On-the-Fly Adaptation of Regression Forests for Online Camera Relocalisation
SUPPLEMENTARY MATERIAL

Tommaso Cavallari Stuart Golodetz* Nicholas A. Lord*
Julien Valentin Luigi Di Stefano Philip H. S. Torr

1. Analysis of Failure Cases

As shown in the main paper, our approach is able to achieve highly-accurate online relocalisation in under 150ms, from novel poses and without needing extensive offline training on the target scene. However, there are inevitably still situations in which it will fail. In this section, we analyse two interesting failure cases, so as to help the reader understand the underlying reasons in each case.

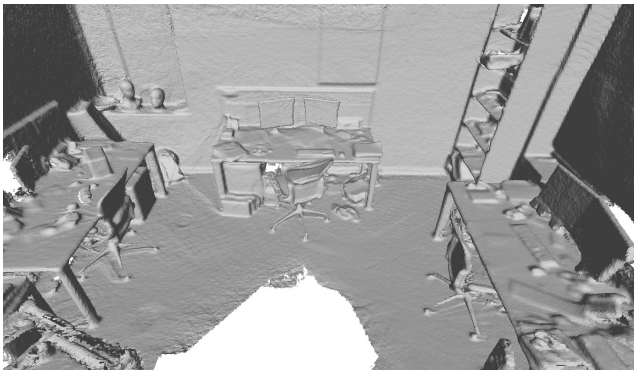
1.1. Office

The first failure case we analyse is from the *Office* scene in the 7-Scenes dataset [1]. This scene captures a typical office that contains a number of desks (see Figure 1). Unfortunately, these desks appear visually quite similar: they are made of the same wood, and have similar monitors and the same associated chairs. This makes it very difficult for a relocaliser such as ours to distinguish between them: as a result, our approach ends up producing a pose that faces the wrong desk (see Figure 1(d)).

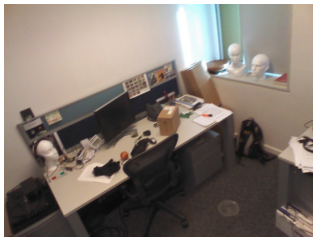
On one level, the pose we produce is not entirely unreasonable: indeed, it looks superficially plausible, and is oriented at roughly the right angle with respect to the incorrect desk. Nevertheless, in absolute terms, the pose is obviously very far from the ground truth.

To pin down what has gone wrong, we visualise the last 16 surviving camera pose hypotheses for this instance in Figure 2, in descending order (left-to-right, top-to-bottom). We observe that whilst the top candidate selected by RANSAC relocalises the camera to face the wrong desk, any of the next five candidates would have relocalised the camera successfully. The problem in this case is that the energies computed for the hypotheses are fairly similar for both the correct and incorrect poses.

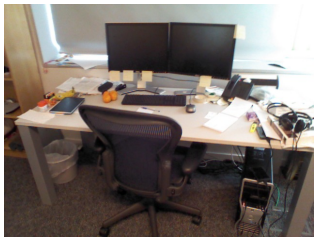
Although we do not investigate it here, one potential way of fixing this might be to score the last few surviving hypotheses based on the photometric consistencies between colour raycasts from their respective poses and the colour input image.



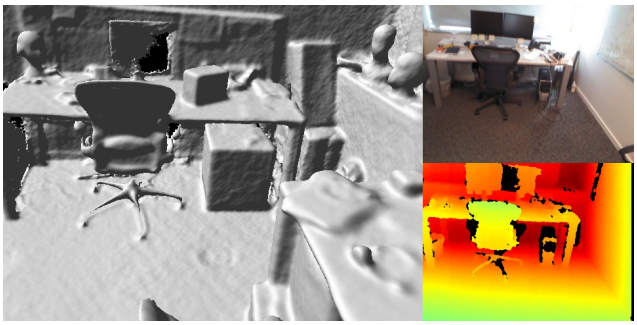
(a)



(b)



(c)



(d)

Figure 1: The *Office* scene from the 7-Scenes dataset [1] (a) contains multiple desks, e.g. (b) and (c), that can appear visually quite similar, making it difficult for the relocaliser to distinguish between them. In (d), for example, the relocaliser incorrectly chooses a pose facing the desk in (b), whilst the RGB-D input actually shows the desk in (c).

