Additional File 1

S1 Lateral resolution and depth of field

Supplementary Fig. S1(a) shows in-focus images of a USAF target captured by one of the 54 cameras, which are nominally identical. Supplementary Fig. S1(b) characterizes the DOF of our system, based on acquiring a z-stack of a calibration target consisting of white noise printed on paper adhered to a glass substrate. The plot shows the mean image gradient as a function of axial position. This figure is analyzed in Sec. 5.1.

S2 Sample scan repeatability and calibration errors

Since we are stitching multiple images into a single composite, seams, or stitching boundaries, may appear and may be geometric (spatial misalignment) or photometric (patch-to-patch intensity or color ratio variation). The former in principle can be pre-calibrated by accurately estimating the camera poses of all 3456 views (54 cameras $\times$ 64 sample scan positions) based on a flat calibration target. The latter, however, in general cannot be fully accounted for due to differing scattering properties of and

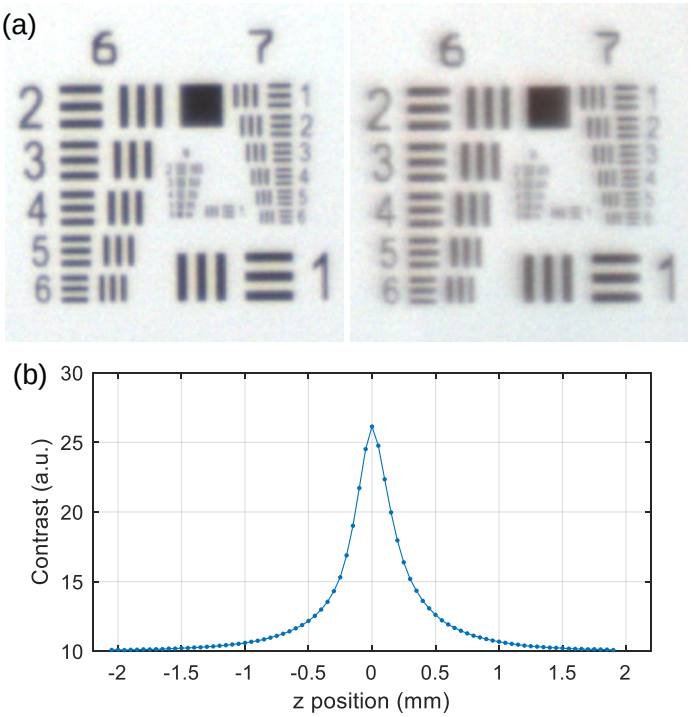

Fig. S1 Lateral resolution (a) and DOF (b) characterization of our imaging system. In (a), the left USAF target was placed at the center of the FOV, while the right USAF target was placed at the edge of the FOV.

shadows cast by the sample, especially considering that the sample is scanned relative to the light sources.

To reduce photometric seams, when we accumulate patches during gigamosaic generation (Sec. 4.2.2), in regions where patches overlap, we have the option to blend (average overlapped pixels). However, when there are calibration errors leading to imperfect registration and geometric seams, blending can cause blurring and therefore reduce spatial resolution. The reason for imperfect registration is that we fix the calibrated camera poses and distortions during 3D height estimation. Thus, any calibration errors remain uncorrected. These calibration errors do not result from our registration algorithm described in Sec. 4.1, which contains no visible seams or blurring across the full FOV. Rather, the calibration errors come from the repeatability of the sample scanning, which may be affected by the accuracy, precision, or repeatability 3D sample stage, how rigid the sample during data acquisition, and how stationary the object remains relative to the stage (e.g., if it whole object slips or subcomponents of the object dislodge). This calibration issue can be addressed through a more rigid sample mount, better stage (with specifications better than the expected resolution), or scanning the MCAM instead of the object to reduce the impact of sample idiosyncrasies.

