Supplementary Material for CPIT

Yuli Sun, Lin Lei, Gangyao Kuang

I. SPECTRAL DOMAIN ANALYSIS OF CHANGES

We exploit the spectral properties of changes on different graph models. In graph signal processing (GSP), the graph Laplacian matrix $\mathbf{L}$ can be factorized as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, where $\mathbf{U}$ represents an orthonormal matrix whose columns consist of the eigenvectors $\mathbf{u}_k$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues λ_k. The graph Fourier transform (GFT) and inverse GFT (IGFT) are defined as [1], [2]:

$$\begin{aligned} \hat{\mathbf{f}} &= \mathrm{GFT}\left(\mathbf{f}\right) = \mathbf{U}^{-1}\mathbf{f} \\ \mathbf{f} &= \mathrm{IGFT}(\hat{\mathbf{f}}) = \mathbf{U}\hat{\mathbf{f}} \end{aligned} \tag{S.1}$$

Arranging the eigenvalues of the graph Laplacian matrix $\mathbf{L}$ in ascending order as $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N_S}$, the smallest eigenvalue λ_1 corresponds to the lowest frequency component, while the largest eigenvalue λ_{N_S} represents the highest frequency component [1], [2].

A. Spectral properties of images on KNN and KFN graphs

Our previous study [3] approached HeCD as a GSP problem by treating feature matrix as graph signal defined on graph. Within this framework, we analyze the spectral properties of different images on similarity-based KNN graph and dissimilarity based KFN graph. Specifically,

(i) on KNN graph G_n^{t1}, the transformed signal $\mathbf{Y}'$ and the original signal $\mathbf{X}$ exhibit approximate low-pass characteristics; likewise, $\mathbf{X}'$ and $\mathbf{Y}$ show similar low-pass behavior on KNN graph G_n^{t2} [3], [4];

(ii) on KFN graph G_f^{t1}, the transformed signal $\mathbf{Y}'$ and the original signal $\mathbf{X}$ exhibit similar high-frequency characteristics; likewise, $\mathbf{X}'$ and $\mathbf{Y}$ show similar high-frequency characteristics on KFN graph G_f^{t2} [3], [4].

In the CPIT model, we use regularization terms of $\mathcal{R}\left(\mathbf{X},\mathbf{Y}'\right)$ and $\mathcal{R}\left(\mathbf{Y},\mathbf{X}'\right)$ to satisfy these spectral properties of transformed images, e.g., $\mathcal{R}\left(\mathbf{X},\mathbf{Y}'\right)$ requires the transformed image $\tilde{\mathbf{Y}}'$ to have the same similarity relationship and dissimilarity relationships as the original image $\tilde{\mathbf{X}}$ on the KNN graph G_{hn}^{t1} and KFN graph G_{hf}^{t1}, respectively.

B. Spectral properties of change signals on KNN and KFN graphs

Then, we analyze the properties of change signals ($\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$) on KNN and KFN graphs.

(i) Change signals on the KNN graphs: In the KNN graph G_n^{t1}, each connected pair $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ belongs to the same object class, i.e., $P_i^{t1} = P_j^{t1}$ for $(i,j) \in E_n^{t1}$. However, for the connected vertices $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$ in the same graph G_n^{t1}, they may correspond to either the same or different object classes, i.e., $P_i^{t1} = P_j^{t1}$ or $P_i^{t1} \neq P_j^{t1}$, for $(i,j) \in E_n^{t1}$. According to (4a) and (4c), we may have $C_i^{t1\to t2} = C_j^{t1\to t2}$ or $C_i^{t1\to t2} \neq C_j^{t1\to t2}$. Consequently, the change signals $\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$ exhibit both low- and high-frequency components on the KNN graph G_n^{t1}. Similarly, on the KNN graph G_n^{t2}, the change signals also contain both low- and high-frequency components.

(ii) Change signals on the KFN graphs: In the KFN graph G_f^{t1}, each connected pair $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{X}}_j$ belongs to different object classes, i.e., $P_i^{t1} \neq P_j^{t1}$ for $(i,j) \in E_f^{t1}$. However, for the connected nodes $\tilde{\mathbf{Y}}_i$ and $\tilde{\mathbf{Y}}_j$ in the same graph G_f^{t1}, they may correspond to either the same or different object classes, i.e., $P_i^{t1} = P_j^{t1}$ or $P_i^{t1} \neq P_j^{t1}$, for $(i,j) \in E_f^{t1}$. According to (4b) and (4d), we may have $C_i^{t1\to t2} = C_j^{t1\to t2}$ or $C_i^{t1\to t2} \neq C_j^{t1\to t2}$. Consequently, the change signals $\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$ also exhibit both low- and high-frequency components on the KFN graph G_f^{t1}. Similarly, on the KFN graph G_f^{t2}, the change signals also contain both low- and high-frequency components.

