Expert Systems With Applications 227 (2023) 120181

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Phase-based fine-grained change detection

Xuzhi Wang, Liang Wan *, Di Lin, Wei Feng

College of Intelligence and Computing, Tianjin University, Tianjin, China

ARTICLE INFO

Keywords:
Complex steerable pyramid
Change detection
Image registration
Structure information

ABSTRACT

Detecting fine-grained changes among a set of images taken at different times is a challenging problem, which is important for the applications such as high-value scene monitoring and structural inspection. Existing methods use a geometry registration and lighting compensation pipeline to alleviate the ingredients that affect change detection. However, lighting compensation may introduce a lot of artifacts, e.g., inaccurate local lighting conditions and fine-grained change disappearance. The distortion introduced by general registration methods also makes the process of lighting correction more challenging. To solve these problems, in this paper, we propose a new phase-based change detection method, called as pb-FGCD, which first aligns two observation of different times and then extract structure information to handle lighting difference. Benefiting from the proposed weighted amplitude filter and iterative alignment strategy, we could obtain a better registration performance. The structure information extracts the lighting irrelevant components of an image and is more sensitive to structure changes compared with the lighting compensation based method, which is vital for detecting the changes with low intensity difference. Furthermore, we contribute a new change detection dataset, which contains 100 groups of the real-world data of the ancient Mogao murals and the Terra-Cotta. It is the first fine-grained change detection dataset focusing on the ancient mural. Our approach significantly improves the accuracy by more than 190% higher F1-measure than the state-of-the-art methods.

1. Introduction

One important task in cultural heritage protection is to accurately detect slow and subtle changes on the surface of culture heritage relics, e.g., sculptures and murals, over time (months or years). This fine-grained change detection of high-value objects is a challenging task (Feng et al., 2015; Stent et al., 2016), as the example shown in Fig. 1. First, it is difficult to restore the exact same location and pose of the camera to take images at different times. This inevitably leads to geometric differences, which usually appear with non-linear distortion of the mural surface. Second, the images were often taken under discrepant lighting conditions. Third, the object may have subtle changes, such as cracking and shedding on the surface, which show complex patterns due to the scene damage over time.

It should be noted that trivially computing the difference between two images leads to the problematic change map (see Fig. 1(d)), which is very sensitive to object edges. Even if there is little difference between viewpoints, the change map can be very sensitive to lighting conditions at different periods (see Fig. 1(e)). Figs. 1(g-i) show the detection results of state-of-the-art methods, among which Feng et al. (2015) obtains a better result. The existing methods (Feng et al., 2015; Stent et al., 2016) formulate the fine-grained change detection as a two-stage task, which conduct geometry registration and lighting correction to handle the above challenges. The detected changes are finally obtained by the intensity difference of one image and the other adjusted image. What is more, we notice that there is an inherent dilemma in the process of lighting correction (Feng et al., 2015; Hou et al., 2021; Luan et al., 2017). On one hand, we aim to compensate the lighting condition of last observation to be the same as current observation. On the other hand, these effects should not make the changes between two images disappear. Especially, the general geometry registration methods may introduce a lot of distortions in the fine-grained area, which makes the lighting correction more challenging.

To address these problems, we propose a phase-based registration method which can accurately manipulate subtle movements by adjusting their phases in the frequency domain. We then extract structure information in the frequency domain to handle lighting differences between two observations of different times. Note that the structure information is more sensitive to structure changes compared with lighting correction based method, which is vital for detecting the fine-grained changes with low intensity difference. Many fine-grained changes in the mural scene have low intensity changes as shown in Fig. 6.

The registration process of our method belongs to phase-based methods (Davis et al., 2015, 2014; Elgharib et al., 2015; Meyer et al.,

* Corresponding author.
E-mail addresses: wangxuzhi@tju.edu.cn (X. Wang), lwan@tju.edu.cn (L. Wan), di.lin@tju.edu.cn (D. Lin), wfeng@tju.edu.cn (W. Feng).

https://doi.org/10.1016/j.eswa.2023.120181
Received 27 January 2022; Received in revised form 26 March 2023; Accepted 15 April 2023
Available online 23 April 2023
0957-4174/© 2023 Elsevier Ltd. All rights reserved.

