Robust Cloud Removal from Optical Satellite Images Using Synthetic Aperture Radar and Multimodal Embedding Prior

Taishin Miura*, Shunsuke Ono†, and Ryo Matsuoka‡
*‡The University of Kitakyushu, Japan
†Institute of Science Tokyo, Japan
E-mail: ‡r-matsuoka@kitakyu-u.ac.jp

***Abstract*—Cloud contamination and noise degradation, including Gaussian and impulsive noise, significantly limit the utility of optical satellite imagery by obscuring surface features and degrading image quality. We propose a robust restoration method that integrates robust principal component analysis (RPCA) with a multimodal regularization based on locally linear embedding (LLE), using synthetic aperture radar (SAR) with speckle noise suppressed by RPCA as structural references. These references then guide the decomposition of optical images into low-rank backgrounds and sparse components representing clouds and impulsive noise. The LLE-based prior enforces local structural consistency and spatial smoothness between the optical and SAR modalities. The resulting optimization problem, incorporating nuclear norm and sparsity constraints, is efficiently solved using the plug-and-play alternating direction method of multipliers (PnP-ADMM) framework. Experimental results show that the proposed method effectively removes clouds and noise, preserves structural details, and improves the visibility and interpretability of the restored images.**

I. Introduction

Optical satellite imagery is essential for applications such as environmental monitoring, precision agriculture, and disaster response. However, its utility is frequently hindered by cloud cover, which obscures surface features and introduces uncertainty in downstream tasks like land cover classification and urban planning. This issue is particularly severe in tropical and subtropical regions, where persistent cloudiness is common.

Synthetic aperture radar (SAR) provides a complementary imaging modality that is unaffected by weather and lighting conditions. By capturing backscattered signals, SAR conveys structural information about the Earth's surface. Integrating SAR with optical imagery offers a promising solution for cloud removal. Nevertheless, the fusion of these modalities remains challenging due to their fundamentally different characteristics: optical images contain rich spectral-spatial details, while SAR data represent structural intensity in a different signal domain. Robust principal component analysis (RPCA) has been widely used for cloud removal [1]–[7], decomposing multi-temporal optical image sequences into low-rank backgrounds and sparse cloud components. Although effective, RPCA-based methods often require a large number of frames, which are not always available due to satellite revisit cycles or persistent cloud cover. Furthermore, in scenarios where clouds remain or overlap across frames, RPCA struggles to separate them accurately, resulting in incomplete restorations. These challenges are exacerbated by additional degradations, such as Gaussian and impulsive noise, which are common in optical imagery and further compromise restoration performance.

In recent years, deep learning, leveraging its powerful nonlinear representation capability, has been prevalently employed for cloud removal in remote sensing images [8]–[12]. However, deep learning-based methods typically require large training data sets and may not adapt well to test data that differ from the training data in terms of spatial resolution, number of spectral bands, or presence or absence of noise.

To address these limitations, we propose a robust cloud and noise removal method that incorporates SAR images as structural guidance. The approach combines RPCA with a multimodal regularizer based on locally linear embedding (LLE), which enforces local geometric consistency between optical and SAR modalities. To enhance the reliability of SAR guidance, speckle noise is first suppressed via RPCA. The entire framework is formulated as a constrained optimization problem that integrates RPCA decomposition, LLE-based regularization, and a reconstruction constraint that implicitly models Gaussian noise. The problem is efficiently solved using the plug-and-play alternating direction method of multipliers (PnP-ADMM) [13]–[17], allowing flexible incorporation of the LLE prior as a denoising operator. Experimental results on simulated cloud-contaminated imagery demonstrate that the proposed method achieves superior cloud removal and structural preservation compared to existing approaches, both quantitatively and qualitatively, even with fewer input frames.

The rest of this paper is organized as follows. Section II introduces the mathematical preliminaries. Section III presents the proposed cloud removal framework. Section IV reports experimental results and comparisons. Finally, Section V concludes the paper.

II. Preliminaries

Vectors and matrices are denoted by bold letters. $\mathbb{R}^N$ and $\mathbb{R}^{N\times M}$ denote real-valued vectors and matrices, respectively.

