— Supplementary Material —
Planar Affine Rectification from Local Change of Scale and Orientation

1. Degenerate Cases

Here, we discuss the degenerate configurations of the proposed solvers.

Collinear Points. Assume the coordinates of two scale features and the coordinates of a finite vanishing point in a dehomogenized representation, derived from two orientation features, are collinear, *i.e*. they can be expressed as (x, y), $(x + \lambda_1\delta_x, y + \lambda_1\delta_y)$ and $(x + \lambda_2\delta_x, y + \lambda_2\delta_y)$, respectively. Let h_7 and h_8 be the parameters of the projective transformation, as in Eq. (1). The matrix

$$\mathbf{A} = \begin{bmatrix} x & y & -h_7x - h_8y + 1 \\ x + \lambda_1\delta_x & y + \lambda_1\delta_y & -h_7(x + \lambda_1\delta_x) - h_8(y + \lambda_1\delta_y) + 1 \\ x + \lambda_2\delta_x & y + \lambda_2\delta_y & 0 \end{bmatrix} \quad (25)$$

is constructed from these features, as described in Eq. (13). The determinant is as follows:

$$\det(\mathbf{A}) = \lambda_1(y\delta_x - x\delta_y)\left[-h_7(x + \lambda_2\delta_x) - h_8(y + \lambda_2\delta_y) + 1\right] \quad (26)$$

Since $(x + \lambda_2\delta_x, y + \lambda_2\delta_y)$ are the coordinates of the vanishing point and $\begin{bmatrix} -h_7 & -h_8 & 1 \end{bmatrix}$ is the vanishing line, we receive that $\det(\mathbf{A}) = 0$ and $\mathbf{A}$ is singular with the vector

$$\mathbf{n} = \begin{bmatrix} h_7(y\delta_x - x\delta_y) + \delta_y \\ h_8(y\delta_x - x\delta_y) - \delta_x \\ y\delta_x - x\delta_y \end{bmatrix} \quad (27)$$

spanning its null space, *i.e*. $\mathbf{An} = \mathbf{0}$. If the vanishing point is at infinity, we get that

$$\det(\mathbf{A}) = \lambda_1(y\delta_x - x\delta_y), \quad (28)$$

and $\mathbf{A}$ is singular when $y\delta_x = x\delta_y$, which happens when the line which intersects the three points also intersects the origin. Then, the null space of $\mathbf{A}$ is spanned by

$$\mathbf{n} = \begin{bmatrix} \delta_y, & -\delta_x, & 0 \end{bmatrix}^T \quad (29)$$

In both cases, the vector $h = \begin{bmatrix} h_7 & h_8 & 1 \end{bmatrix}^T$ solves the equation $\mathbf{Ah} = \mathbf{b}$. Therefore, there is a one-dimensional affine space of solutions given by

$$\{\mathbf{h} + \lambda\mathbf{n} \ : \lambda \in \mathbb{R}\}. \quad (30)$$

Zero Vanishing Point. Consider two scale features with coordinates (x, y) and $(x + \delta_x, y + \delta_y)$, and two orientation features, (x_1, y_1, θ_1) and (x_2, y_2, θ_2), from which two lines, $\mathbf{l}_1 = \mathbf{l}(x_1, y_1, \theta_1)$ and $\mathbf{l}_2 = \mathbf{l}(x_2, y_2, \theta_2)$ (as in Eq. (9)), are derived whose intersection $\mathbf{v} = \mathbf{l}_1 \times \mathbf{l}_2 = \mathbf{0}$ is a degenerate vanishing point. Let h_7 and h_8 be the parameters of the projective transformation, as in Eq. (1). The matrix

$$\mathbf{A} = \begin{bmatrix} x & y & -h_7x - h_8y + 1 \\ x + \delta_x & y + \delta_y & -h_7(x + \delta_x) - h_8(y + \delta_y) + 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (31)$$

is constructed from these features, as described in Eq. (13). The matrix $\mathbf{A}$ is singular with the same vector $\mathbf{n}$ as in Eq. (27) as its null space. Similarly to the previous degenerate case, the vector $\mathbf{h} = \begin{bmatrix} h_7 & h_8 & 1 \end{bmatrix}^T$ solves the equation $\mathbf{Ah} = \mathbf{b}$, and the family of solutions detailed in Eq. (30) is applicable here as well. A vanishing point equates to zero if

$$\theta_2 = \theta_1 + \pi k_1 \text{ and } \theta_1 = \varphi + \pi k_2, \text{ for } k_1, k_2 \in \mathbb{Z} \quad (32)$$

where $\varphi = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$ is the angle of the line that passes through (x_1, y_1) and (x_2, y_2). In other words, the condition is met if the angles θ_1, θ_2, and φ are pairwise equal or opposite.