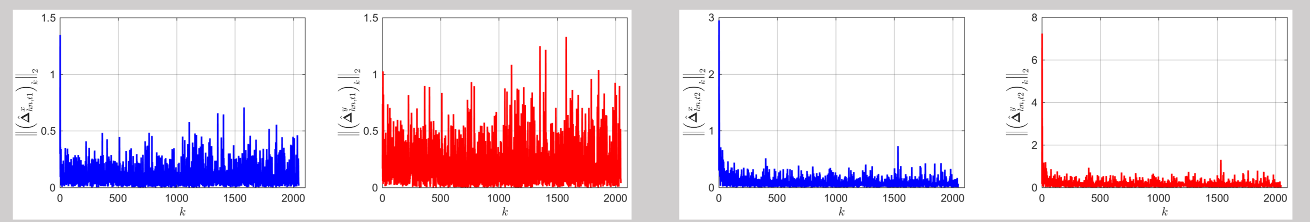
From the above analysis, we can observe that using only KNN or KFN graphs, we cannot effectively obtain the spectral properties of the change signals, i.e., KNN or KFN graphs cannot discriminate between the low- and high-frequency components of change signals. In Fig. 1(a)-(d), we plot the spectral curves of $\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$ on different graphs, where the spectra are defined as $\hat{\mathbf{\Delta}}_{hn,t1}^x = \mathbf{U}_{hn,t1}^{-1}(\mathbf{\Delta}^x)^T$, $\hat{\mathbf{\Delta}}_{hn,t2}^x = \mathbf{U}_{hn,t2}^{-1}(\mathbf{\Delta}^x)^T$, $\hat{\mathbf{\Delta}}_{hf,t1}^x = \mathbf{U}_{hf,t1}^{-1}(\mathbf{\Delta}^x)^T$, and $\hat{\mathbf{\Delta}}_{hf,t2}^x = \mathbf{U}_{hf,t2}^{-1}(\mathbf{\Delta}^x)^T$, where the orthonormal matrices of $\mathbf{U}_{hn,t1}$, $\mathbf{U}_{hn,t2}$, $\mathbf{U}_{hf,t1}$, and $\mathbf{U}_{hf,t2}$ are calculated from the Laplacian matrices of graphs G_{hn}^{t1}, G_{hn}^{t2}, G_{hf}^{t1}, and G_{hf}^{t2}, respectively. As illustrated in Fig. 1(a)-(d), both similarity-based KNN graph models (G_{hn}^{t1} and G_{hn}^{t2}) and dissimilarity-based KFN graph models (G_{hf}^{t1} and G_{hf}^{t2}) fail to effectively separate the frequency components of the change signals ($\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$).

C. Spectral properties of change signals on weighting matrix induced graphs

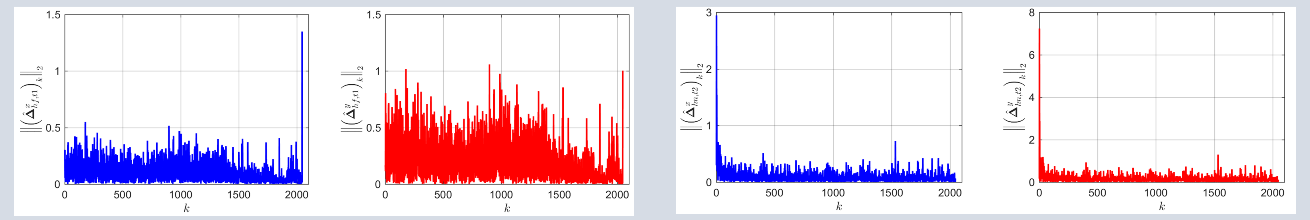
Next, we analyze the characteristics of the change signals on graphs with weighting matrices $\mathbf{W}^{sc}$ (12) and $\mathbf{W}^{si}$ (21).