979-8-3315-7206-8/25/$31.00

A. Plug-and-Play Alternating Direction Method of Multipliers (PnP-ADMM)

ADMM [18]–[20] is a proximal splitting method for solving convex problems of the form:

$$\min_{\mathbf{x}\in\mathbb{R}^{N_1},\,\mathbf{z}\in\mathbb{R}^{N_2}} F(\mathbf{x}) + G(\mathbf{z}) \quad \text{subject to} \quad \mathbf{z} = \mathbf{L}\mathbf{x}, \tag{1}$$

where F and G are proximable functions, and $\mathbf{L} \in \mathbb{R}^{N_2 \times N_1}$ has full column rank.

Given initial values $\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)}$, $\mathbf{b}^{(0)}$, and $\rho > 0$, the iterations proceed as:

$$\begin{cases} \mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} \left\{ F(\mathbf{x}) + \dfrac{\rho}{2}\|\mathbf{z}^{(t)} - \mathbf{L}\mathbf{x} - \mathbf{b}^{(t)}\|_2^2 \right\}, \\ \mathbf{z}^{(t+1)} = \arg\min_{\mathbf{z}} \left\{ G(\mathbf{z}) + \dfrac{\rho}{2}\|\mathbf{z} - \mathbf{L}\mathbf{x}^{(t+1)} - \mathbf{b}^{(t)}\|_2^2 \right\}, \\ \qquad = \mathrm{prox}_{1/\rho G}(\mathbf{L}\mathbf{x}^{(t+1)} + \mathbf{b}^{(t)}), \\ \mathbf{b}^{(t+1)} = \mathbf{b}^{(t)} + \mathbf{L}\mathbf{x}^{(t+1)} - \mathbf{z}^{(t+1)}, \end{cases} \tag{2}$$

where the superscript (t) denotes the iteration number. The $\mathbf{z}$-update employs the proximity operator[1] of $G(\cdot)$, and the iterates converge efficiently.

In PnP-ADMM [13]–[17], the sub-problem associated with $\mathbf{z}$ (assuming that $\mathbf{L}$ is the identity matrix) is replaced with a noise removal algorithm, commonly a denoising operator that assumes additive white Gaussian noise (AWGN) with standard deviation σ. This update is given by:

$$\mathbf{z}^{(t+1)} = \mathcal{D}_\sigma\left(\mathbf{x}^{(t+1)} + \mathbf{b}^{(t)}\right), \tag{4}$$

where $\mathcal{D}_\sigma$ denotes the corresponding Gaussian denoiser.

B. Locally Linear Embedding

LLE is a manifold learning technique that assumes each high-dimensional data point can be approximated as a linear combination of its nearest neighbors. Let $\{\mathbf{x}_i\}_{i=1}^N$ be a set of high-dimensional vectors. For each point $\mathbf{x}_i$, its k-nearest neighbors $\mathcal{N}(i)$ are first identified using a suitable distance metric. Then, reconstruction weights $w_{i,j}$ are computed to best approximate $\mathbf{x}_i$ as a linear combination of its neighbors:

$$\mathbf{x}_i \approx \sum_{j\in\mathcal{N}(i)} w_{i,j}\mathbf{x}_j, \quad \text{s.t.} \quad \sum_{j\in\mathcal{N}(i)} w_{i,j} = 1, \;\; w_{i,j} \geq 0. \tag{5}$$

Finally, a low-dimensional embedding $\{\mathbf{y}_i\}_{i=1}^N$ is obtained by requiring that each $\mathbf{y}_i$ is similarly reconstructed using the same weights:

$$\mathbf{y}_i \approx \sum_{j\in\mathcal{N}(i)} w_{i,j}\mathbf{y}_j.$$

[1]The proximity operator [21] of a function f is defined as:

$$\mathrm{prox}_{\gamma f}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathbb{R}^N} f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{x} - \mathbf{y}\|^2. \tag{3}$$