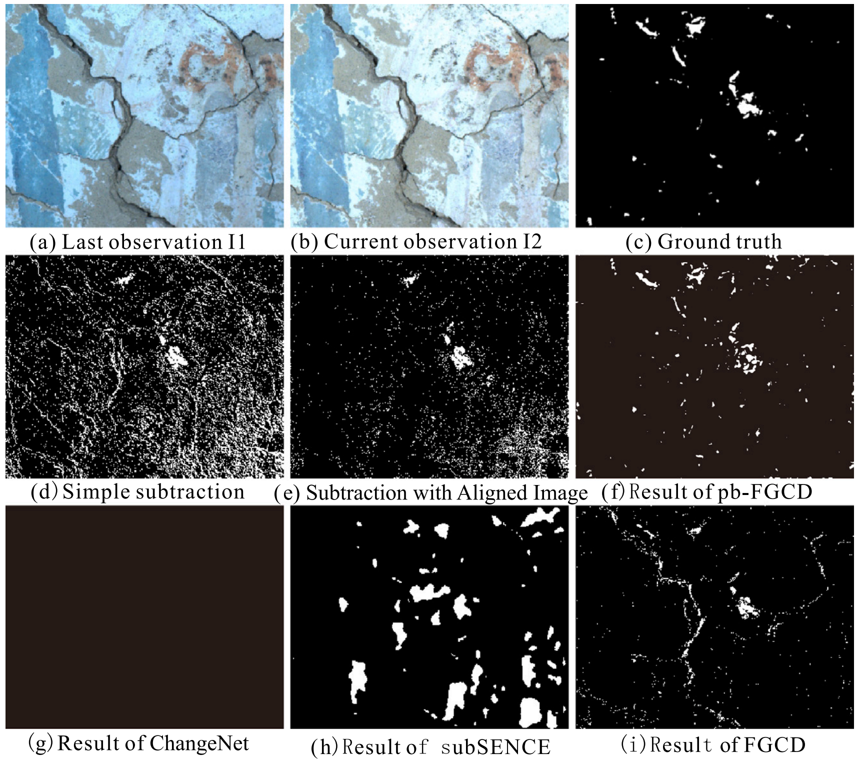


Fig. 1. A real case of fine-grained change detection of mural in the Mogao Grottoes.

2018; Takeda et al., 2018; Zhang et al., 2017). We first decompose the input images into different scale and orientation subbands in the frequency domain by applying the complex steerable pyramid. Each subband contains two components, i.e. phase and amplitude. We then adjust the phase of reference observation in different scale and orientation subbands by the proposed weighted amplitude filter guided by current observation. The above registration process runs in a recursive way to improve the registration result. We next use the complex steerable pyramid to extract phase congruency features, which is used as image structure information that can handle lighting differences. At last, we compare the structure information of the two observations of different times to obtain the final change map. The contributions of this paper are:

- We propose a novel and effective framework that conducts phase-based registration and structure information extraction for fine-grained change detection. To the best of our knowledge, it is the first work that applies the phase-based method to fine-grained change detection.
- We introduce weighted amplitude filter and iterative alignment strategy which allows the phase-based method for accurate registration and change detail preservation. We extract structure information to obtain change detection results in different lighting conditions. Our approach increases the F1-measure by more than 190% over those state-of-the-art methods.
- We built a real-world dataset to benchmark fine-grained change detection of misaligned scenes under varied lighting conditions of the ancient mural and the Terra-Cotta, containing 100 scenes.

The rest of the paper is organized as follows. Section 2 reviews the related works. The proposed method is introduced in Section 3. The experimental results and analysis of the results are detailed in Section 4. Section 5 draws the conclusions of the proposed method. We also give the intuition for the phase-based method in Appendix.

2. Related work

2.1. Change detection

There are a number of methods developed for change detection, including background models (Chen et al., 2017; Lin et al., 2017), nonparametric models (Elgammal et al., 2000) and low rank models (Bouwmans et al., 2017), which is desirable in many applications such as moving object detection (Lin et al., 2017), video surveillance (Elgammal et al., 2000), video indexing (Chen et al., 2017; Gargi et al., 2000; Urhan et al., 2006), to name a few. For instance, Lin et al. (2017) proposed a method for detecting the changes caused by motion based on attention mechanism. Bescos (Bescós, 2004) proposed a module for video temporal segmentation which is able to detect both abrupt transitions and all types of gradual transitions in real-time. Urhan et al. (2006) proposed a change detection method for hard-cut of archive film. They used phase-correlation to calculate the similarities between consecutive frames, and low correlation indicates a candidate hard-cut. PCA-K-Means (Celik, 2009) proposes a multitemporal satellite image change detection method by principal component analysis and k-means clustering. NPSG (Sun et al., 2021) proposes a change detection method using similarity measurement. The basic idea of NPSG is to represents the image structure by the nonlocal path similarity based graph and compare the two representations for change detection.

Recently, convolutional neural networks methods have been developed for change detection. Nguyen et al. (2004) proposed a data-driven method to detect changes related to object motion.

However, we have to note that CNN-based methods require a sufficiently large training set. Although data augmentation (Zhao et al., 2017), weakly supervised learning (Zhao et al., 2018), few-shot learning (Zhou et al., 2018) et al. do help, the existing fine-grained change detection datasets of cultural heritage, are too small to train a network on par with other change detection tasks. Below, we review the most relevant works for fine-grained change detection.

Fine-grained change detection (FGCD) is different from video surveillance, as it focuses on the fine-grained change of background (e.g., the mural peeling and the color changes of the mural). Feng et al. (2015) used the geometry correction (Liu et al., 2011), the lighting differences (Barron & Malik, 2015) and the change mask to estimate the result in a coarse-to-fine manner. It was the first work that focuses on fine-grained change detection for high-value monitoring. These methods modeled the fine-grained change detection as a joint optimization of the viewpoint variation, the photometric variation and the changes in the scene. Stent et al.'s method (Stent et al., 2016) compensates for the variations of the viewpoint, photometry and resolution/focal setting, and effectively improves the performance of FGCD. Note that these works heavily depend on the optical flow to reduce the effect of different variations on the detection. But the interaction between the lighting correction and the computation of optical flow, which were done independently, easily led to artifacts of the final detection results.