- For the weighting matrix $\mathbf{W}^{sc}$ induced graph: Based on the construction process of change semantic consistency-based $\mathbf{W}^{sc}$ (12), it can be observed that for edges with large weight $W_{i,j}^{sc}$, the conditions $P_i^{t1} = P_j^{t1}$ and $P_i^{t2} = P_j^{t2}$ hold with high probability. Consequently, according to (4a), we have $C_i^{t1\to t2} = C_j^{t1\to t2}$. In this case, the differences in the change signals between connected vertices—measured by $\left\|\mathbf{\Delta}_i^x - \mathbf{\Delta}_j^x\right\|_2$ and $\left\|\mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y\right\|_2$—are expected to be very small, implying that $\mathbf{\Delta}^x$ and $\mathbf{\Delta}^y$ behave as low-pass signals on the graph induced by $\mathbf{W}^{sc}$.

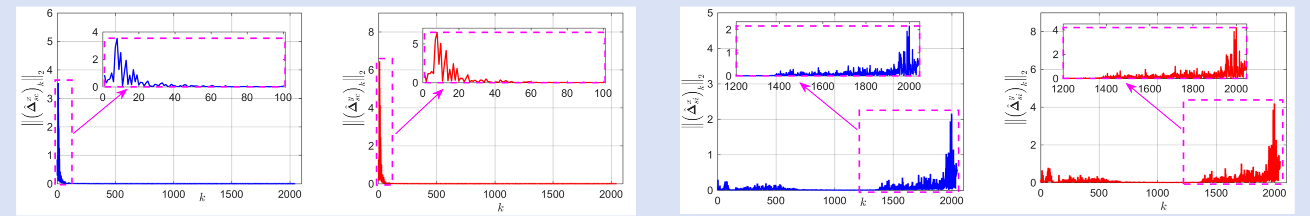
- For the weighting matrix $\mathbf{W}^{si}$ induced graph: In contrast, for the change semantic inconsistency-based matrix $\mathbf{W}^{si}$ constructed in (21), large weights $W_{i,j}^{si}$ typically correspond to



(a) $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ on the KNN graph G_{hn}^{t1} (b) $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ on the KNN graph G_{hn}^{t2}



(c) $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ on the KFN graph G_{hf}^{t1} (d) $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ on the KFN graph G_{hf}^{t2}



(e) $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ on the $\mathbf{W}^{sc}$ induced graph (f) $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ on the $\mathbf{W}^{si}$ induced graph

Fig. 1. Spectral curves of graph signals on different graphs: (a) the KNN graph G_{hn}^{t1}; (b) the KNN graph G_{hn}^{t2}; (c) the KFN graph G_{hf}^{t1}; (d) the KFN graph G_{hf}^{t2}; (e) the $\mathbf{W}^{sc}$ induced graph; (f) the $\mathbf{W}^{si}$ induced graph. The graph signals and graphs are constructed from the multi-temporal images of Dataset #1.

edges where either $P_i^{t1} = P_j^{t1}$ and $P_i^{t2} \neq P_j^{t2}$, or $P_i^{t1} \neq P_j^{t1}$ and $P_i^{t2} = P_j^{t2}$. According to (4b) and 4(c), these cases imply $C_i^{t1\to t2} \neq C_j^{t1\to t2}$. As a result, the differences in the change signals across these edges are expected to be very large, indicating that $\boldsymbol{\Delta}^x$ and $\boldsymbol{\Delta}^y$ exhibit significant high-frequency components on the graph induced by $\mathbf{W}^{si}$.

In Fig. 1(e)-(f), we plot the spectral curves of $\hat{\boldsymbol{\Delta}}^x_{sc} = \mathbf{U}_{sc}^{-1}(\boldsymbol{\Delta}^x)^T$ and $\hat{\boldsymbol{\Delta}}^x_{si} = \mathbf{U}_{si}^{-1}(\boldsymbol{\Delta}^x)^T$, where the orthonormal matrices of $\mathbf{U}_{sc}$ and $\mathbf{U}_{si}$ are calculated from the Laplacian matrices of $\mathbf{W}^{sc}$ and $\mathbf{W}^{si}$ induced graphs, respectively. As illustrated in Fig. 1, compared to the chaotic spectral components on the KNN and KFN graphs, the change signals exhibit distinguishable spectral characteristics on the graphs induced by $\mathbf{W}^{sc}$ and $\mathbf{W}^{si}$, that is, the spectral components of the change signals become separable on these induced graphs, as depicted in Fig. 1(e) and Fig. 1(f).