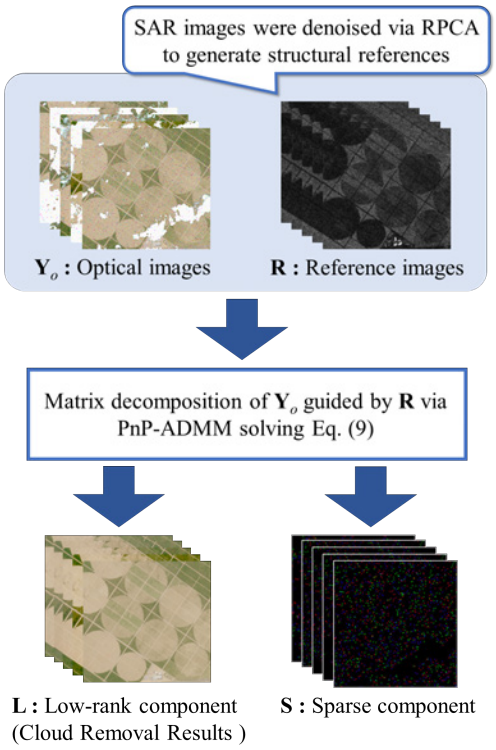

Fig. 1. Overview of our cloud removal framework. For visualization, the sparse component was enhanced using brightness scaling and sharpness adjustment to improve perceptibility.

III. Proposed Method

Figure 1 illustrates the overall workflow of the proposed method, which restores cloud-contaminated optical satellite images using structural guidance from SAR data. The framework consists of the following two main steps:

(i) Reference generation: Multi-temporal SAR images are denoised using RPCA [1] to suppress speckle noise, yielding clean reference images that preserve structural details.

(ii) Cloud removal: Using the denoised SAR images as structural references, the method decomposes optical images into a low-rank background and sparse component via a constrained optimization problem. A multimodal LLE prior (see Section III-A) enforces local geometric consistency, and the optimization is solved via the PnP-ADMM framework (see Section III-B).

A. Multimodal LLE Prior for Optical and SAR Images

1) Prior Assumptions: Consider two co-registered satellite images of the same geographic region: an optical RGB image and a SAR image. The optical image is represented as three single-channel vectors $\mathbf{u}_o^{(c)} \in \mathbb{R}^N$ for $c \in \{r, g, b\}$, where $u_{o,i}^{(c)}$ denotes the pixel value at location i in channel c. The SAR image is represented as a single-channel vector $\mathbf{u}_s \in \mathbb{R}^N$, with $u_{s,i}$ as the pixel intensity at the same location i.

For each pixel i, we form paired data $\left\{u_{o,i}^{(c)}, u_{s,i}\right\}$, preserving spatial and spectral relationships across modalities. To en-

force structural consistency, we assume that the neighborhood structure $\mathcal{N}(i)$ and the reconstruction weights $\omega_{i,j}$ are shared between the optical and SAR domains. That is,

$$u_{o,i}^{(c)} \approx \sum_{j\in\mathcal{N}(i)} \omega_{i,j}\, u_{o,j}^{(c)}, \quad u_{s,i} \approx \sum_{j\in\mathcal{N}(i)} \omega_{i,j}\, u_{s,j}.$$

This shared reconstruction scheme allows the model to capture local geometric consistency across modalities, which is critical for our multimodal regularization framework.

2) Matrix Representation of Local Reconstruction: Let each image layer be indexed by c (e.g., $c \in \{r, g, b, s\}$ for the RGB and SAR channels), and assume that each image consists of N pixels. For each spatial location i, we define a local region (patch) containing L neighboring pixels. By shifting the ℓ-th neighbor in each patch to the center position i, we construct shifted image vectors $\mathbf{u}_\ell^{(c)} \in \mathbb{R}^N$ for $\ell = 1, \ldots, L$.

To express the LLE-based reconstruction in matrix form, we introduce the following definitions:

Construction of $\mathbf{U}_c$: We define the matrix $\mathbf{U}_c \in \mathbb{R}^{N\times LN}$ by horizontally concatenating L diagonal matrices, each constructed from a shifted vector: $\mathbf{U}_c := [\mathrm{diag}(\mathbf{u}_1^{(c)}), \ldots, \mathrm{diag}(\mathbf{u}_L^{(c)})]$, where $\mathrm{diag}(\mathbf{u}_\ell^{(c)})$ denotes an $N\times N$ diagonal matrix with the elements of $\mathbf{u}_\ell^{(c)}$ on the diagonal.