S3 Accounting for low-contrast objects with confidence maps and weighted sharpness

Both stereo- and sharpness-based methods require the sample to have fine texture or features. For stereo methods, these features are registered from different views, while for sharpness methods, these features must have different appearances when translated to different axial positions. Thus, in regions of the sample that are low contrast, the reconstruction may result in larger errors. To account for this, we introduce two strategies. The first is the use of a weighted sharpness loss (Eq. 3), described in Sec. 4.2.1, with the idea that regions of low sharpness (or low contrast) should contribute less or not at all to the the total loss. The other strategy is to compute a coregistered confidence map, based on the max sharpness values across the z-stack dimension. Fig. S2(b) shows that specific PCB components (e.g., the row of 16 components on the righthand side) as well as white text, lines, and stickers are particularly low-contrast, thus cautioning the user that the height values may be inaccurate. We can see evidence of this in the argmax-only height map in Fig. S2(d), which erroneously assigns inflated height values to the white text and lines, unlike the height map generated by the full version of our method that also incorporates the weighted sharpness loss and stereo (Fig. S2(c)). Thus, the confidence map can be reported alongside our reconstructions to safeguard or warn the user when parts of the sample of interest are low contrast, and therefore whose height estimates may be inaccurate.

S4 CNN architecture

The nearly symmetric encoder-decoder CNN architecture used to map from z-stacks to height maps is summarized in Fig. S3. The input is a z-stack with flexible lateral

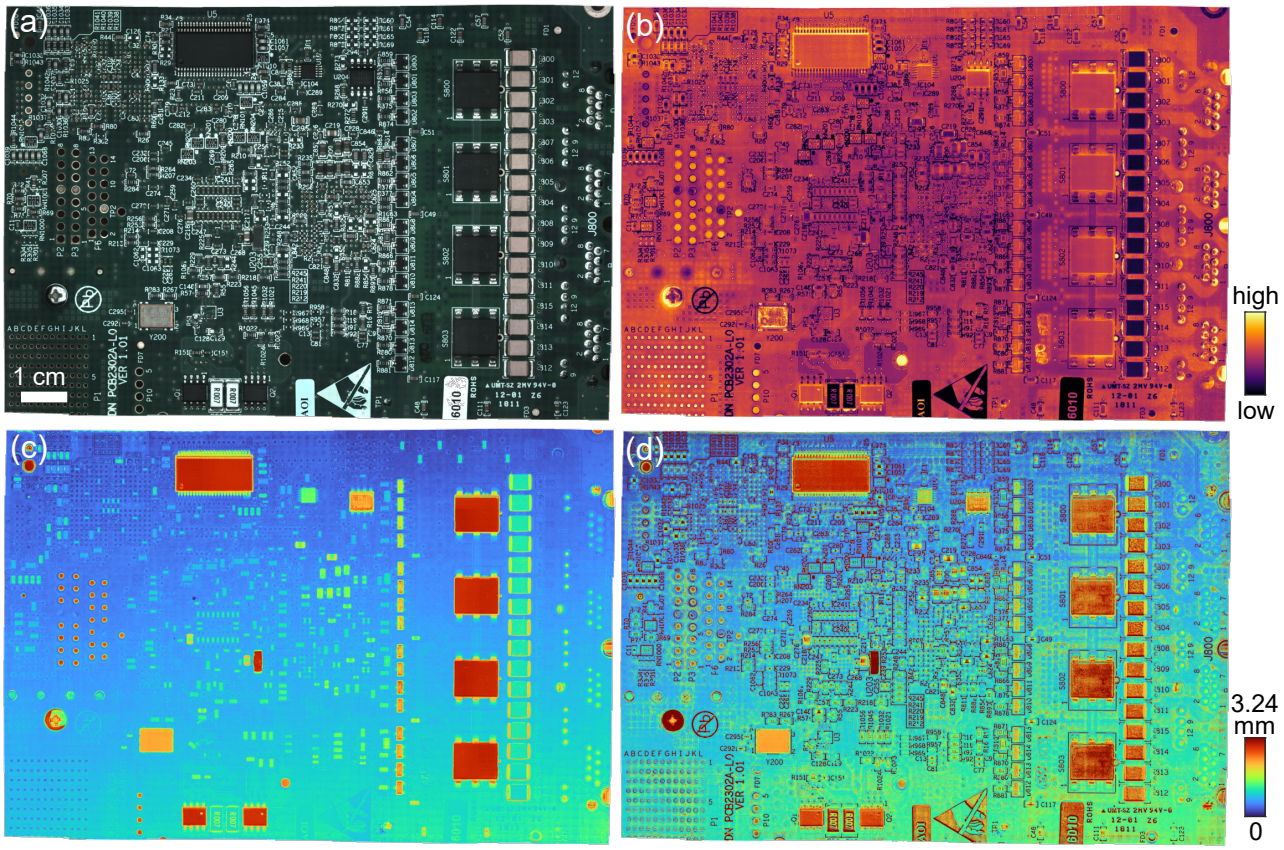

Fig. S2 PCB analysis. (a) All-in-focus photometric gigamosaic. (b) Confidence map, based on max sharpness across the z-stack dimension. (c) 3D height map reconstruction. (d) 3D height map reconstructed using only argmax across the z-stack dimension.

