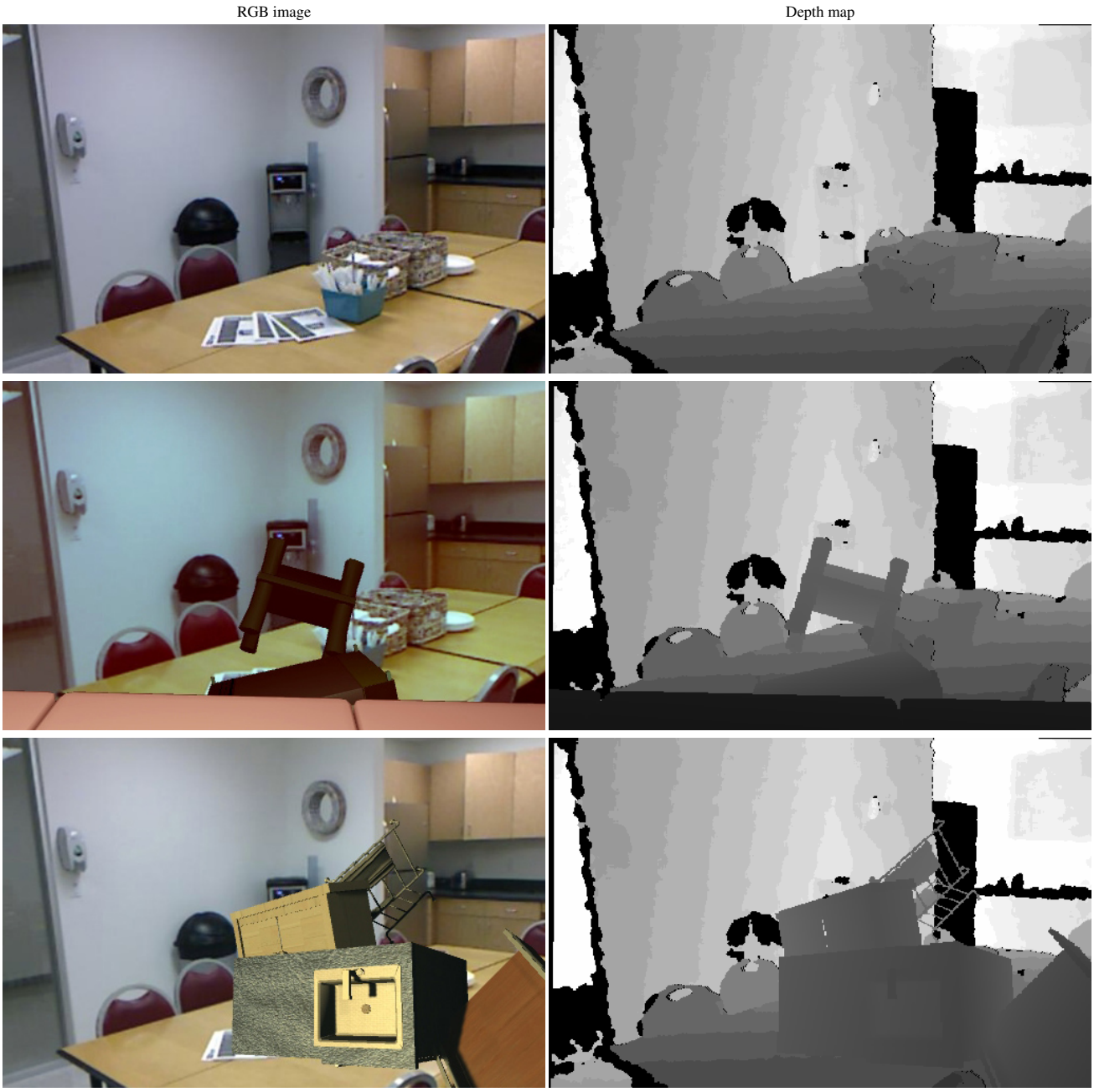

Figure 24. RGB-D image pairs: the original one (top) and enriched with virtual objects (the rest).