Weight vector $\mathbf{w}$: We define the full weight vector as $\mathbf{w} := [\mathbf{w}_1^\top, \ldots, \mathbf{w}_L^\top]^\top \in \mathbb{R}^{LN}$, where each subvector $\mathbf{w}_\ell \in \mathbb{R}^N$ contains the spatially varying weights associated with the ℓ-th neighbor. The i-th entry of $\mathbf{w}_\ell$ indicates the contribution of the ℓ-th neighbor to reconstructing the i-th pixel.

Reconstruction model: The reconstructed image $\mathbf{u}^{(c)} \in \mathbb{R}^N$ is obtained by applying the weights to the shifted patches: $\mathbf{u}^{(c)} \approx \mathbf{U}_c\mathbf{w} = \sum_{\ell=1}^{L} \mathrm{diag}(\mathbf{u}_\ell^{(c)})\mathbf{w}_\ell$. This formulation linearly combines the neighboring pixel values using spatially adaptive weights, in accordance with the LLE assumption.

3) Multimodal LLE Weight Optimization with TV Regularization: To compute reconstruction weights that are both spatially smooth and consistent across modalities, we formulate the following optimization problem[2]:

$$\begin{aligned} &\min_{\mathbf{w}} \sum_{c=1}^{4} \left\{ \left\|\mathbf{u}^{(c)} - \mathbf{U}_c\mathbf{w}\right\|_2^2 + \lambda\, \mathrm{TV}(\mathbf{U}_c\mathbf{w}) \right\}, \\ &\text{s.t. } \mathbf{0} \le \mathbf{w}_\ell \le \mathbf{1} \ \ \forall_\ell, \quad \sum_{\ell=1}^{L} \mathbf{w}_\ell = \mathbf{1}, \end{aligned} \tag{6}$$

where $\lambda > 0$ controls the strength of the TV regularization. The first term ensures reconstruction fidelity based on LLE, while the second promotes spatial smoothness.

The TV term is defined using discrete vertical and horizontal gradient operators $\mathbf{D}_v$ and $\mathbf{D}_h \in \mathbb{R}^{N\times N}$ as:

$$\mathrm{TV}(\mathbf{x}) = \sum_{i=1}^{N} \sqrt{(\mathbf{D}_v\mathbf{x})_i^2 + (\mathbf{D}_h\mathbf{x})_i^2}. \tag{7}$$

[2]We build upon the LLE-based, TV-regularized weight optimization in [22], [23] and extend it to SAR-guided multimodal regularization for optical image restoration.

This formulation minimizes abrupt intensity variations while preserving local structure.

The constraints in (6) follow the standard LLE assumptions: weights are non-negative and sum to one across each neighborhood.

We solve the proposed problem using the ADMM algorithm. In the cloud removal framework described later, this weight optimization is incorporated into the regularization term $\mathcal{R}_{\mathrm{LLE},\lambda}$ and functions as a denoising prior to encourage local structural alignment between optical and SAR images.

B. Problem Formulation and Optimization

This study aims to restore cloud-contaminated optical satellite images by leveraging multi-temporal optical–SAR image pairs using RPCA-based decomposition and multimodal regularization.

Given K co-registered pairs of RGB optical images and single-channel SAR images[3], we define the optical observation matrix $\mathbf{Y}_o \in \mathbb{R}^{3N\times K}$ and SAR observation matrix $\mathbf{Y}_s \in \mathbb{R}^{N\times K}$ as: $\mathbf{Y}_o = [\mathbf{u}_o^{(1)}, \ldots, \mathbf{u}_o^{(K)}]$, $\mathbf{Y}_s = [\mathbf{u}_s^{(1)}, \ldots, \mathbf{u}_s^{(K)}]$, where each optical vector $\mathbf{u}_o := [\mathbf{u}_{o,r}^\top\, \mathbf{u}_{o,g}^\top\, \mathbf{u}_{o,b}^\top]^\top \in \mathbb{R}^{3N}$ is formed by stacking the RGB channels at time k. Each SAR vector $\mathbf{u}_s^{(k)} \in \mathbb{R}^N$ captures the spatial backscatter profile at time k, providing structural information complementary to the optical observations.