*S. Golodetz and N. Lord assert joint second authorship.

Figure 2: The top 16 pose candidates (left-to-right, top-to-bottom) corresponding to the failure case on the *Office* scene shown in Figure 1(d). The coloured points indicate the 2D-to-3D correspondences that are used to generate the initial pose hypotheses. Note that whilst the top candidate selected by RANSAC relocalises the camera to face the wrong desk, any of the next five candidates would have relocalised the camera correctly.

1.2. Stairs

The second failure case we analyse is from the *Stairs* scene in the 7-Scenes dataset [1]. This is a notoriously difficult scene containing a staircase that consists of numerous visually-identical steps (see Figure 3). When viewing the scene from certain angles (see Figure 4), the relocaliser is able to rely on points in the scene that can be identified unambiguously to correctly estimate the pose, but from viewpoints such as that in Figure 3(d), it is forced to use more ambiguous points, e.g. those on the stairs themselves or the walls. When this happens, relocalisation is prone to fail, since the relocaliser finds it difficult to tell the difference between the different steps.

As in the previous section, we can visualise the top 16 camera pose hypotheses for this instance to pin down what has gone wrong (see Figure 7). It is noticeable that in this case, none of the top 16 hypotheses would have successfully relocalised the camera. As suggested by the points predicted in the 3D scene for each hypothesis (which are often in roughly the right place but on the wrong stairs), this is because the points at the same places on different stairs tend to end up in similar leaves, making the modes in the leaves less informative (see Figure 5) and significantly reducing the probability of generating good initial hypotheses.

Unlike in the *Office* case, the problem here cannot be fixed by a late-stage consistency check, since none of the last few surviving hypotheses are of any use. Instead, one potential way of fixing this might be to improve the way in which the initial set of hypotheses is generated so as to construct a more diverse set and increase the probability of one of the initial poses being in roughly the right place. An alternative might be to adaptively increase the number of hypotheses generated in difficult conditions.