II. OPTIMIZATION OF CPIT MODEL

Here, we rewrite the forward transformation model of CPIT as follows

$$\begin{aligned}\min_{\mathbf{Y}',\boldsymbol{\Delta}^y}\ & 2\mathrm{Tr}\left(\mathbf{Y}'\mathbf{L}^{t1}\mathbf{Y}'^T\right) + 2\mathrm{Tr}\left(\boldsymbol{\Delta}^y\mathbf{L}^{sc}\left(\boldsymbol{\Delta}^y\right)^T\right) + \lambda\|\boldsymbol{\Delta}^y\|_{2,1}\\ &+\beta\sum_{(i,j)\in E_{hf}^{t1}}\varphi\left(\left\|\mathbf{Y}'_i-\mathbf{Y}'_j\right\|\right) + \alpha\sum_{i,j\in\mathcal{I}}\varphi\left(\left\|\boldsymbol{\Delta}^y_i-\boldsymbol{\Delta}^y_j\right\|\right)W^{si}_{i,j}\\ & s.t.\quad \mathbf{Y}=\mathbf{Y}'-\boldsymbol{\Delta}^y\end{aligned} \tag{S.2}$$

We solve the transformation model (S.2) by utilizing ADMM, which obtains the augmented Lagrangian function by introducing an auxiliary variable $\mathbf{P}\in\mathbb{R}^{4B_y\times N_S}$ with $\boldsymbol{\Delta}^y = \mathbf{P}$ as

$$\begin{aligned}\Theta\left(\mathbf{Y}',\boldsymbol{\Delta}^y,\mathbf{R}\right) =\ & 2\mathrm{Tr}\left(\mathbf{Y}'\mathbf{L}^{t1}\mathbf{Y}'^T\right) + 2\mathrm{Tr}\left(\mathbf{P}\mathbf{L}^{sc}\mathbf{P}^T\right)\\ &+\beta\sum_{(i,j)\in E_{hf}^{t1}}\varphi\left(\left\|\mathbf{Y}'_i-\mathbf{Y}'_j\right\|\right) + \alpha\sum_{i,j\in\mathcal{I}}\varphi\left(\left\|\mathbf{P}_i-\mathbf{P}_j\right\|\right)W^{si}_{i,j}\\ &+\mathrm{Tr}\left(\mathbf{R}_1^T\left(\mathbf{Y}'-\mathbf{Y}-\boldsymbol{\Delta}^y\right)\right)+\frac{\mu_1}{2}\left\|\mathbf{Y}'-\mathbf{Y}-\boldsymbol{\Delta}^y\right\|_F^2\\ &+\mathrm{Tr}\left(\mathbf{R}_2^T\left(\boldsymbol{\Delta}^y-\mathbf{P}\right)\right)+\frac{\mu_2}{2}\left\|\boldsymbol{\Delta}^y-\mathbf{P}\right\|_F^2+\lambda\|\boldsymbol{\Delta}^y\|_{2,1}\end{aligned} \tag{S.3}$$

where $\mathbf{R}_1,\mathbf{R}_2\in\mathbb{R}^{4B_y\times N_S}$ denote the Lagrange multipliers and $\mu_1,\mu_2>0$ denote the penalty parameters. The minimization of (S.3) can be decomposed into four subproblems.

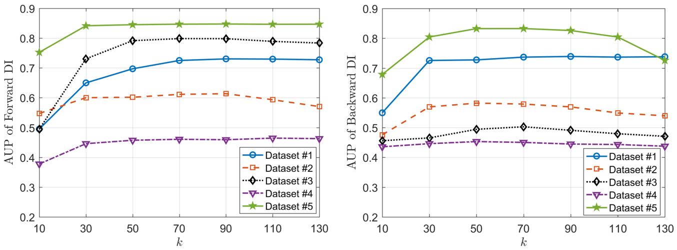
1) $\mathbf{Y}'$-subproblem. Minimizing (S.3) w.r.t. $\mathbf{Y}'$ can be expressed as

$$\begin{aligned}\min_{\mathbf{Y}'} 2\mathrm{Tr}\left(\mathbf{Y}'\mathbf{L}^{t1}\mathbf{Y}'^T\right)+\beta\sum_{(i,j)\in E_{hf}^{t1}}\varphi\left(\left\|\mathbf{Y}'_i-\mathbf{Y}'_j\right\|\right)\\ +\frac{\mu_1}{2}\left\|\mathbf{Y}'-\mathbf{Y}-\boldsymbol{\Delta}^y+\mathbf{R}_1/\mu_1\right\|_F^2\end{aligned} \tag{S.4}$$

Then, we can iteratively update $\mathbf{Y}'$ with an inner loop by using the gradient descent method.