To suppress speckle noise and extract reliable structural information, RPCA [1] is applied to $\mathbf{Y}_s$, yielding a low-rank estimate $\mathbf{R} \in \mathbb{R}^{N\times K}$ of clean SAR images. This structural reference is used to guide the decomposition of $\mathbf{Y}_o$ into background and sparse components via multimodal LLE-based regularization.

The restoration task is formulated as the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{L},\mathbf{S}} \quad & \mathcal{R}_{\mathrm{LLE},\lambda}(\mathbf{L},\mathbf{R}) + \alpha\|\mathbf{L}\|_* + \beta\|\mathbf{\Omega}_{\epsilon,1}\mathbf{S}\|_1 \\ \text{s.t.} \quad & \|\mathbf{M}\mathbf{Y}_o - \mathbf{M}\mathbf{L} - \mathbf{M}\mathbf{S}\|_F \le \delta, \end{aligned} \tag{8}$$

where $\mathbf{L}$ and $\mathbf{S}$ denote the low-rank and sparse components of the optical images, respectively, and α and β are balancing weights of the three terms. The regularization term $\mathcal{R}_{\mathrm{LLE},\lambda}(\mathbf{L},\mathbf{R})$ imposes LLE constraints to ensure local structural similarity between the optical and SAR modalities.

The nuclear norm penalty $\|\mathbf{L}\|_*$ promotes low-rankness[4], capturing background structures. The ℓ_1-norm term $\|\mathbf{\Omega}_{\epsilon,1}\mathbf{S}\|_1$ enforces sparsity to isolate clouds, where $\mathbf{\Omega}_{\epsilon,1}$ is a weighting matrix assigning small values ϵ to cloud regions and values near 1 elsewhere, modulating the sparsity constraint based on cloud presence. The constraint $\|\mathbf{M}\mathbf{Y}_o - \mathbf{M}\mathbf{L} - \mathbf{M}\mathbf{S}\|_F \le \delta$ ensures reconstruction fidelity within a tolerance δ. The parameter δ sets the allowable reconstruction error, accommodating Gaussian noise and other degradations.

[3]Optical and SAR data are captured by different satellites on independent schedules; their acquisition times typically differ by a few days. Image pairs are formed within a short temporal window to preserve spatial and structural consistency.

[4]$\|\cdot\|_*$ is the nuclear norm, defined as the sum of all singular values of the argument.