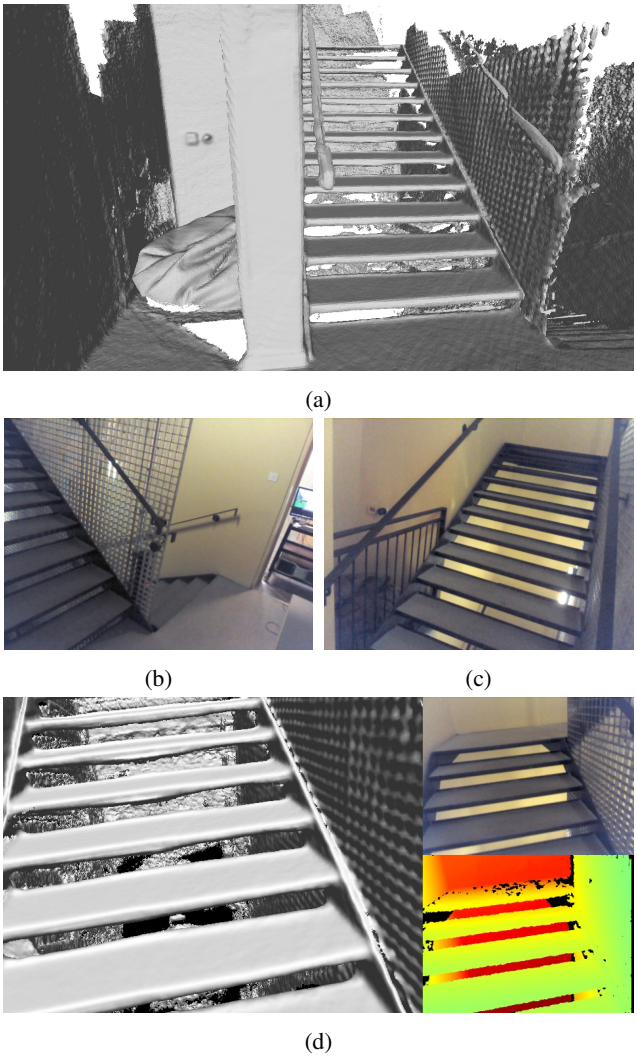


Figure 3: The *Stairs* scene from the 7-Scenes dataset [1] (a) is notoriously difficult, containing a staircase that consists of numerous visually-identical steps (see (b) and (c)). In (d), many of the 2D-to-3D correspondences predicted by the forest are likely to be of a low quality, since it is hard to distinguish between similar points on different stairs. This significantly reduces the probability of generating good initial hypotheses, leaving RANSAC trying to pick a good hypothesis from an initial set that only contains bad ones.

Figure 4: From certain angles in the *Stairs* scene, the relocaliser is able to rely on points in the scene that can be identified unambiguously to estimate the pose.

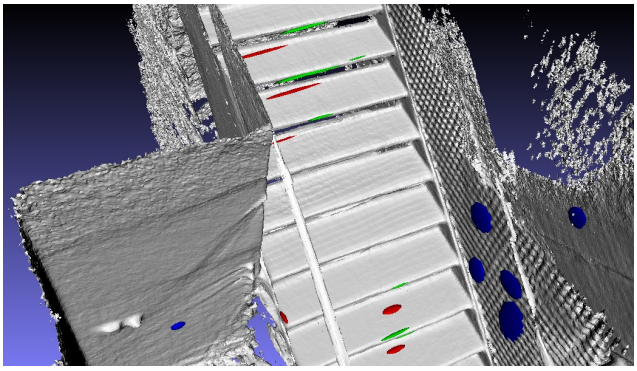


Figure 5: The modal clusters contained in the leaves for the points in the optimal camera pose hypothesis from Figure 7. It is noticeable that points at the same places on different stairs end up in the same leaves, making the distributions in those leaves less informative.

2. Further Successful Examples

Some further examples of successful relocalisation, this time in the *Fire* scene from the 7-Scenes dataset [1], can be seen in Figure 6. As in Figure 4, it is noticeable that the relocaliser tries to rely on points in the scene that can be identified unambiguously where these are available, something that is clearly easier in sequences such as *Fire* that contain many easily-distinguished objects.

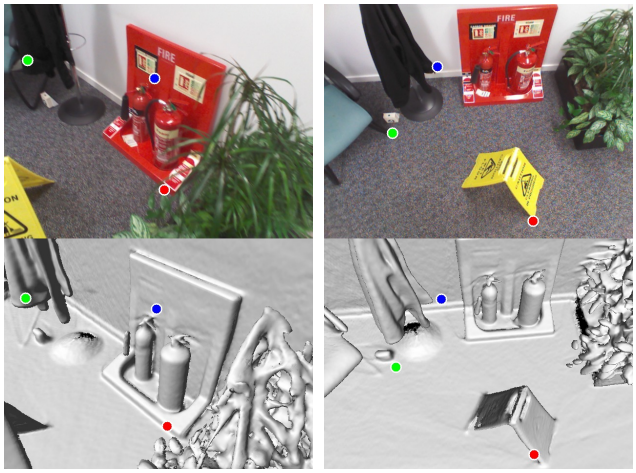


Figure 6: Further examples of successful relocalisation in the *Fire* scene from the 7-Scenes dataset [1]. To estimate the pose, the relocaliser tries to rely on points in the scene that can be identified unambiguously.