2) $\mathbf{P}$-subproblem. Minimizing (S.3) w.r.t. $\mathbf{P}$ can be expressed as

$$\begin{aligned}\min_{\mathbf{P}} 2\mathrm{Tr}\left(\mathbf{P}\mathbf{L}^{sc}\mathbf{P}^T\right)+\alpha\sum_{i,j\in\mathcal{I}}\varphi\left(\left\|\mathbf{P}_i-\mathbf{P}_j\right\|\right)W^{si}_{i,j}\\ +\frac{\mu_2}{2}\left\|\boldsymbol{\Delta}^y-\mathbf{P}+\mathbf{R}_2/\mu_2\right\|_F^2\end{aligned} \tag{S.5}$$



(a) (b)

Fig. 2. Sensitivity analysis of parameter k.

Then, $\mathbf{P}$ can also be iteratively updated by using the gradient descent method.

3) $\mathbf{\Delta}^y$-subproblem. Minimizing (S.3) w.r.t. $\mathbf{\Delta}^y$ can be expressed as

$$\min_{\mathbf{\Delta}^y} \lambda \|\mathbf{\Delta}^y\|_{2,1} + \frac{\mu_1 + \mu_2}{2} \|\mathbf{\Delta}^y - \mathbf{Q}\|_F^2 \tag{S.6}$$

with

$$\mathbf{Q} = \frac{\mu_1 \mathbf{Y}' - \mu_1 \mathbf{Y} + \mathbf{R}_1 + \mu_2 \mathbf{P} - \mathbf{R}_2}{\mu_1 + \mu_2} \tag{S.7}$$

The closed-form solution of (S.6) is given by [5] as

$$\mathbf{\Delta}_i^y = \max \left\{ \|\mathbf{Q}_i\|_2 - \frac{\lambda}{\mu_1 + \mu_2}, 0 \right\} \frac{\mathbf{Q}_i}{\|\mathbf{Q}_i\|_2} \tag{S.8}$$

where we follow the convention $0 \cdot (0/0) = 0$.

4) $\mathbf{R}_1, \mathbf{R}_2$-subproblems. Finally, the Lagrangian multipliers are updated by

$$\begin{aligned} \mathbf{R}_1 &\leftarrow \mathbf{R}_1 + \mu_1 \left(\mathbf{Y}' - \mathbf{Y} - \mathbf{\Delta}^y \right) \\ \mathbf{R}_2 &\leftarrow \mathbf{R}_2 + \mu_2 \left(\mathbf{\Delta}^y - \mathbf{P} \right) \end{aligned} \tag{S.9}$$

III. PARAMETER ANALYSIS OF k

The parameter k plays an important role in defining the neighborhood structure of the KNN and KFN graphs. A very small k results in fewer connected vertex pairs, meaning that fewer pairwise relationships are exploited during optimization. This reduces the effectiveness of the change semantic consistency and inconsistency rules, as well as the structure-consistency-based regression regularization. Conversely, a very large k may introduce over-connections, which could decrease the accuracy of the regularization and increase computational complexity.

Considering $N_S = 2500$ in CPIT, we evaluated k values from 10 to 130 in steps of 20, and the corresponding change detection results are shown in Fig. 2. The results indicate that while variations in k cause minor fluctuations in the quantitative metric (AUP), the overall performance remains stable for k between 50 and 110. Specifically, the average AUP ranged from $0.670 < AUP_{forward} < 0.689$ and $0.603 < AUP_{backward} < 0.620$, demonstrating robustness to the choice of k. Considering both change detection performance and computational cost, we set $k = \sqrt{N_S}$ in CPIT. This choice is consistent with empirical guidelines from density estimation [6], KNN classification [7], and optimization of KNN-based density estimation [8], [9].