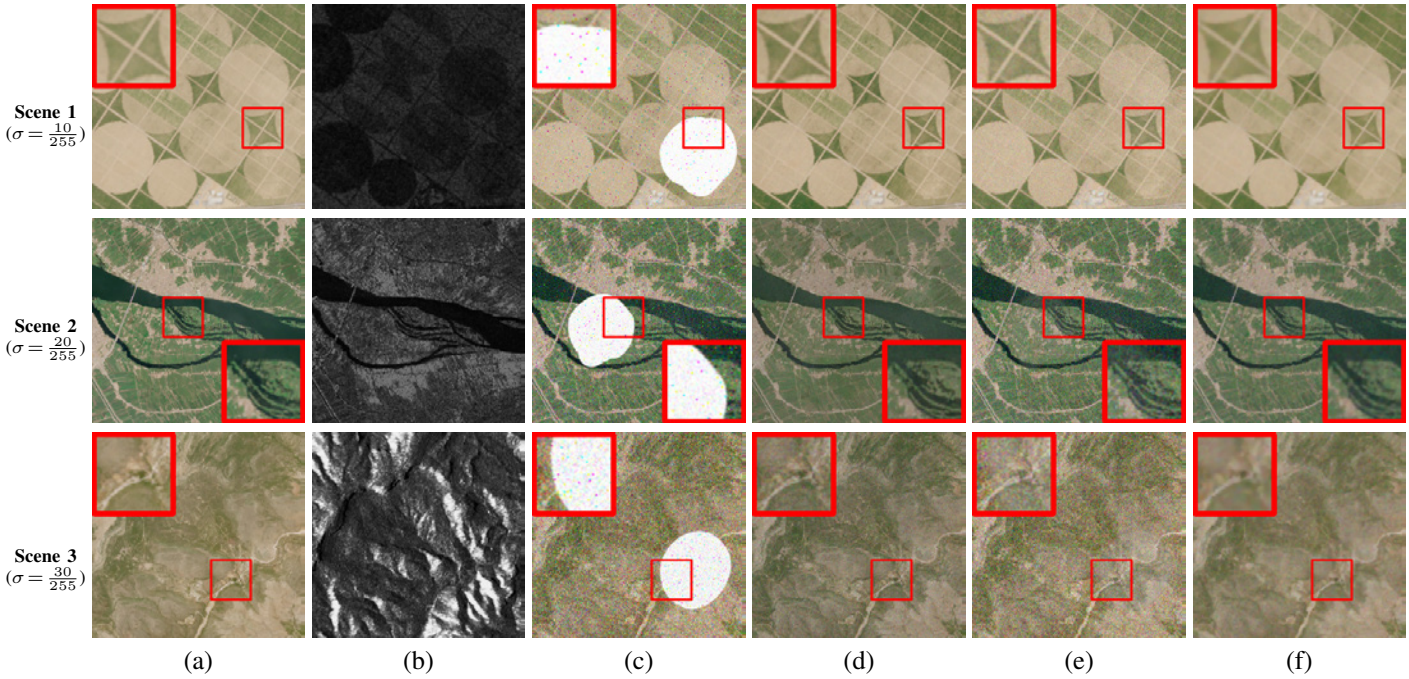

Fig. 2. Results of cloud removal on three simulated scenes. (a) Ground truth, (b) Reference SAR, (c) Input, (d) RPCA [1], (e) DRPCA [3], and (f) Our method.

To simplify the constraint, we introduce an indicator function over a Frobenius-norm ball and rewrite the problem as:

$$\begin{aligned}\min_{\mathbf{L},\mathbf{S}} \quad & \mathcal{R}_{\mathrm{LLE},\lambda}(\mathbf{L},\mathbf{R}) + \alpha\|\mathbf{L}\|_* + \beta\|\mathbf{\Omega}_{\epsilon,1}\mathbf{S}\|_1 \\ & + \iota_{\mathcal{B}^F_{\mathbf{M}\mathbf{Y}_o,\delta}}(\mathbf{M}\mathbf{L}+\mathbf{M}\mathbf{S}),\end{aligned} \tag{9}$$

where $\iota_{\mathcal{B}}$ is the indicator function of the set

$$\mathcal{B}^F_{\mathbf{M}\mathbf{Y}_o,\delta} = \{\mathbf{X} \in \mathbb{R}^{N\times K} \mid \|\mathbf{M}\mathbf{Y}_o - \mathbf{X}\|_F \le \delta\}.$$

To enable efficient optimization, we introduce auxiliary variables $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}_3$, and $\mathbf{Z}_4$, we rewrite the minimization problem into the following equivalent expression:

$$\begin{aligned}\min_{\mathbf{L},\mathbf{S},\mathbf{Z}_i} \quad & \mathcal{R}_{\mathrm{LLE},\lambda}(\mathbf{Z}_1,\mathbf{R}) + \alpha\|\mathbf{Z}_2\|_* + \beta\|\mathbf{Z}_3\|_1 + \iota_{\mathcal{B}^F_{\mathbf{M}\mathbf{Y}_o,\delta}}(\mathbf{Z}_4) \\ \text{s.t.} \quad & \mathbf{Z}_1 = \mathbf{L}, \mathbf{Z}_2 = \mathbf{L}, \mathbf{Z}_3 = \mathbf{\Omega}_{\epsilon,1}\mathbf{S}, \mathbf{Z}_4 = \mathbf{M}\mathbf{L}+\mathbf{M}\mathbf{S}.\end{aligned} \tag{10}$$

The minimization problem (10) is solved using the PnP-ADMM [13]–[16].

IV. Experiments

The proposed method was evaluated on three satellite image scenes using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [24], where higher values indicate better reconstruction quality. We compared our approach with two baseline methods: RPCA [1] and DRPCA [3]. All method parameters were manually tuned to yield visually satisfactory results and ensure fair comparison.