In contrast to the existing methods, we develop a phase-based scheme to address both geometric and photometric variations. We exploit structure information, which is less sensitive to varying lighting conditions, to achieve robust detection of subtle changes.

2.2. Phase-based methods

The phase-based methods have been used in many applications (Davis et al., 2015, 2014; Elgharib et al., 2015; Meyer et al., 2018; Takeda et al., 2018; Wang et al., 2022; Wu et al., 2012; Zhang et al., 2017), which recognized the vibration of the object surface, or tackle small movements (Meyer et al., 2015; Wadhwa et al., 2013). These applications relied on the ability of the complex steerable pyramid to manipulate subtle movements by adjusting their phase in a video sequence.

For example, to enable larger amplification factors for the motion magnification, Wadhwa et al. (2013) proposed a phase-based video magnification method, in which they employed the complex-steerable pyramid to decompose the image to obtain the phase variations of each pixel. They encoded the motion information into the phase shift and amplified the small movements. In these approaches (Davis et al., 2015; Elgharib et al., 2015), the complex steerable pyramid was used for extracting the local motion signal of the object surface. Meyer et al. (2015) proposed a phase-based method for the frame interpolation of the video. They reduced the phase variation to generate intermediate frames, which provided smoother transitions between the interpolated images.

We employ the phase-based scheme for aligning images. Different from the video in which the appearances of the same object in different frames are the same, the image pairs, which were captured in different periods, can have locally different appearances. Although the interpolation of the video frames can yield smooth motion, it usually losses the details of changes. In our work, we develop a weighted amplitude filter to align the image pair. This is done by an iterative alignment process, which is good at preserving the detailed information of changes.

2.3. 2D image alignment

Many existing works adopt variational methods for optical flow estimation to achieve the alignment of 2D images (Horn & Schunck, 1981). Many works (Brox & Malik, 2011; Liu et al., 2011; Weinzaepfel et al., 2013) focused on the accurate estimation of optical flow, which better captured the large displacements of pixels or dramatic changes in appearance. Weinzaepfel et al. (2013) proposed the correlated multi-scale patches that were matched for producing the optical flow. Compared to the conventional variational methods, our weighted amplitude filter provides better property of motion smoothness.

In recent years, CNN-based methods for optical flow estimation have been developed (Dosovitskiy et al., 2015). Ilg et al. (2017) used the cascaded architecture to improve the accuracy of flow estimation but at cost of increasing the computational overhead. The recent SpyNet (Ranjan & Black, 2017), PWC-Net (Sun et al., 2018), and LiteFlowNet (Hui et al., 2018) were lightweight networks that can achieve competitive performance on the flow estimation. However, to reduce the ghosting effects in the warped images, the learning-based methods require a lot of training data, which significantly increases the manual labeling efforts.

Different from the above 2D registration methods, our method is relied on the phase-based scheme. Since in the case of fine-grained change detection, the viewpoint differences of two observations of different times are not large and the scene is fine-grained, the phase-based scheme can manipulate subtle movements accurately by adjusting their phases.

2.4. Illumination equivalence

Many methods, such as color constancy (Gijsenij et al., 2011; Yang et al., 2015), intrinsic image decomposition (Hauagge et al., 2013), and some deep learning-based method (Fourure et al., 2016), can be used to do lighting correction, given the images under different lighting conditions.

Inspired by structural similarity which is used for comparing the local patterns of the pixel intensities for image quality assessment (Wang et al., 2004; Zhang et al., 2011), we use the complex steerable pyramid to extract phase congruency which implies the structure information irrelevant to lighting variation. Experiments show that structural information is good at handling lighting differences for fine-grained change detection.

3. The proposed method

Given two observations of different time denoted as I_{ref} and I_{cur}, where I_{ref} represents reference observation and I_{cur} represents current observations, the goal of fine-grained change detection is to estimate the change map C. For the existing fine-grained change detection methods, C can be formulated by:

$$C = L(R(I_{\text{ref}})) - I_{\text{cur}}, \tag{1}$$

where R(·) represents the geometry registration process and L(·) represents the lighting compensation process, which aligns the geometry and lighting condition of I_{ref} to I_{cur}. Lighting compensation can be seen as adding a "virtual" light to reference observation to correct its lighting differences to the current observation.

Since lighting compensation may introduce many false-positive results and changes between two observations may disappear during the lighting compensation process, we propose to extract structure information which is irrelevant to lighting conditions. The framework is shown in Fig. 2. We first use the complex steerable pyramid to align the two images taken at different times (say, with one-year interval). We then exploit the complex steerable pyramid to extract structure information to handle the lighting difference between the two observations and finally obtain the change map.

In our method, the change map C is formulated as

$$C = P(R(I_{\text{ref}})) - P(I_{\text{cur}}), \tag{2}$$

where R(·) and P(·) represent the registration operation and structure information extraction. In the following, we introduce the details of our fine-grained change detection method.