IV. INTEGRATING SPATIAL PRIORS INTO THE TRANSFORMATION MODEL

In CPIT, the change smoothness and consistency priors are incorporated into the difference images (DI) fusion and segmentation stage rather than directly into the image transformation model. Here, we discuss an alternative fusion strategy, in which they are directly incorporated into the image transformation model.

$$\begin{aligned} \min_{\mathbf{Y}', \mathbf{X}', \mathbf{\Delta}^y, \mathbf{\Delta}^x} & 2\mathrm{Tr}\left(\mathbf{Y}' \mathbf{L}^{t1} \mathbf{Y}'^T \right) + 2\mathrm{Tr}\left(\mathbf{X}' \mathbf{L}^{t2} \mathbf{X}'^T \right) \\ & + 2\mathrm{Tr}\left(\mathbf{\Delta}^y \mathbf{L}^{sc} \left(\mathbf{\Delta}^y \right)^T \right) + 2\mathrm{Tr}\left(\mathbf{\Delta}^x \mathbf{L}^{sc} \left(\mathbf{\Delta}^x \right)^T \right) \\ & + \beta_1 \sum_{(i,j) \in E_{hf}^{t1}} \varphi\left(\left\| \mathbf{Y}_i' - \mathbf{Y}_j' \right\| \right) + \beta_2 \sum_{(i,j) \in E_{hf}^{t2}} \varphi\left(\left\| \mathbf{X}_i' - \mathbf{X}_j' \right\| \right) \\ & + \alpha_1 \sum_{i,j \in \mathcal{I}} \varphi\left(\left\| \mathbf{\Delta}_i^y - \mathbf{\Delta}_j^y \right\| \right) W_{i,j}^{si} + \alpha_2 \sum_{i,j \in \mathcal{I}} \varphi\left(\left\| \mathbf{\Delta}_i^x - \mathbf{\Delta}_j^x \right\| \right) W_{i,j}^{si} \\ & + \lambda_1 \|\mathbf{\Delta}^y\|_{2,1} + \lambda_2 \|\mathbf{\Delta}^x\|_{2,1} + \gamma_1 f\left(\mathbf{\Delta}^y \right) + \gamma_2 f\left(\mathbf{\Delta}^x \right) \\ & + \mu \phi\left(\mathbf{\Delta}^x, \mathbf{\Delta}^y \right) \\ & s.t. \quad \mathbf{Y} = \mathbf{Y}' - \mathbf{\Delta}^y, \quad \mathbf{X} = \mathbf{X}' - \mathbf{\Delta}^x \end{aligned} \tag{S.10}$$

where $\beta_1, \beta_2, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \gamma_1, \gamma_2, \mu > 0$ are penalty parameters, the function $f(\cdot)$ denotes change spatial smoothness based regularization and the function $\phi(\cdot)$ is change spatial consistency based regularization. For example, we can design the smoothness based function as

$$\begin{aligned} f\left(\mathbf{\Delta}^y \right) &= \mathrm{Tr}\left(\mathbf{\Delta}^y \mathbf{L}^{adj} \left(\mathbf{\Delta}^y \right)^T \right) \\ f\left(\mathbf{\Delta}^x \right) &= \mathrm{Tr}\left(\mathbf{\Delta}^x \mathbf{L}^{adj} \left(\mathbf{\Delta}^x \right)^T \right) \end{aligned} \tag{S.11}$$

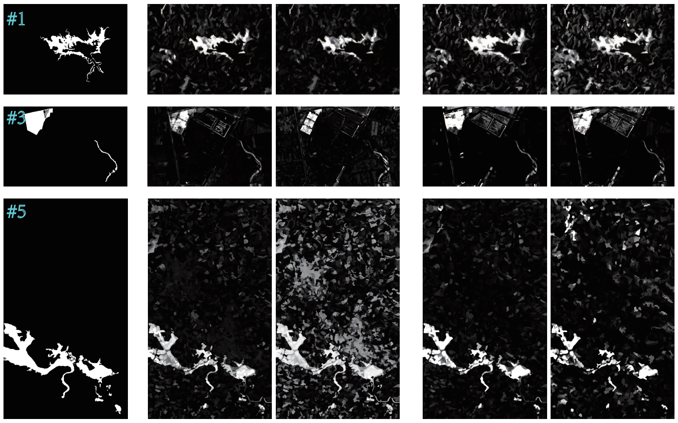
with $\mathbf{L}^{adj}$ being the Laplacian matrix of graph that connectes each superpixel with its spatial neighborhood, e.g., superpixel pairs whose boundaries intersect or whose centroid distance is below a threshold. We can choose the spatial consistency based penalty function as