To simulate realistic cloud contamination, cloud masks were generated by thresholding bright regions from real cloud textures, followed by morphological dilation. The corresponding SAR images were used to guide the restoration process[5]. To evaluate robustness, AWGN with standard deviations $\sigma = 10/255$, $20/255$, and $30/255$ was added to the optical images, along with salt-and-pepper noise at a fixed probability of $p = 0.01$.

[5]All optical and SAR images were normalized to the $[0, 1]$ range by dividing by their respective maximum values.

TABLE I
PSNR/SSIM on simulated cloud-contaminated images. The parameter σ denotes the standard deviation of AWGN; salt-and-pepper noise was fixed at $p = 0.01$.

Scene	σ	RPCA [1]		DRPCA [3]		Ours	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	10/255	28.84	0.9061	29.16	0.6433	**31.65**	**0.9266**
	20/255	25.00	0.7882	23.41	0.3787	**29.60**	**0.8208**
	30/255	22.21	0.6748	20.02	0.2467	**24.87**	**0.7410**
2	10/255	26.49	0.8736	28.97	0.7473	**32.29**	**0.8847**
	20/255	24.19	0.7876	23.21	0.4635	**29.94**	**0.8125**
	30/255	21.30	0.6714	19.83	0.2994	**26.48**	**0.7393**
3	10/255	26.27	0.8213	27.24	0.7700	**29.68**	**0.8854**
	20/255	22.78	0.7418	22.51	0.5691	**26.02**	**0.8102**
	30/255	21.92	0.6376	19.38	0.4210	**24.19**	**0.7172**

The proposed method was applied using 5 paired optical–SAR frames, whereas RPCA and DRPCA were applied to 15 optical-only frames. Despite using fewer frames, our method consistently outperformed both baselines, demonstrating the effectiveness of multimodal structural guidance provided by SAR data.

Table I reports the PSNR and SSIM scores for three scenes under varying noise levels. The proposed method consistently outperforms RPCA and DRPCA, particularly under severe noise ($\sigma = 30/255$), where it outperforms DRPCA by approximately 6dB in PSNR and 0.4 in SSIM over DRPCA in Scene 2.

Figure 2 shows qualitative results for Scenes 1 and 2. The

proposed method produces visually coherent reconstructions, suppressing noise and cloud artifacts while preserving fine structures. In contrast, DRPCA exhibits residual noise in non-cloud regions due to its lack of noise modeling and selective sparsity. Although RPCA achieves moderate denoising, it leaves visible artifacts and tends to significantly alter the background colors. Both baselines show boundary distortions near cloud edges, which are mitigated in our approach by leveraging SAR-guided priors and local regularization. Our method further benefits from a reconstruction constraint bounded by the Frobenius norm, which implicitly models Gaussian noise and enables robust separation of noise and cloud components.

These results validate the effectiveness of combining multimodal guidance with locally adaptive regularization under challenging conditions.

V. Conclusion

The proposed restoration framework for cloud-contaminated optical satellite imagery leverages SAR images as structural priors. It couples RPCA-based decomposition for separating clouds and impulsive noise with the LLE-based multimodal regularizer, which enforces local geometric consistency and spatial smoothness. The resulting optimization is efficiently solved via the PnP-ADMM framework. Experimental results demonstrated that the proposed approach effectively restores optical images by removing clouds as well as Gaussian and impulsive noise, while preserving fine structural details. It consistently outperformed conventional RPCA-based methods in both quantitative metrics and visual quality, confirming robustness across diverse degradation conditions.

Future work will focus on accelerating the optimization, extending the framework to jointly optimize optical and SAR modalities, and validating performance on real-world cloudy scenes with fewer frames and more complex structures (e.g., urban areas).

ACKNOWLEDGMENT

This work was partially supported by the FAIS Research and Development Project Support Program, Satellite Data Utilization Demonstration Project and JSPS KAKENHI Grant Number 24K15075.

References

[1] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[2] F. Wen, Y. Zhang, Z. Gao, and X. Ling, "Two-pass robust component analysis for cloud removal in satellite image sequence," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1090–1094, 2018. DOI: 10.1109/LGRS.2018.2829028.