3.1. Phase-based registration

Given an input image I, it is decomposed into complex-valued responses $S_{\omega,\theta}$ of different scales and orientations by applying the steerable filters $\psi_{\omega,\theta}$ to I,

$$\begin{aligned} S_{\omega,\theta} &= (I * \psi_{\omega,\theta})(x,y), \\ &= A_{\omega,\theta}(x,y)e^{i\phi_{\omega,\theta(x,y)}}, \\ &= E_{\omega,\theta}(x,y) + iO_{\omega,\theta}(x,y), \end{aligned} \tag{3}$$

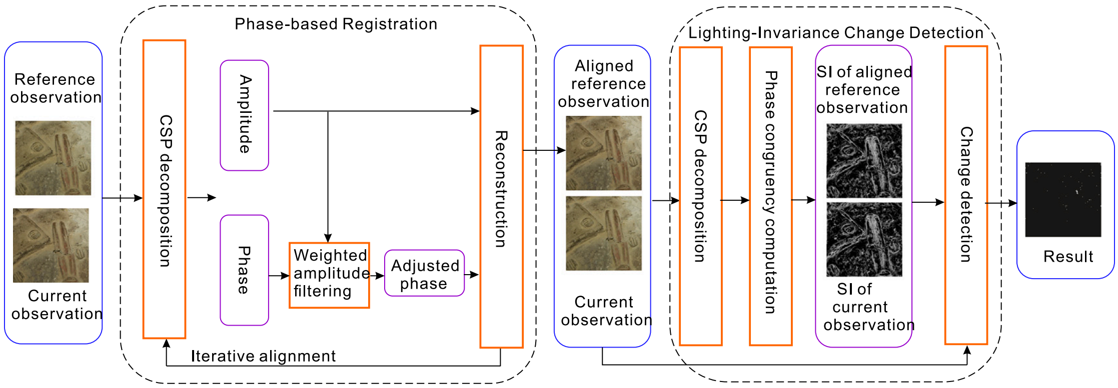


Fig. 2. The framework of the proposed method.

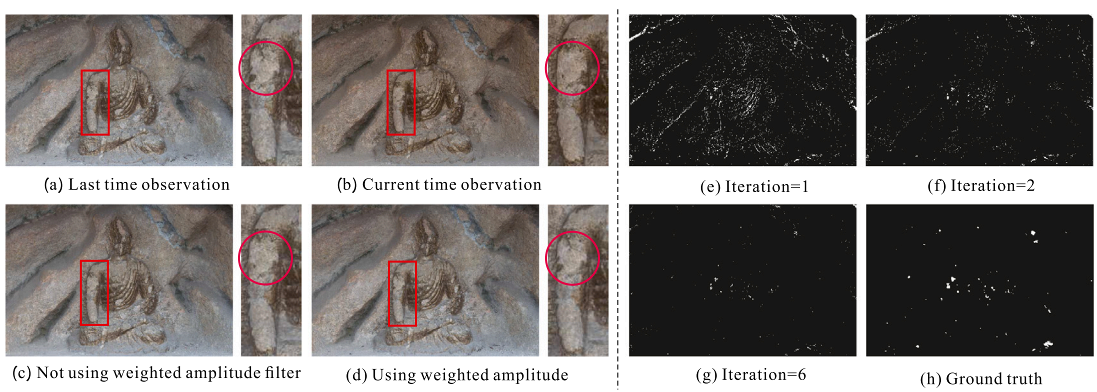


Fig. 3. Registration results with/without weighted amplitude filter and iterative alignment process.

where $E_{\omega,\theta}$ is the even-symmetric filter response, and $O_{\omega,\theta}$ is the odd-symmetric filter response. The amplitude is computed by $A_{\omega,\theta}(x,y) = \sqrt{E^2_{\omega,\theta}(x,y) + O^2_{\omega,\theta}(x,y)}$, and the phase is computed by $\phi_{\omega,\theta}(x,y) = \arctan(O_{\omega,\theta}(x,y)/E_{\omega,\theta}(x,y))$.

The decomposition of the image pair (I_{ref}, I_{cur}) into phase and amplitude can be written as:

$$I_{\text{ref}}(x,y) \rightarrow A_{\text{ref}}(\omega,\theta,x,y)e^{i\phi_{\text{ref}}(\omega,\theta,x,y)}, \tag{4}$$

$$I_{\text{cur}}(x,y) \rightarrow A_{\text{cur}}(\omega,\theta,x,y)e^{i\phi_{\text{cur}}(\omega,\theta,x,y)}. \tag{5}$$

We then compute the phase variation between two observations by subtracting the local phase ϕ of the current observation in each position, orientation and scale in the complex steerable pyramid from the reference observation, given by

$$\phi_{\text{var}}(\omega,\theta,x,y) = \phi_{\text{ref}}(\omega,\theta,x,y) - \phi_{\text{cur}}(\omega,\theta,x,y). \tag{6}$$

Next, the aligned reference image with respect to the current observation can be computed by simply subtracting ϕ_{var} from ϕ_{ref}, as follow,

$$\begin{aligned} I^{\text{reg}}_{\text{ref}} &= R(A_{\text{ref}}(\omega,\theta,x,y)e^{\phi_{\text{ref}}-\phi_{\text{var}}}), \\ &= R(A_{\text{ref}}(\omega,\theta,x,y)e^{\phi_{\text{cur}}}). \end{aligned} \tag{7}$$

This will produce a seemingly good registration result shown in Fig. 3(c), for which the reference observation in Fig. 3(a) is aligned to the current observation in Fig. 3(b). But A close inspection reveals that Fig. 3(c) loses many details of the reference image compared to Fig. 3(a) (look at the circled regions in the blow-ups). It is because the arrangement of the phase encodes the structure information of the image, and Eq. (7) directly uses the phase value including the changed area in each scale and orientation subband of the current observation. Hence, Fig. 3(c) looks like the current observation in details.

