# SGDE: Self-supervised Geometry Degradation Estimation Framework for Coded Aperture Compressive Spectral Imaging

## Supplementary Material

This appendix includes: (1) Sensing matrix derivation (Sec. A), (2) theoretical analysis of the estimation range (Sec. B), (3) supplementary results for SD-CASSI (Sec. C), and (4) additional experiments for DD-CASSI (Sec. D).

## A. Sensing Matrix Derivation

The sensing matrix $\Phi$ in Eq. (4) is constructed from the mask and the prism dispersion. We first vectorize the mask $\boldsymbol{M}$ and form a diagonal modulation matrix:

$$J = \text{Diag}(\text{vec}(\boldsymbol{M})) \tag{11}$$

where $\text{vec}(\boldsymbol{M}) \in \mathbb{R}^{HW}$, $J \in \mathbb{R}^{HW \times HW}$. Then, for the $c$-th spectral channel, the prism introduces a wavelength-dependent shift of $D(c)$ pixels along the dispersion direction. The shifted modulation matrix is given by:

$$J_c(i + D(c), j) = J(i, j) \tag{12}$$

where $J_c \in \mathbb{R}^{H(W+D(C-1)) \times HW}$. The full sensing matrix is constructed by concatenating all shifted matrices:

$$\Phi = [J_1, J_2, ..., J_C] \tag{13}$$

where $\Phi \in \mathbb{R}^{H(W+D(C-1)) \times HWC}$ represents the combined mask modulation and prism-induced spectral dispersion.

## B. Theoretical Analysis of Estimation Range

Consider a mask with the smallest coding unit spanning $e$ pixels. A horizontal shift of $e$ pixels produces a new mask:

$$\boldsymbol{M}'(:, j) = \begin{cases} \boldsymbol{M}(:, j + e) & \text{if } j < W, \\ \boldsymbol{0} & \text{if} j \geq W - e. \end{cases} \tag{14}$$

Define the mask difference:

$$\Delta \boldsymbol{M} = \boldsymbol{M}' - \boldsymbol{M}. \tag{15}$$

Assuming each coding unit in $\boldsymbol{M}$ follows a random binary distribution with $P(M(i, j) = 1) = 0.5$, the expected energy of $\boldsymbol{M}$ is:

$$\mathbb{E}(||\boldsymbol{M}||^2) = \sum_{i=1}^{H} \sum_{j=1}^{W} P(M(i, j) = 1) = 0.5HW. \tag{16}$$

For $W \gg e$, $\boldsymbol{M}'$ has approximately the same expected energy:

$$\mathbb{E}(||\boldsymbol{M}'||^2) \approx \mathbb{E}(||\boldsymbol{M}||^2). \tag{17}$$

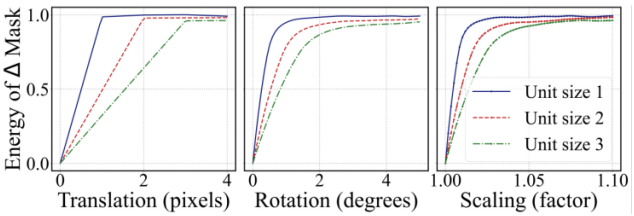

Figure 10. Mask difference energy under different perturbations.

Pixels within one coding unit are correlated, but different units are independent, hence:

$$\begin{aligned} P(\boldsymbol{M}(i, j) = 1, \boldsymbol{M}(i, j + e) = 1) \approx 0.25 \\ P(\boldsymbol{M}(i, j) = 1, \boldsymbol{M}'(i, j) = 1) \approx 0.25 \end{aligned} \tag{18}$$

and thus:

$$\mathbb{E}\langle \boldsymbol{M}' \cdot \boldsymbol{M} \rangle \approx 0.25HW. \tag{19}$$