$$\begin{aligned} \phi^1\left(\mathbf{\Delta}^x, \mathbf{\Delta}^y \right) &= \left(\|\mathbf{\Delta}_i^x\|_2 - \|\mathbf{\Delta}_i^y\|_2 \right)^2 \\ \phi^2\left(\mathbf{\Delta}^x, \mathbf{\Delta}^y \right) &= -\|\mathbf{\Delta}_i^x\|_2 \|\mathbf{\Delta}_i^y\|_2 \\ \phi^3\left(\mathbf{\Delta}^x, \mathbf{\Delta}^y \right) &= \exp\left(-\|\mathbf{\Delta}_i^x\|_2 \|\mathbf{\Delta}_i^y\|_2 \right) \end{aligned} \tag{S.12}$$

The minimization of $\phi\left(\mathbf{\Delta}^x, \mathbf{\Delta}^y \right)$ requires that whenever either change level, $\|\mathbf{\Delta}_i^x\|_2$ or $\|\mathbf{\Delta}_i^y\|_2$, is large, the other one should also be large, which means that $\|\mathbf{\Delta}_i^x\|_2$ and $\|\mathbf{\Delta}_i^y\|_2$ are aligned.

In Fig. 3, we show the DIs generated by CPIT and the fused model of (S.10) with ϕ^1, where the hyperparameters have been carefully adjusted. It can be observed that after introducing the constraint of change spatial consistency, the forward and backward DIs become more consistent and more similar, as shown in Fig. 3(d) and (e), which further validates the effectiveness of the fused model of (S.10). However, this constraint is not always effective, especially under conditions where the model is difficult to optimize, as reflected by the AUP comparisons reported in Table I.

Based on the above analysis and experimental results, applying change smoothness and consistency priors directly to the image transformation stage, such as the fused model in (S.10), can achieve higher detection accuracy in certain cases.



(a) (b) (c) (d) (e)

Fig. 3. DIs generated by CPIT and fused model of (S.10). (a) Ground truth; (b) Forward DI of CPIT; (b) Backward DI of CPIT; (c) Forward DI of fused model with (S.10); (d) Backward DI of fused model with (S.10).

TABLE I
THE AUP OF DIS GENERATED BY CPIT AND THE FUSED MODEL OF (S.10).

Methods	Dataset #1		Dataset #3		Dataset #5	
	DI^y	DI^x	DI^y	DI^x	DI^y	DI^x
CPIT	0.687	0.717	0.794	0.500	0.846	0.833
Fused (S.10)	0.711	0.622	0.734	0.742	0.836	0.539

However, this strategy also involves two major risks. First, it introduces more hyperparameters, making the model harder to tune and control, and requiring a more complex parameter adjustment process. Second, it leads to a more challenging optimization problem with higher computational complexity and increased runtime. Considering these trade-offs, we adopt the CPIT fusion strategy in this work. In future studies, we will explore more effective fusion schemes that can directly incorporate spatial change priors into the transformation model while maintaining stable and efficient optimization.

V. COMPUTATIONAL COMPLEXITY

The CPIT framework is composed of three main stages: preprocessing, relationship representation, and image transformation.

- Preprocessing: The complexity of the superpixel segmentation method [10] is linear with respect to the image size, i.e., $\mathcal{O}(MN)$. Given that the average superpixel contains approximately MN/N_S pixels, the computational cost of extracting mean features is $\mathcal{O}(MNB)$, where $B = B_x + B_y$ denotes the total number of image bands. For the extraction of three quartile features, which typically involves sorting operations, the computational complexity is $\mathcal{O}(MNB\log(MN/N_S))$.
- Relationship representation: Computing pairwise feature distance matrices requires $\mathcal{O}(N_S^2 B)$; sorting the distance matrix column-wise has a complexity of $\mathcal{O}(N_S^2 \log N_S)$ by using efficient sorting algorithms; building high-order graphs requires $\mathcal{O}(N_S^2)$. As such, the total cost of relationships representation is dominated by $\mathcal{O}(N_S^2 \log N_S)$.
- Image transformation: This stage involves solving the optimization problem of (S.2). For example, updating $\mathbf{Y}'$ with (S.4) and updating $\mathbf{P}$ with (S.5) both require $\mathcal{O}(N_S^2 B_y)$ for each inner loop; updating $\boldsymbol{\Delta}^y$ using the closed-form solution (S.8) can be done column-wise, yielding $\mathcal{O}(N_S B_y)$; updating the Lagrange multipliers via (S.9) also requires $\mathcal{O}(N_S B_y)$. Overall, the image transformation has a total complexity of $\mathcal{O}(N_S^2 B)$, considering both forward and backward transformations.