Weighted amplitude filtering. To handle the above problem, let us analyze the phase variation. The phase variation between the reference observation and the current observation in each subband of the steerable pyramid is caused by two main factors: viewpoint variation and scene changes. Scene changes are indeed the real change of the objects. It is the variance of the object appearance regardless of viewpoint variation, lighting difference, etc. Therefore, if we can eliminate the phase variation caused by scene changes, we will obtain the phase variation caused by viewpoint variation and do the alignment more accurately.

Here, we propose a weighted amplitude filtering to eliminate the phase variation. Considering the prior knowledge that the structure moves in the same way in the spatial neighborhood (Brox & Malik, 2011; Horn & Schunck, 1981; Weinzaepfel et al., 2013), we introduce the smoothness constraint in a filtering manner, given by

$$\phi^{\text{filtered}}_{\text{var}} = \frac{(\gamma\phi_{\text{var}}A_{\text{ref}}) * G_\rho}{A_{\text{ref}}}, \tag{8}$$

where G_ρ is a Gaussian kernel given by $e^{-\frac{x^2+y^2}{\rho^2}}$, the operator * denotes convolution operation, A_{ref} is the amplitude of the reference observation, and γ is a factor that controls the displacement in the spatial domain which we can manipulate by shifting the phase. The Gaussian kernel in the weighted amplitude filter is to make the pixels move with a similar distance to their neighborhood. We further use the amplitude as a weight since each part in a scene usually has a similar amplitude in each band of the steerable pyramid and each part moves in a similar way. In this way, the alignment will not change the arrangement of the pixels and thus preserve the detail of the scene changed area. In experiments, we find that a small amount of phase variation can better reflect the real motion between two images. Specifically, we choose γ=0.5 in our work which makes a trade-off between the amount of displacement that can be successfully represented and the computation efficiency.

Iterative alignment. Let us examine the ideal case of image alignment. Since the offsets between two images are proportional to their phase variation, intuitively, the two images are well aligned when their phase variation equals zero. However, applying the weighted amplitude filtering for one time cannot make the phase variation equals zero. As shown in Fig. 3(e), the image pair is not well aligned, which has many unaligned edges. To handle this, we apply the weighted amplitude filtering in an iterative manner. That is, we iteratively process ϕ_{ref} to make it approach to ϕ_{cur},

$$\phi_{\text{ref}}^{k+1} = \phi_{\text{ref}}^{k} - \phi_{\text{var}}^{\text{filtered}\,k}. \tag{9}$$

In this way, the current observation is warped successively by each iteration and finally aligned to the reference observation. The aligned image can be computed by

$$I_{\text{ref}}^{\text{reg}} = R(A_{\text{ref}}^{k} e^{\phi_k}), \tag{10}$$

where k represents kth iteration of the base algorithm.

Fig. 3(d) shows the aligned result, which is well-aligned with the current observation and also well preserves the original details. The quantitative performance of the iterative registration is shown in Fig. 8.

For the observation images of murals and sculptures, the geometry difference usually contains deformation caused by non-linear deformation of the surface and displacement caused by viewpoint variation. The weighted amplitude filter and the iterative process can be regarded as a decomposition of the complex movement. In the experiment, we find that the former iterations most compensates for the displacement caused by viewpoint variation, and the latter iterations most compensate for the deformation caused by non-linear deformation (as demonstrated in Fig. 3(e–g)).

3.2. Lighting-invariance change detection

When we align the image pair taken at different times, we can estimate changes by subtracting the aligned images in the spatial domain, given by

$$C_{\text{im}} = \begin{cases} 1, & \zeta \leq |L(I_{\text{ref}}^{\text{reg}}) - I_{\text{cur}}| \\ 0, & otherwise \end{cases}. \tag{11}$$

Existing methods detect the changes between two observations by comparing their intensity. However, this spatial change may contain some false-negative results (as demonstrated in Fig. 1(d)), which is caused by the lighting difference. And this spatial change may contain some false-negative result for some changes may have low intensity changes between two observations.

Notice that the phase of the complex steerable pyramid implies structure information that is robust to lighting conditions. Further, structure information is more sensitive to structure changes even the changes has little color variance. We propose to extract the structure information to handle the lighting difference between two observations of different times.

Specifically, we first extract the phase congruency based on complex steerable pyramid, which indicates the structure information (Sampat et al., 2009; Zhang et al., 2011), at position (x, y) as follows:

$$P(x, y) = \frac{\sum_{\theta} E_{\theta(x,y)}}{\epsilon + \sum_{\omega} \sum_{\theta} A_{\omega,\theta}(x, y)}, \tag{12}$$

where ϵ is a small positive constant used to prevent division by zero. Here, we choose $\epsilon = 0.0001$. Let the phase congruency of the reference observation be P_{ref}, the current observation be P_{cur}, and the registered observation be P_{ref}^{reg}, respectively.