[3] Y. Zhang, F. Wen, Z. Gao, and X. Ling, "A coarse-to-fine framework for cloud removal in remote sensing image sequence," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5963–5974, 2019. DOI: 10.1109/TGRS.2019.2903594.

[4] S. Imran, M. Tahir, Z. Khalid, and M. Uppal, "A deep unfolded prior-aided rpca network for cloud removal," *IEEE Signal Process. Lett.*, vol. 29, pp. 2048–2052, 2022.

[5] T.-Y. Ji, D. Chu, X.-L. Zhao, and D. Hong, "A unified framework of cloud detection and removal based on low-rank and group sparse regularizations for multitemporal multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022. DOI: 10.1109/TGRS.2022.3152630.

[6] L.-Y. Li, T.-Z. Huang, Y.-B. Zheng, *et al.*, "Thick cloud removal for multitemporal remote sensing images: When tensor ring decomposition meets gradient domain fidelity," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023. DOI: 10.1109/TGRS.2023.3276414.

[7] Y. Chen, M. Chen, W. He, J. Zeng, M. Huang, and Y.-B. Zheng, "Thick cloud removal in multitemporal remote sensing images via low-rank regularized self-supervised network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–13, 2024. DOI: 10.1109/TGRS.2024.3358493.

[8] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, Aug. 2020. DOI: 10.1016/j.isprsjprs.2020.06.007.

[9] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018. DOI: 10.1109/TGRS.2018.2819190.

[10] Q. Zhang, Q. Yuan, J. Li, Z. Li, H. Shen, and L. Zhang, "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 148–160, Apr. 2020. DOI: 10.1016/j.isprsjprs.2020.02.008.

[11] S. Ji, P. Dai, M. Lu, and Y. Zhang, "Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 732–748, Jan. 2021. DOI: 10.1109/TGRS.2020.2991544.

[12] B. Jiang *et al.*, "A deep-learning reconstruction method for remote sensing images with large thick cloud cover," *International Journal of Applied Earth Observation and Geoinformation*, vol. 115, p. 103 079, Dec. 2022. DOI: 10.1016/j.jag.2022.103079.

[13] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process. (GlobalSIP)*, IEEE, 2013, pp. 945–948.

[14] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play admm for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, 2016.

[15] Z. Meng, X. Yuan, and S. Jalali, "Deep unfolding for snapshot compressive imaging," *Int. J. Comput. Vis.*, vol. 131, no. 11, pp. 2933–2958, 2023.

[16] H. Yamamoto, S. Anami, and R. Matsuoka, "Optimizing dynamic mode decomposition for video denoising via plug-and-play alternating direction method of multipliers," *Signals*, vol. 5, no. 2, pp. 202–215, 2024, ISSN: 2624-6120. DOI: 10.3390/signals5020011.

[17] A. Emoto and R. Matsuoka, "Unsupervised anomaly detection in hyperspectral imaging: Integrating tensor robust principal component analysis with autoencoding adversarial networks," *IEEE Access*, vol. 13, pp. 21 422–21 433, 2025. DOI: 10.1109/ACCESS.2025.3534981.

[18] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[20] R. Matsuoka and M. Okuda, "Beyond staircasing effect: Robust image smoothing via ℓ_0 gradient minimization and novel gradient constraints," *Signals*, vol. 4, no. 4, pp. 669–686, 2023, ISSN: 2624-6120. DOI: 10.3390/signals4040037.

[21] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris*, vol. 255, pp. 2897–2899, 1962.

[22] R. Matsuoka, S. Ono, and M. Okuda, "High dynamic range image generation based on convolutional weight optimization robust to mixed noise removal," in *Proc. APSIPA Annu. Summit Conf. (APSIPA ASC)*, 2018, pp. 1066–1070. DOI: 10.23919/APSIPA.2018.8659491.

[23] R. Matsuoka, S. Ono, and M. Okuda, "Transformed-domain robust multiple-exposure blending with huber loss," *IEEE Access*, vol. 7, pp. 162 282–162 296, 2019. DOI: 10.1109/ACCESS.2019.2951817.

[24] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.