Although the proposed algorithm requires $\mathcal{O}(N_S^2 \log N_S)$ for relationship representation (no need for updates and iterations) and $\mathcal{O}(N_S^2 B)$ for ADMM based image transformation by using iterative gradient descent, the optimization is performed at the superpixel level rather than the pixel level, which significantly reduces the problem size. This advantage is also empirically validated by the running time reported in Table II. In addition, several factors help keep the overall runtime at a reasonable level, such as the structural consistency based initialization strategy and the early-stopping strategy described in Section IV-C.3 "Convergence behavior" of the manuscript.

In Table II, we report the execution times of CPIT on all datasets. Specifically, t_{pre} denotes the time for preprocessing, t_{rr} corresponds to relationship representation, t_{it} represents the time required for forward and backward image transformations, t_{ce} denotes the time spent on change extraction, and t_{total} represents the total computational time of CPIT. As shown in Table II, among all components, the image transformation stage is the most time-consuming component, as it involves iterative optimization via gradient descent. To further improve efficiency, future work will explore more advanced optimization strategies for solving the underlying image transformation models. Moreover, replacing the full-graph models (KNN and KFN graphs) with anchor-graph models could further reduce graph complexity. For large-scale images, region-wise parallel processing is also a promising strategy to further improve scalability.

TABLE II
THE COMPUTATIONAL TIME (SECONDS) OF CPIT ON DIFFERENT DATASETS.

Datasets	Image size	t_{pre}	t_{rr}	t_{it}	t_{ce}	t_{total}
Dataset #1	$300 \times 412 \times 1(3)$	0.5	3.9	34.1	0.2	38.7
Dataset #2	$2000 \times 2000 \times 3(3)$	3.8	4.9	41.8	1.8	52.3
Dataset #3	$593 \times 921 \times 1(3)$	1.2	4.5	39.9	0.5	46.1
Dataset #4	$3500 \times 2000 \times 11(3)$	6.6	6.4	40.5	2.9	56.4
Dataset #5	$4135 \times 2325 \times 3(1)$	7.9	5.7	42.3	4.0	59.9

REFERENCES

[1] L. Stankovic, D. P. Mandic, M. Dakovic, I. Kisil, E. Sejdic, and A. G. Constantinides, "Understanding the basis of graph signal processing via an intuitive example-driven approach [lecture notes]," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 133–145, 2019. 1

[2] L. Stanković, M. Daković, and E. Sejdić, "Introduction to graph signal processing," in *Vertex-Frequency Analysis of Graph Signals*. Springer, 2019, pp. 3–108. 1

[3] Y. Sun, L. Lei, D. Guan, G. Kuang, and L. Liu, "Graph signal processing for heterogeneous change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–23, 2022. 1

[4] Y. Sun, L. Lei, Z. Li, and G. Kuang, “Similarity and dissimilarity relationships based graphs for multimodal change detection,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 208, pp. 70–88, 2024. 1

[5] J. Yang, W. Yin, Y. Zhang, and Y. Wang, “A fast algorithm for edge-preserving variational multichannel image restoration,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009. 3

[6] D. O. Loftsgaarden and C. P. Quesenberry, “A nonparametric estimate of a multivariate density function,” *The Annals of Mathematical Statistics*, vol. 36, no. 3, pp. 1049–1051, 1965. 3

[7] P. Mitra, C. A. Murthy, and S. K. Pal, “Unsupervised feature selection using feature similarity,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 301–312, 2002. 3

[8] K. Fukunaga and L. Hostetler, “Optimization of k nearest neighbor density estimates,” *IEEE Transactions on information theory*, vol. 19, no. 3, pp. 320–326, 2003. 3

[9] L. Devroye, L. Györfi, and G. Lugosi, *A probabilistic theory of pattern recognition*. Springer Science & Business Media, 2013, vol. 31. 3

[10] Z. Ban, J. Liu, and L. Cao, “Superpixel segmentation using gaussian mixture model,” *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4105–4117, 2018. 4