The expected energy of the mask difference is:

$$\begin{aligned} &\mathbb{E}(||\Delta \boldsymbol{M}||^2) = \mathbb{E}(||\boldsymbol{M}' - \boldsymbol{M}||^2) \\ &= \mathbb{E}(||\boldsymbol{M}'||^2) + \mathbb{E}(||\boldsymbol{M}||^2) - 2\mathbb{E}\langle \boldsymbol{M}' \cdot \boldsymbol{M} \rangle \\ &\approx 0.5HW = \mathbb{E}(||\boldsymbol{M}||^2) \end{aligned} \tag{20}$$

This indicates that $\boldsymbol{M}'$ and $\boldsymbol{M}$ become nearly decorrelated when the shift exceeds one coding unit. Thus, shifts within one unit can be reliably estimated, while larger displacements cannot. A similar argument can apply to rotation and scaling. As shown in Fig. 10, the difference energy $\mathbb{E}[|\Delta \boldsymbol{M}|^2]$ reaches 95% of $\mathbb{E}[|\boldsymbol{M}|^2]$ under only $2^\circ$ rotation or 5% scaling. Consequently, the effective degradation estimation range is relatively small. Moreover, the practical scenario may couple different perturbation types, which can significantly limit the effective estimation range.

## C. Supplementary Results on SD-CASSI

### C.1. Affine Parameter Visualization

The estimated affine transformations for TSA-Scene 01-05 and our self-built Scene 06 are shown in Fig. 11. If the system does not undergo perturbations, a single estimation is enough and can be reused.

### C.2. Real-world Physical Validation

Our estimated degradation is at the $\mu$m-level. In real systems, such perturbations are hard to reproduce. We perform a physical validation: first capture a reference mask,

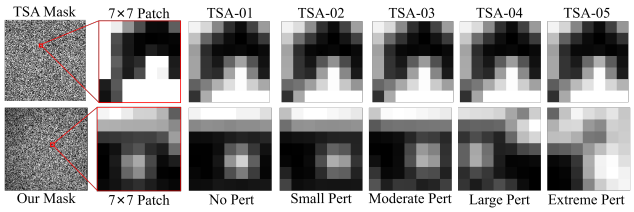


Figure 11. Mask patches after applying estimated transformation. The TSA mask appears slightly blurred due to sub-pixel warping.

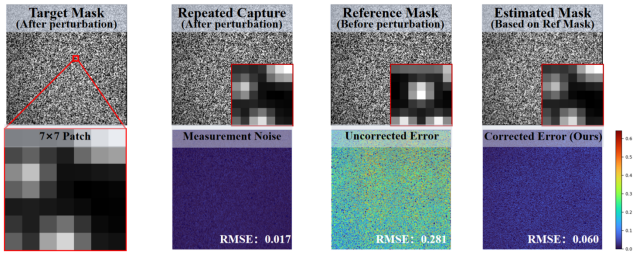


Figure 12. Real-world physical validation under perturbation.

then introduce a slight perturbation and acquire a measurement, and finally capture the after-perturbation mask twice to quantify noise. We use the measurement and the reference mask to estimate the after-perturbation mask. As shown in Fig. 12, the estimated mask is geometrically consistent with the after-perturbation mask, indicating the estimated degradation reflects real mask drift.

### C.3. Additional Ablation Experiments

**Ablation on Switching Iterations**. We analyze the impact of the switching iterations (when multi-kernel strategy transitions to single-kernel). We test 0-500 switching iterations under four settings: no perturbation, 2-pixel shift, 0.8° rotation, and 0.008 scaling, with results shown in Tab. 4. For scenes without perturbation, early switching (around 100 iterations) is already sufficient. However, under perturbations, delaying the switching point generally leads to more accurate degradation estimation. For instance, under rotation, the model is correctly aligned at 300 iterations with a PSNR of 35.37 dB, but switching at 400 iterations results in estimation failure with a PSNR of 24.78 dB. In practice, if resources allow, a later switching point typically leads to more accurate degradation estimation.