As we project two observations of different times into one feature space which is irrelevant to lighting difference and sensitive to structure information. Then the changes between the two observations of different times can be estimated by,

$$P_{\text{c}} = P_{\text{cur}} - P_{\text{ref}}^{\text{reg}}. \tag{13}$$

Note that P_{ref}^{reg} may contain some residual components which are not removed by the phase-based registration stage. We also compute the second difference to compensate the residual components, as follows,

$$P_{\text{r}} = P_{\text{cur}} - P_{\text{ref}}. \tag{14}$$

Given P_{c} and P_{r}, we estimate the changes in the phase domain as,

$$P = \begin{cases} 1, & \delta \leq |P_{\text{c}} - \lambda * P_{\text{r}}| \\ 0, & otherwise \end{cases}, \tag{15}$$

where λ is a parameter used to adjust the relative importance of P_{c} and P_{r}.

Since the phase domain and the spatial domain are complementary, and they reflect different aspects of changes of the input image pair. Phase domain is more sensitive to structure changes between two observations and spatial domain are more sensitive to intensity changes between two observations. We fuse the detected changes from the phase domain and the spatial domain to get accurate results.

The final change map is computed as the intersection between C_{im} and P, given by

$$C_{\text{final}} = C_{\text{im}} \cap P. \tag{16}$$

4. Experimental results

4.1. Experimental settings

We compare our methods with five state-of-the-art FGCD methods including PCA-K-Means (Celik, 2009), ChangeNet (Varghese et al., 2018), SubSENCE (St-Charles et al., 2014), FGCD (Feng et al., 2015) and NPSG (Sun et al., 2021), on four real-world datasets, including three public datasets (Feng et al., 2015) and a newly collected one (Fig. 4). (1) The first dataset, denoted as D_s, contains 4 small statues and artifacts. (2) The second dataset, denoted as D_p, contains two sculpture scenes, which are taken in the Summer Palace with one year interval. (3) The third dataset, denoted as D_b, provides a sequence of 10 sets of laboratory scenes to simulate the deterioration of a mural over time. Each sequence has 7 images with 7 different illuminations, including one environment lighting (EL) and six directional side lightings (DSLs).

(4) Apart from the three datasets, we construct a new dataset that contains 100 scenes, named HCD100 (heritage change detection in 100 scenes). Each scene has a counterpart that was taken a year ago, satisfying the need of detecting the fine-grained changes in the practices of cultural heritage preservation. More details can be found in the next subsection.

The change detection results are evaluated in terms of the F1-measure (F1), recall (Re), precision (Pr), specificity (Sp), false positive rate (FPR), false-negative rate (FNR) and percentage of wrong classifications (PWC).

Additionally, we quantify the registration performance with the AFD score (Tian et al., 2018) based on the average feature-point displacement flow :

$$AFD(f_{\text{ref}}, f_{\text{cur}}) = \frac{1}{m} \sum_{i=1}^{m} \| f_{\text{cur}}^{\text{i}} - f_{\text{ref}}^{\text{i}} \|_2, \tag{17}$$

where f_{ref} and f_{cur} are the matched feature-point coordinates in the reference observation and the current observation, m is the number of matches. Given an image pair, the matched feature points are determined by using SIFT feature descriptor and RANSC robust matching as in Tian et al. (2018) .

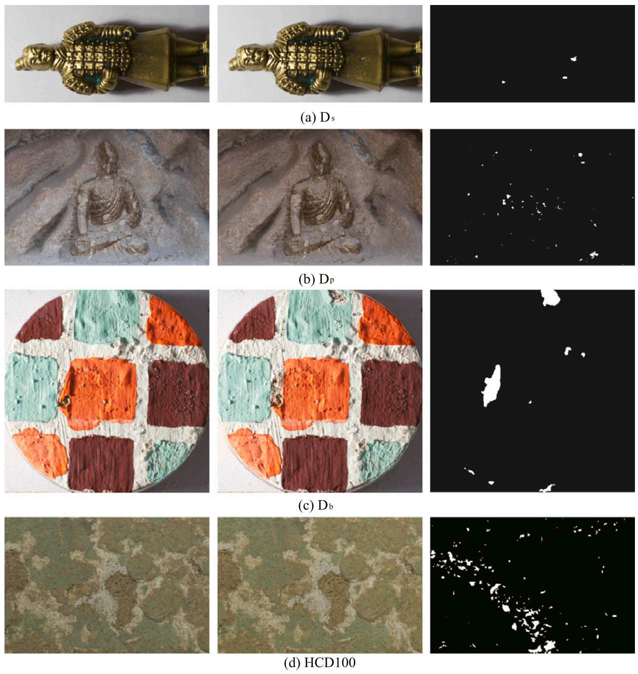


Fig. 4. Some examples in HCD100, D_s, D_p, and D_b dataset. The first and second column are the last observation and current observation. The third column is the GT.

Table 1
HCD100, D_s, D_p, and D_b datasets.

Dataset	Scenes	Heritage sites	Type
HCD100	100	Mogao Grottoes and Terra Cotta Warriors	Ancient mural and painted sculpture
D_s	4	Laboratory scene	Laboratory sculpture
D_p	2	Summer Palace	Sculpture
D_b	10	Laboratory scene	Laboratory testing block

4.2. The proposed HCD100 dataset

Dataset HCD100 is the first dataset focusing on fine-grained change detection of ancient murals. It contains 100 scenes with diverse heritage sites and different degradation sources.