References

[1] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, and A. Fitzgibbon. Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images. In *CVPR*, pages 2930–2937, 2013. 1, 3, 4

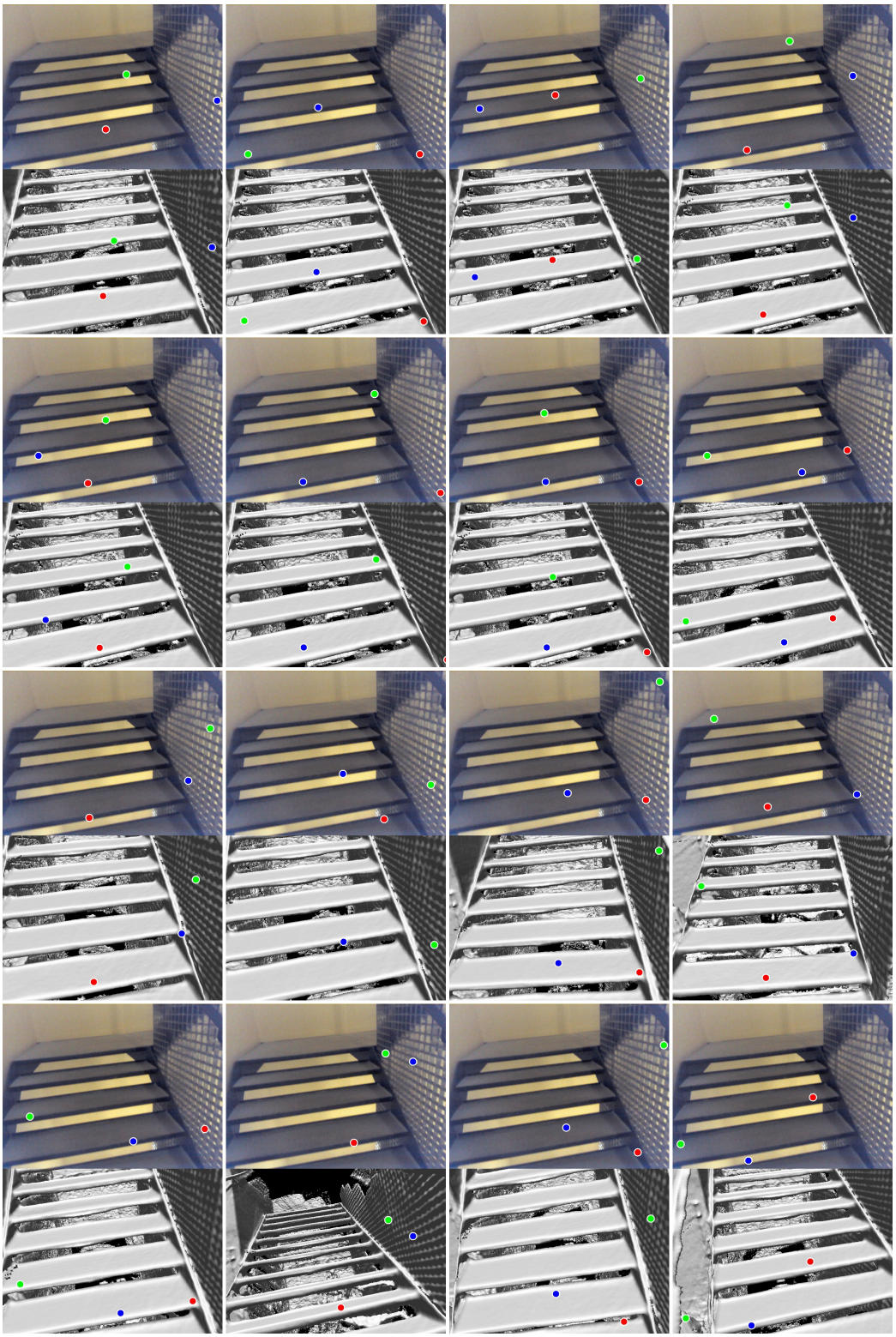


Figure 7: The top 16 pose candidates (left-to-right, top-to-bottom) corresponding to the failure case on the *Stairs* scene shown in Figure 3(d). The coloured points indicate the 2D-to-3D correspondences that are used to generate the initial pose hypotheses. Note that in this case, none of the candidates would relocalise the camera successfully. This is likely because the points at the same places on different stairs tend to end up in similar leaves, making the modes in the leaves less informative and significantly reducing the probability of generating good initial hypotheses.