**Ablation on Preset Degradation Estimation Range**. We evaluate how different preset degradation estimation ranges affect estimation performance. Six range configurations are tested under four conditions: no perturbation, 0.5-pixel shift, 0.5 ° rotation, and 0.005 scaling, as shown in Tab. 5. Translation range is the most sensitive parameter: an overly large range (e.g., 6 pixels) causes incorrect estimation even without perturbation. Rotation and scaling tolerate moderately larger ranges (up to 4° and 0.1 scaling), but extending the range does not improve correction and may fail under small perturbations. Overall, the preset ranges should be chosen based on measured degradation estimation range rather than enlarged blindly.

Table 4. Ablation study on switching iterations across different perturbation types.

| Type | Metric | Switching Iterations. | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | 100 | 200 | 300 | 400 | 500 |
| None | PSNR | 27.22 | 37.89 | **38.21** | 38.02 | 37.92 | 37.57 |
| | SSIM | 0.718 | 0.955 | **0.956** | 0.953 | 0.952 | 0.951 |
| Trans: 2 | PSNR | 27.59 | 23.87 | 37.88 | 37.89 | **38.19** | 37.52 |
| | SSIM | 0.730 | 0.548 | 0.954 | 0.954 | 0.953 | **0.956** |
| Rot: 0.8 | PSNR | 28.35 | 28.44 | 25.02 | **35.37** | 24.78 | 35.14 |
| | SSIM | 0.774 | 0.769 | 0.614 | **0.931** | 0.597 | 0.930 |
| Scale: 0.008 | PSNR | **36.05** | 27.32 | 36.01 | 35.82 | 35.96 | 35.95 |
| | SSIM | **0.938** | 0.709 | 0.939 | 0.936 | 0.940 | 0.939 |

Table 5. Ablation study on degradation estimation ranges across different perturbation types.

| Type | Metric | Preset Calibration Ranges | | | | | |
|---|---|---|---|---|---|---|---|
| | | Trans: 4<br>Rot: 0.8<br>Scale: 0.008 | 6<br>0.8<br>0.008 | 2<br>4<br>0.008 | 2<br>5<br>0.008 | 2<br>0.8<br>0.1 | 2<br>0.8<br>0.15 |
| None | PSNR | 37.79 | 26.58 | **38.10** | 37.81 | 37.50 | 23.15 |
| | SSIM | 0.953 | 0.661 | **0.955** | 0.954 | 0.950 | 0.476 |
| Trans: 0.5 | PSNR | 35.12 | 28.11 | **35.43** | 34.68 | 34.52 | 34.83 |
| | SSIM | 0.932 | 0.753 | 0.935 | 0.925 | 0.925 | 0.927 |
| Rot: 0.5 | PSNR | **29.90** | 25.22 | 27.27 | 27.94 | 27.02 | 27.26 |
| | SSIM | **0.803** | 0.591 | 0.687 | 0.705 | 0.671 | 0.680 |
| Scale: 0.005 | PSNR | 28.42 | 25.33 | 28.77 | **28.92** | 28.89 | 27.56 |
| | SSIM | 0.753 | 0.627 | 0.764 | **0.768** | 0.762 | 0.754 |

## D. Additional Experiments for DD-CASSI

To assess the cross-system generalization of our method, we further conduct experiments on the Dual-Disperser CASSI (DD-CASSI) system.

### D.1. Imaging model of DD-CASSI

As illustrated in Figure 13, the DD-CASSI system adds a reverse disperser compared to the SD-CASSI system. The hyperspectral datacube $\boldsymbol{X} \in \mathbb{R}^{H\times W\times C}$ is first dispersed by a prism:

$$\boldsymbol{X'}(i, j + D(c), c) = \boldsymbol{X}(i, j, c) \tag{21}$$

where $\boldsymbol{X'} \in \mathbb{R}^{H\times(W+D(C-1))\times C}$. It is then modulated by the mask $\boldsymbol{M} \in \mathbb{R}^{H\times(W+D(C-1))}$:

$$\boldsymbol{X''}(:, :, c) = \boldsymbol{M} \odot \boldsymbol{X'}(:, :, c) \tag{22}$$