Also note that for the real heritage sites, the changes between two observations of different times can be rather tiny and complex. Figs. 5 and 6 show some cases in our dataset. We compare HCD100 with D_s, D_p, and D_b in Table 1.

High quality. We collected the data in Dunhuang Mo Gao Grotto and Terra Cotta Warriors in Mausoleum of the First Qin Emperor with one-year interval. Empirically, the experts of the heritage sites showed where the representative changing cases are. One year later, we went to the heritage sites again to collect the same scene. We relocated the current camera to the similar pose and position of the reference observation I_{ref} and then captured the current observation image I_{cur}. Note that although active camera relocalization was employed in the data capturing (Feng et al., 2015; Tian et al., 2018), the image pair still has some slight viewpoint differences, and even a tiny viewpoint difference may have a big impact on the fine-grained change detection results, as demonstrated in Fig. 1. Here, our dataset focuses on millimeter-level fine-grained changes in the heritage sites, which can be difficult for experienced experts to directly identify with naked eyes.

Diverse scenes. To make the dataset more valuable and covering diverse scenes, we collect the 100 scenes at different places in the two heritage sites and covering different degradation sources, e.g. cracking and shedding on the surface, etc.

Data annotation. It is a challenging task to annotate the image pairs with many tiny changes. To ensure labeling accuracy, we use cross-validation in the annotation process. Two volunteers annotated all the changes between two observations of different times separately, and we invited a specialist in the field of heritage prevention to judge the differences between their annotations. We finally selected the annotations which are not controversy over the three people.

It is worth mentioning that finding tiny changes on the two images is really difficult. Hence we provided the annotators with the warped current observation via registration methods, which are only used for better guidance, and the annotators made annotation directly on the original reference image. To be more specific, given two images I_{ref} and I_{cur}, we used two registration methods (the one proposed in our paper and that used in FGCD) to warp I_{cur} to I_{ref}. The warped images $R_1(I_{cur})$ and $R_2(I_{cur})$ are aligned with I_{ref} with some little distortions. Next we compared $R_1(I_{cur})$ with I_{ref} to generate GT_1 and compare $R_2(I_{cur})$ with I_{ref} to generate GT_2. The annotator compares GT_1, GT_2 with I_{ref} and I_{cur} to select the more accurate change label, and then does manual annotation on the original reference image. By this way, the annotator generates a ground truth.

Table 2
Registration performance comparison. The best result is shown in bold. More explanation is refereed to in text.

Method	w/o registration	FGCD	Our method
AFD	1.3812	0.6607	**0.5374**

4.3. Registration performance comparison

We first compare the registration results of our phase-based method and that from SIFT flow adopted in FGCD method (Feng et al., 2015) by AFD scores. AFD score measures the average feature-point displacement (Eq. (17)). The low AFD score is better.

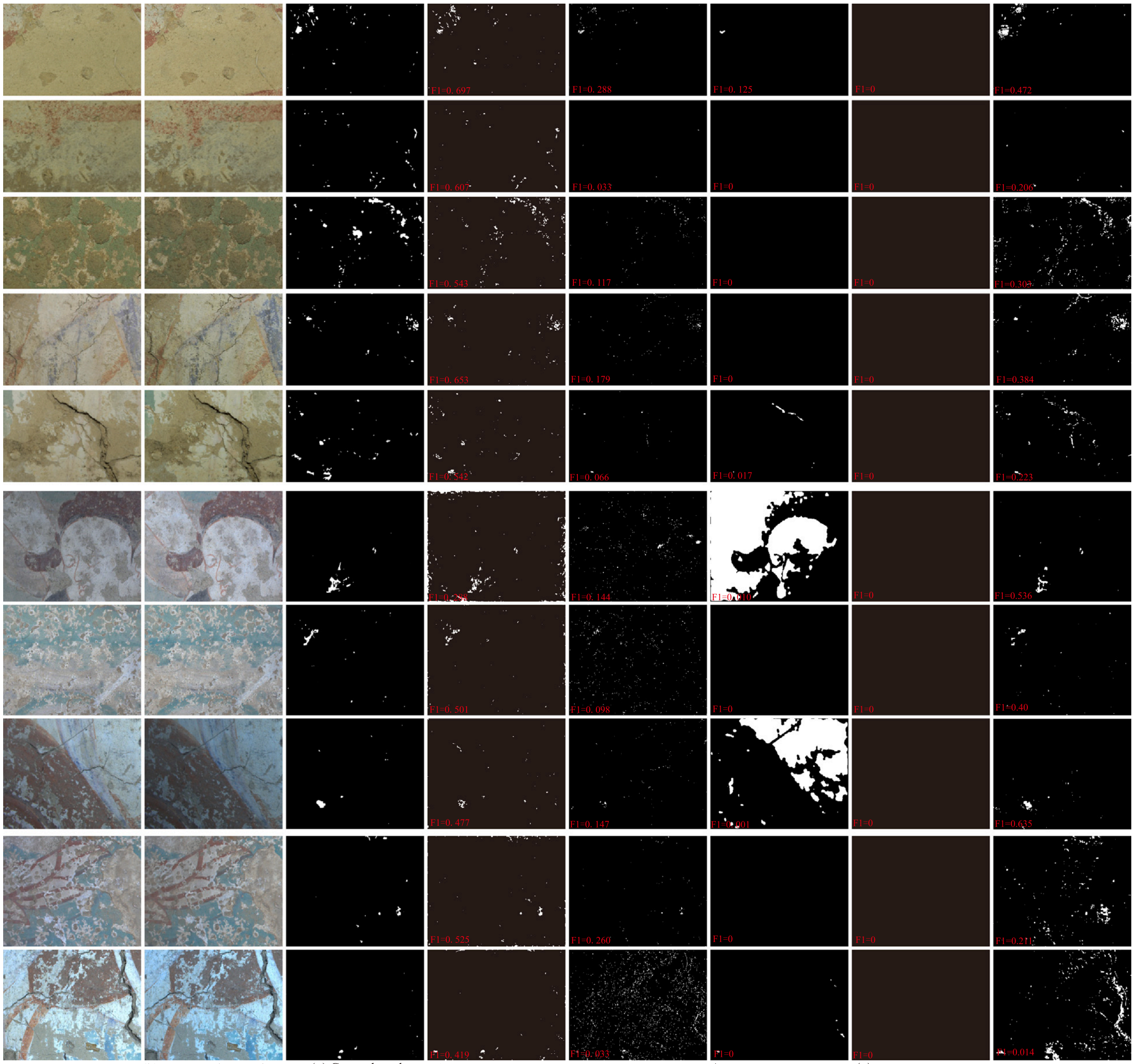
Quantitative comparison results are listed in Table 2. 'w/o registration' represents that last observation and current observation are not registered. 'FGCD' represents that we use the registration step of FGCD to align the last observation with the current observation. 'Our method' represents that we use the phase-based registration method. The AFD score of the unregistration image pair is 1.3812. The AFD score of our method is 0.5374, while the AFD score of FGCD is 0.6607. In comparison, our method achieves the highest registration performance.