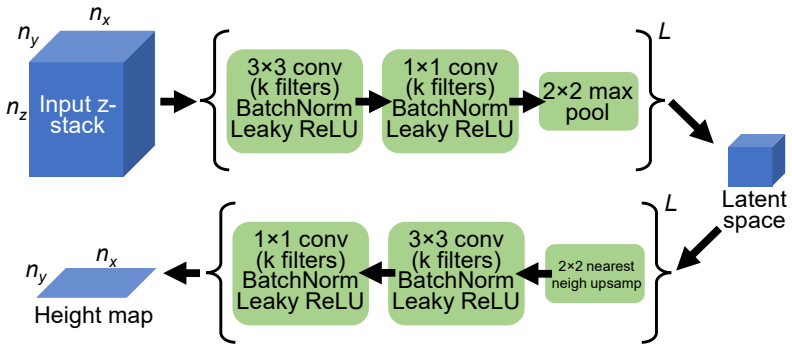

Fig. S3 Architecture of the CNN mapping from z-stacks to height maps.

dimensions (n_x and n_y) and a fixed number of z slices ($n_z = 65$). The encoder path consists of L downsample blocks, each consisting of k 3×3 convolutional filters (where k can vary from block to block), followed by batch normalization and a leaky ReLU activation. The block is terminated with a 2×2 max pooling operation (the convolutions do not alter lateral pixel count). The values for k can be specified by a list of length L (e.g., six downsample blocks could be specified by [32, 32, 32, 64, 64, 64]).

The decoder path consists of blocks of the same components, except a 2×2 nearest neighbor upsampling operation is performed at the beginning of the block (instead of a 2×2 max pooling at the end). The output is an $n_x \times n_y$ height map, after summing the channel dimension of the output of the final block. The number of blocks and filters per block is the same as in the encoder path, except reversed (e.g., if the encoder path is given by [32, 32, 32, 64, 64, 64], then the upsample blocks of the decoder path are summarized by [64, 64, 64, 32, 32, 32]). No skip connections were used.

In this paper, we used a CNN architecture given by [32, 32, 32, 64, 64, 64] for all the objects in this paper, except for the PCB sample, which had a reduced architecture given by [32, 32, 32, 64, 64] for further regularization.

S5 Additional validation experiments on the PCB sample

Here, we report on additional experiments on the PCB sample (Supplementary Figs. S4 and S5). Specifically, we compared the height estimates of STARCAM with those of a commercial 3D surface profiler (Keyence VK-X3050), which features a laser confocal microscope and focal stacking microscope, and an electronic caliper with a 20-µm nominal accuracy. The caliper measurements are of the bulk heights of the components (reported as the mean of 10 independent measurements). We applied these measurement approaches to six arbitrarily chosen flat-top electronic components on the PCB. These results are summarized in Supplementary Figs. S4 and S5.

Overall, STARCAM's height estimates are consistent with those of the Keyence profilometer, apart from when there were artifacts. In particular, the Keyence profilometer exhibits some artifacts from tiling and stitching multiple FOVs (Fig. S4(a,c,d,e)). The Keyence focal stacking method (third column) sometimes has artifacts where the object is highly reflective (Fig. S4(b,d)) or where there are white lines or text (Fig. S4(b,f)). STARCAM, which also uses focal stacking, overcomes the latter artifacts by incorporating sharpness weighting in our reconstruction algorithm (Supplementary Sec. S3), without which we would also have the same white-line artifacts (Supplementary Fig. S2(d)). STARCAM also exhibits some minor artifacts in the PCB substrate that are less obvious in the Keyence profilometric measurements. Nevertheless, STARCAM's height estimates were always consistent with the caliper estimates. Further, we emphasize that STARCAM can maintain accurate profilometric estimates over a much larger FOV (>110 cm^2).