Subsequently, a reverse prism subsequently compensates the initial dispersion:

$$\boldsymbol{X'''}(i, j - D(c), c) = \boldsymbol{X''}(i, j, c), \tag{23}$$

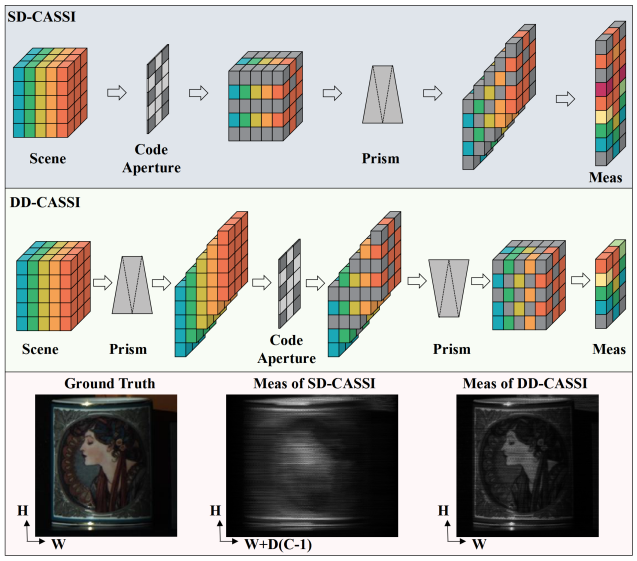


Figure 13. Comparison between SD-CASSI (top) and DD-CASSI (middle). SD-CASSI uses a single disperser, which introduces spectral shearing in the measurement. DD-CASSI employs dual dispersers to cancel this shearing. The corresponding measurements (bottom) show that DD-CASSI preserves spatial structures more clearly than SD-CASSI.

where $\boldsymbol{X}''' \in \mathbb{R}^{H\times W\times C}$. The final measurement is:

$$\boldsymbol{Y} = \sum_{c=1}^{C} \boldsymbol{X}'''(:,:,c) + \mathbf{N} \tag{24}$$

The DD-CASSI system reduces shearing artifacts and produces clearer measurements, generally enabling better reconstruction quality. However, the additional disperser increases system complexity and physical size, and the two prisms are difficult to manufacture in perfectly matched form, which can introduce additional dispersion errors.

## D.2. Reconstruction Performance

We compare SD-CASSI and DD-CASSI on KAIST Scenes 01-10. Average results are shown in Table 6. DD-CASSI consistently outperforms SD-CASSI across all scenes, with an average PSNR gain of 1.25 dB and SSIM improvement of 0.015.

Table 6. Quantitative comparison across different systems.

| System | Metric | LCTC | SK-LCTC | MK-LCTC |
|---|---|---|---|---|
| SD-CASSI | PSNR | 38.51 | 38.58 | **38.75** |
| | SSIM | 0.962 | 0.963 | **0.964** |
| DD-CASSI | PSNR | 39.79 | 39.67 | **40.00** |
| | SSIM | 0.979 | 0.976 | **0.979** |

## D.3. Robustness Evaluation

We evaluate the robustness of SD-CASSI and DD-CASSI on KAIST-scene 02 using LCTC [12]. The systems are tested under various perturbations, including translations (0–2 pixels), rotations (0–1.0°), and scalings (0–0.01), with results shown in Fig. 14. DD-CASSI achieves higher PSNR than SD-CASSI under different perturbations. The difference stems from their optical designs. In SD-CASSI, the single disperser introduces spectral shearing that is not canceled, resulting in coupled spatial–spectral encoding. Consequently, mask misalignment simultaneously distorts both domains. DD-CASSI removes the shearing through dual dispersers, so misalignment mainly affects the spectra. However, DD-CASSI still undergoes noticeable degradation. For instance, a 1-pixel shift reduces PSNR from 38.75 dB to 24.05 dB, showing degradation estimation is still essential.