Discarding Degenerate Sample Configurations. The detection and elimination of these two degenerate sample configurations, resulting in singular solutions, is straightforward. With a set of scale and orientation features, we employ a standard collinearity test to ensure the sample's positions are not all collinear, and remove any orientation feature pairs that meet the criteria outlined in Eq. (32).

2. Additional Real-World Evaluation

We offer additional real-world evaluation of the proposed method (SIFT + LSD) in Fig. 10. Its success in affinely rectifying an image plane is measured and compared with two benchmarks: vanishing point-based estimation (LSD) and scale-based estimation (SIFT). The success of each method is measured by the extent to which lines—which are considered to be parallel in the real-world plane—are parallel in the respective affinely rectified image planes. To quantify this, two sets of lines, each with lines which are parallel to one another in the real-world plane, are estimated.

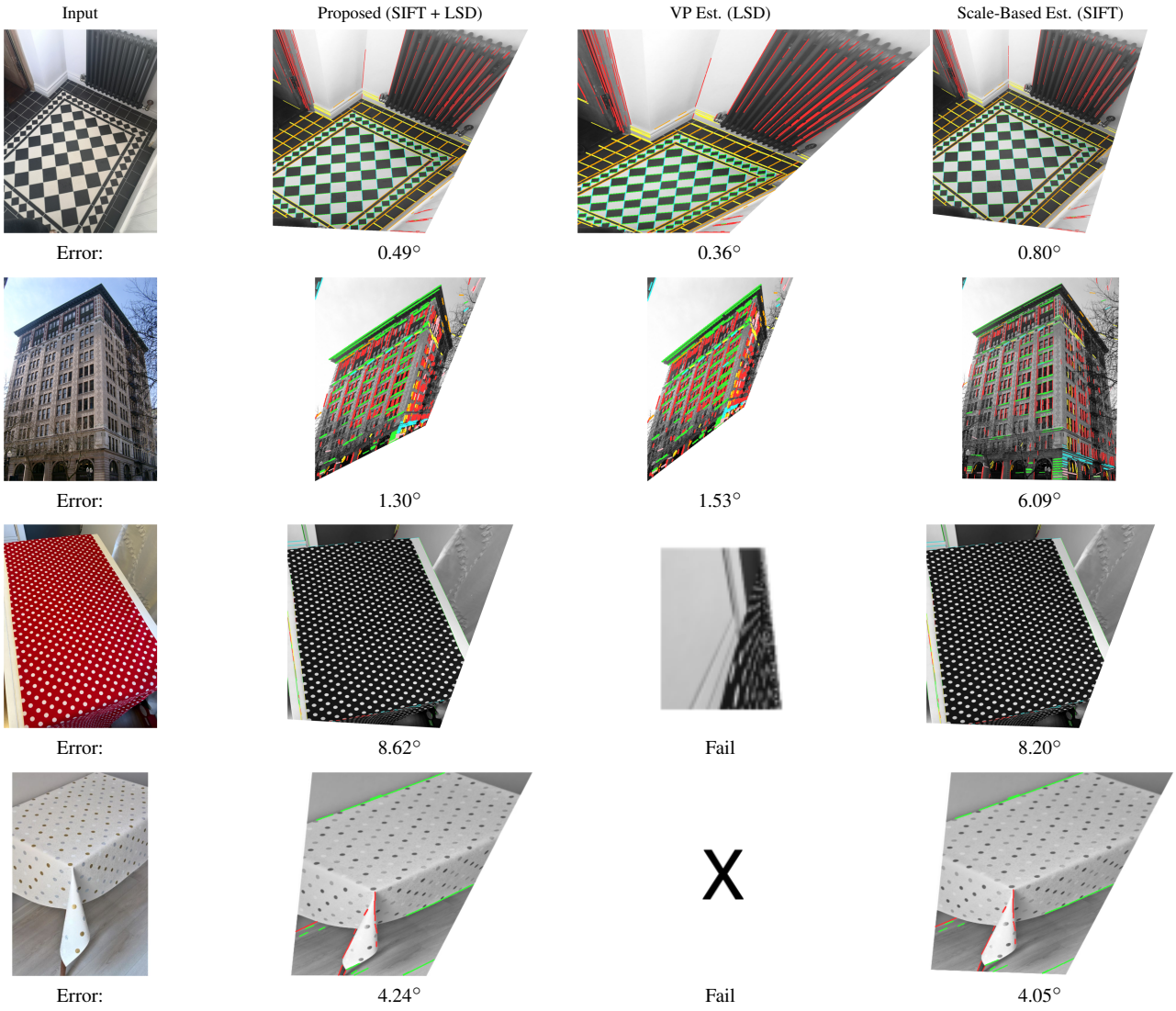

Figure 10. **Real World Evaluation Against Benchmarks.** Each row is comprised of a real-world image followed by an affine rectification of an imaged plane using the proposed method (SIFT + LSD), the VP estimation (LSD) and scale-based estimation (SIFT) benchmarks. Sets of parallel lines are clustered together by estimating a vanishing point they all converge to using the Progressive-X algorithm [2]. The sets are distinguished by different colors. The error is the average pairwise angle difference between all lines in the same set.