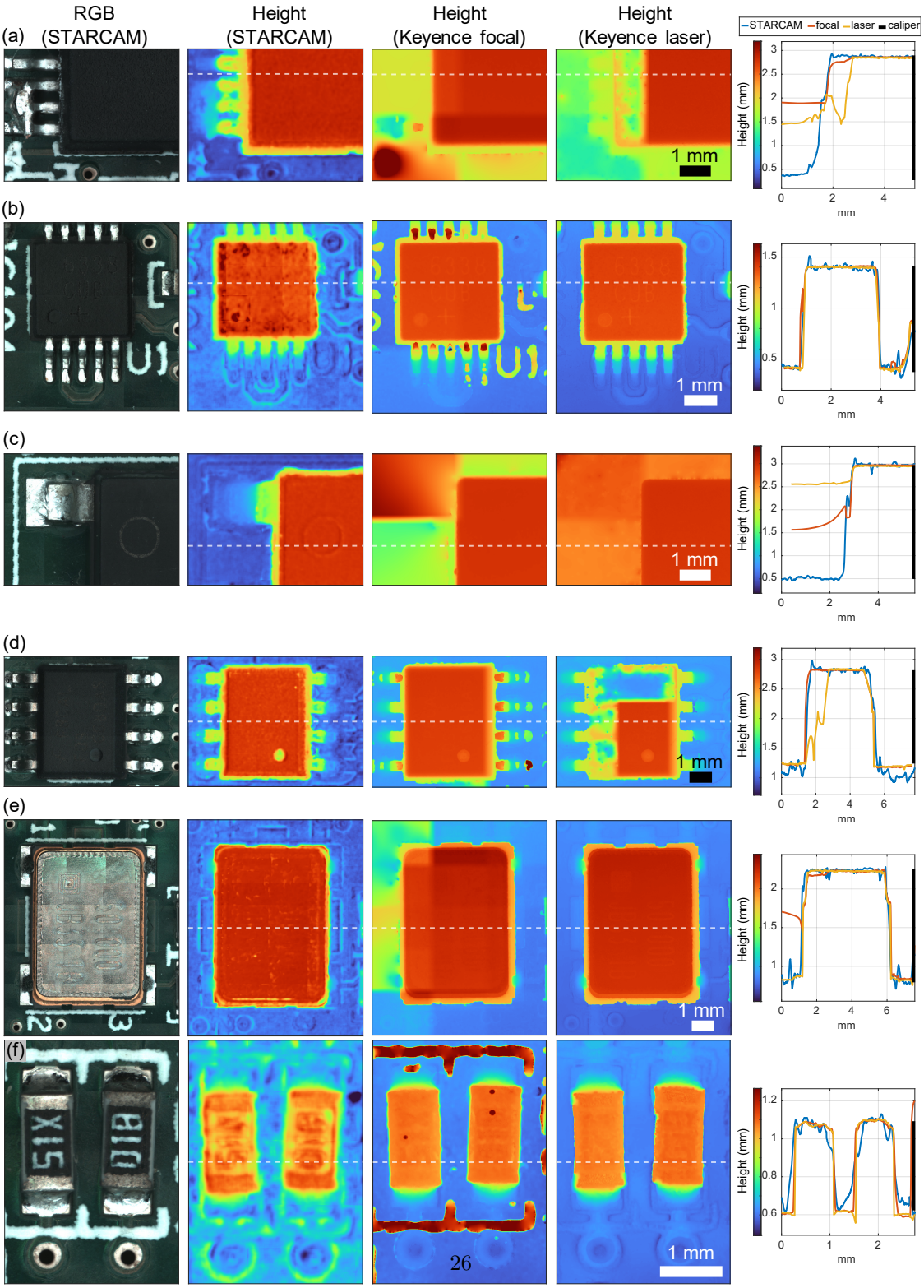

Fig. S4 Comparison of STARCAM with laser confocal microscopy (Keyence), focal stacking (Keyence), and calipers at six locations (a-f) across the PCB sample (Fig. 7). These locations are indicated in Supplementary Fig. S5. The first column shows the photometric (RGB) image. The middle three columns show the height map comparisons, with the same color range in each row. The fifth column shows the 1D height profiles indicated by the dotted lines in the middle three columns. The 1D height profiles are also accompanied by the caliper estimates (based on the mean of 10 independent measurements).