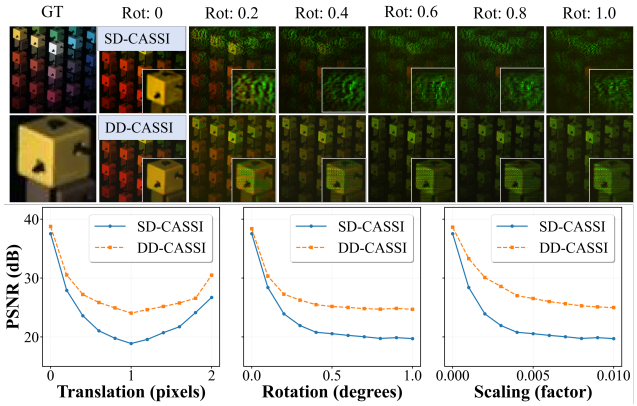


Figure 14. Robustness comparison between SD-CASSI and DD-CASSI systems under various perturbations. DD-CASSI maintains higher reconstruction quality than SD-CASSI across different levels of perturbations.

## D.4. Degradation Estimation Range of DD-CASSI

We further evaluate the degradation Estimation range of DD-CASSI under various perturbations and encoding unit sizes, with results shown in Tab. 7. Compared with the SD-CASSI results in Tab. 2, DD-CASSI enables larger single-kernel estimation ranges. For example, the maximum rotation estimation range increases from 1.6° in SD-CASSI to 4.6° in DD-CASSI. Similar improvements appear in translation and scaling. However, single-kernel degradation estimation still fails under larger perturbations, making multi-kernel estimation strategy necessary for reliable degradation estimation.

Table 7. Ablation study on degradation estimation range across different perturbation levels and encoding unit sizes.

| Method | Metric | Encoding Unit Size: 1 pixel | | | | | | | Encoding Unit Size: 2 pixels | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Trans: 0 | 1.0 | 1.5 | 0 | 0 | 0 | 0 | Trans: 0 | 4.5 | 5.0 | 0 | 0 | 0 | 0 |
| | | Rot: 0 | 0.0 | 0.0 | 0.2 | 0.4 | 0 | 0 | Rot: 0 | 0.0 | 0.0 | 4.6 | 4.8 | 0 | 0 |
| | | Scale: 0 | 0.0 | 0.0 | 0 | 0 | 0.002 | 0.004 | Scale: 0 | 0.0 | 0.0 | 0 | 0 | 0.020 | 0.022 |
| LCTC | PSNR | 40.95 | 25.61 | 28.58 | 30.82 | 27.43 | 33.97 | 31.60 | 36.45 | 28.63 | 27.90 | 23.69 | 23.91 | 25.23 | 24.77 |
| | SSIM | 0.983 | 0.716 | 0.815 | 0.874 | 0.818 | 0.925 | 0.880 | **0.968** | 0.857 | 0.834 | 0.709 | 0.705 | 0.784 | 0.773 |
| SK-LCTC | PSNR | 41.09 | **41.48** | 33.28 | 33.57 | 30.65 | 33.97 | 31.61 | 36.70 | 31.21 | 29.50 | 35.31 | 25.04 | 32.68 | 26.62 |
| | SSIM | 0.983 | **0.984** | 0.929 | 0.926 | 0.873 | 0.945 | 0.914 | **0.968** | 0.899 | 0.858 | 0.962 | 0.744 | 0.917 | 0.819 |
| MK-LCTC | PSNR | **41.43** | 41.23 | **41.53** | **37.24** | **39.02** | **39.31** | **38.02** | 36.46 | **35.29** | **36.30** | **36.37** | **35.21** | **35.88** | **35.22** |
| | SSIM | **0.984** | 0.983 | **0.984** | **0.972** | **0.894** | **0.974** | **0.972** | 0.966 | **0.961** | **0.968** | **0.965** | **0.963** | **0.964** | **0.962** |