First, multiple vanishing points for each imaged plane are estimated using the Progressive-X algorithm [2] on the input images in Fig. 10. Then, the pair of vanishing points with the most inliers is selected. Lines converging to each of the two vanishing points are considered ground truth inliers. The measure of parallelism of the lines in each set is defined as the average angular difference between any line pair. We note that this process may be biased towards line-based estimators.

As expected, the proposed method is on par with the VP-based estimator when linear patterns dominate, achieving marginally worse results in the first row and better in the second. Here, the scale-based estimator is highly inaccurate. In the last two rows, the VP-based estimator fails as there are only a few lines. Our method is on par with the scale-based estimator. This mirrors the conclusion from our main experiments: the proposed hybrid approach maintains good accuracy in scenarios where single-feature methods (line or scale) fail.

3. Method Runtime Comparison

In Fig. 11, the mean and standard deviation for the runtimes of our two proposed methods and the benchmarks, which are used to produce the findings in Fig. 6, are presented.

	Feature Detection [ms]	Model Estimation (1000 iterations) [ms]
Proposed (SIFT + LSD)	394 ± 55	110 ± 42
Proposed (SIFT)	274 ± 42	76 ± 15
VP Estimation (LSD)	177 ± 49	38 ± 36
Scale-Based Estimation (SIFT)	245 ± 66	30 ± 19

Figure 11. **Mean and standard deviation of runtimes** of the experiments on the generated data in Sec. 5, displayed in Fig. 6. Columns (left to right): mean and standard deviation of feature detection run times, and mean and standard deviation of 1000 RANSAC iterations in homography estimation phase. Rows (top to bottom): proposed method with SIFT + LSD features, with SIFT features, vanishing points estimation, and scale-based estimation.

As anticipated, the average runtime for detecting features in our proposed method that employs SIFT features and line segments roughly equals the sum of runtimes for feature detection in the vanishing points estimation method and scale-based estimation method, due to the combined usage of both their feature sets. The average runtime for 1000 RANSAC iterations with our proposed method is approximately 3-4 times longer than the benchmarks. However, as the total runtime of the proposed method remains below 0.5 seconds, it remains efficient for the purpose of planar affine rectification. Moreover, it consistently achieves accurate results where marginally faster alternatives fail. This runtime increase stems from our method's broader feature selection available during the RANSAC minimal fitting stage. In experiments, we limited the number of RANSAC features to 1000, without restricting the number of line segments that can be detected. Our method, in practice, requires significantly fewer SIFT features and line segments for satisfactory results, allowing for a notable reduction of scale and orientation features to optimize runtimes.

4. Insights into Multi-Scale Rectification

In our case study, we examine the synthetic image depicted in Fig. 12a, which comprises examples from two separate feature categories: specifically, black circles with radii of 10 and 20 pixels. Fig. 12b illustrates a projective transformation of Fig. 12a via a homography defined by Eq. (1), with parameters $h_7 = 1 \times 10^{-4}$ and $h_8 = 1 \times 10^{-4}$. This transformation alters the apparent scale of each local feature depending on their positions, causing features from the same category to appear at varying scales. The histogram in Fig. 12c presents these scale distributions of both images.

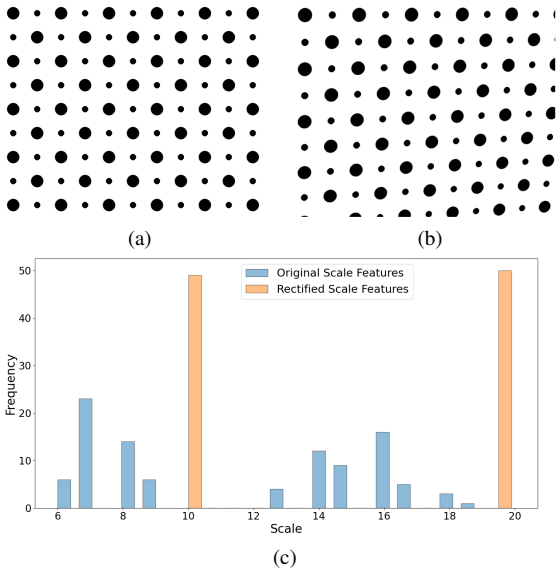

Figure 12. **Illustration of alignment of scales** into clusters of similarly sized features following affine rectification.

We identified that affine rectification that converts Fig. 12b back to Fig. 12a effectively aligns the scales in the input image into clusters of features of similar size, representing distinct categories in the rectified image. In essence, distinct scale feature classes form peaks in the scale histogram in the rectified image.