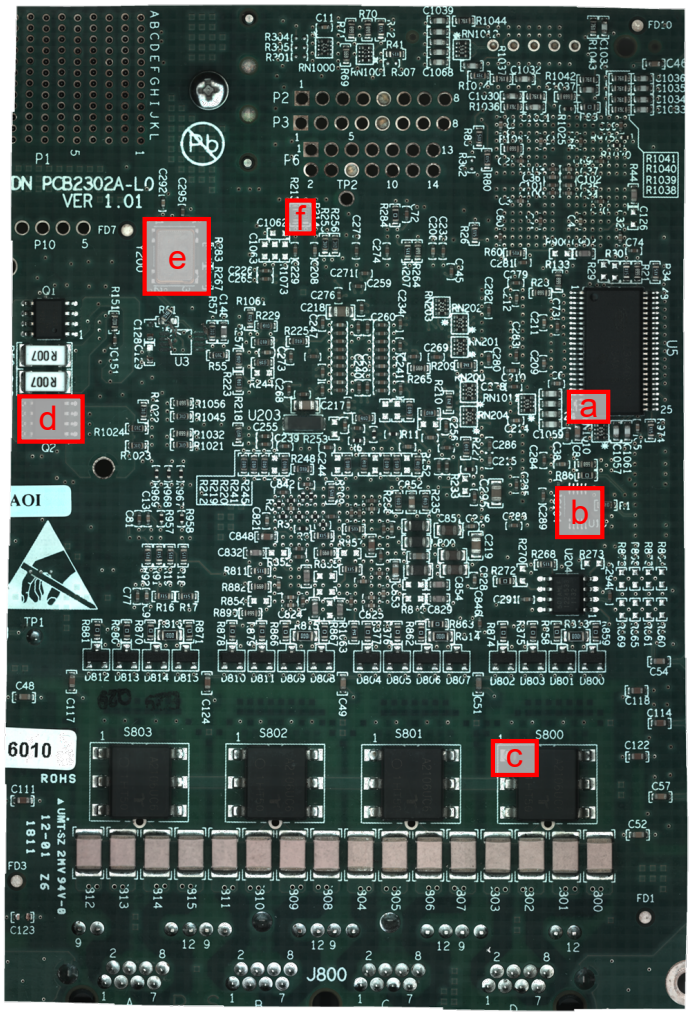

Fig. S5 The locations of the crops analyzed in Supplementary Fig. S4.

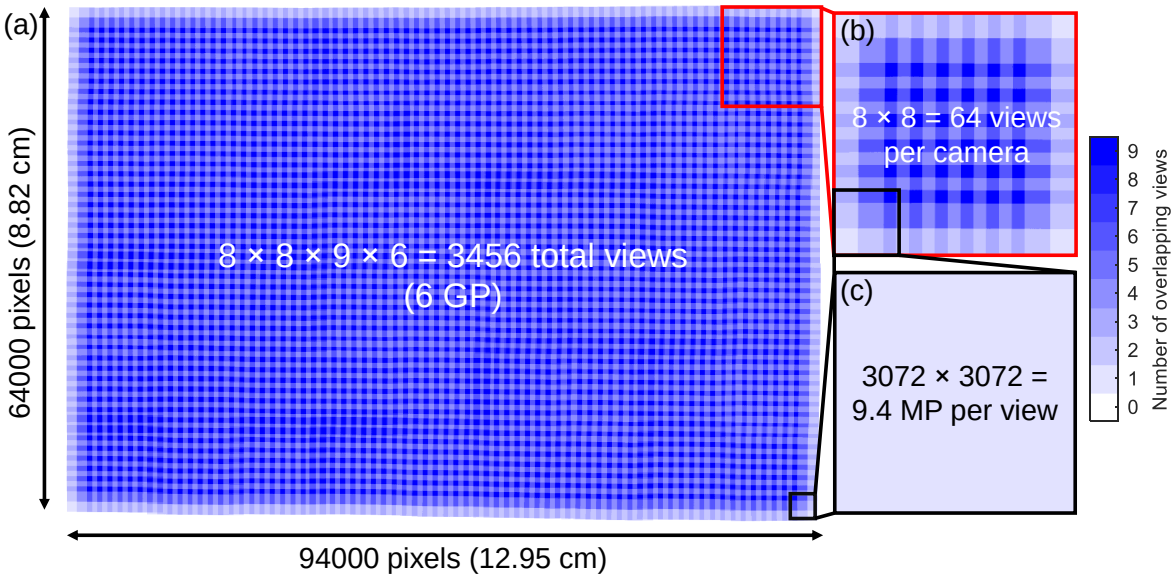

Fig. S6 Overlap maps, showing how many times (0-9) each point in the FOV. (a) The overlap map for all 3456 camera views, based on the joint camera calibration from imaging a flat reference target. (b) The overlap map for one 8×8 scan from a single camera. (c) Each view contains over 9 MP.