4.4. Comparisons with SOTA methods

4.4.1. Visual comparison

Fig. 5 visually compares our method with the state-of-the-art methods, including FGCD, subSENCE, NPSG and ChangeNet, on ten different scenes. Each scene contain two observations of different time, Ground Truth change map, change detection results of three state-of-the-art methods and change detection results of our method. Limited to the space of this paper and the tiny changes in mural scenes, the changes can be difficultly observed by the naked eye. We also provide the zoomed-in version in Fig. 6. The tiny changes can be obviously observed by the zoomed-in patches.

In comparison, subSENCE is sensitive to lighting differences, and predicts large changed regions for the sixth and eighth scenes, which exhibit a little bit obvious difference in lighting conditions. ChangeNet fails to detect such subtle changes for all the tenth scenes. FGCD reports better detection results among the three state-of-the-art methods, however, it still wrongly estimates or misses changed regions. FGCD detects the changes by the intensity difference between the last observation and adjusted current observation, However, the changes in the mural scene may have low intensity difference as shown in Figs. 5 and 6. Our method detects the changes by structure information. Structure information is irrelevant to lighting conditions and more sensitive to structural changes. It is clear that the change detection results of our method look much close to the ground truth, and surpass other methods with large margins in terms of F1 measure (the red numbers in the left-bottom of figures).



(a) Last observation I1 (b) Current observation I2 (c) Ground truth (d) Result of pb-FGCD (e) Result of FGCD (f) Result of subSENCE (g) Result of ChangeNet (h) Results of NPSG

Fig. 5. Ten scenes in HCD100.

Table 3
Results on HCD100 dataset. The best results are shown in bold. More explanation is refereed to in text.

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
PCA-K-Means (Celik, 2009)	0.04	0.53	0.03	0.58	0.42	0.47	42.4
ChangeNet (Varghese et al., 2018)	0.00	0.001	0.00	0.89	0.11	0.99	10.79
SubSENCE (St-Charles et al., 2014)	0.08	0.08	0.27	0.98	0.02	0.92	2.73
FGCD (Feng et al., 2015)	0.19	0.25	0.39	0.99	0	0.75	1.11
NPSG (Sun et al., 2021)	0.26	0.50	0.28	0.97	0.03	0.52	2.80
pb-FGCD	**0.51**	**0.54**	**0.54**	**1**	**0**	**0.46**	**0.63**

4.4.2. Quantitative comparison

Tables 3–6 show the average quantitative performance of different methods on the four datasets. Our method is denoted as pb-FGCD. For the comparison methods, we have tried a series of reasonable parameters, and use the best one in our evaluation.

Dataset HCD100 has finer textures and more subtle changes compared with D_s, D_p, and D_b. It requires more accurate registration ability and it is more challenge to detect the tiny changes from different lighting conditions. We can clearly see that our approach increases the F1-measure by more than 190% over the second-best method NPSG on HCD100 dataset. The proposed method outperforms the FGCD by a large margin due to two reasons: (1) our method can align the two observations of different times more accurately as shown in Table 2, which is vital for locating and detecting the changing area. (2) The lighting compensation based methods detect the changes by intensity difference. FGCD struggles at the changes with low intensity difference. Our method uses structure information which is irrelevant to lighting difference. And structure information is sensitive to structure changes, which is vital for detecting the fine-grained changes with low intensity difference.

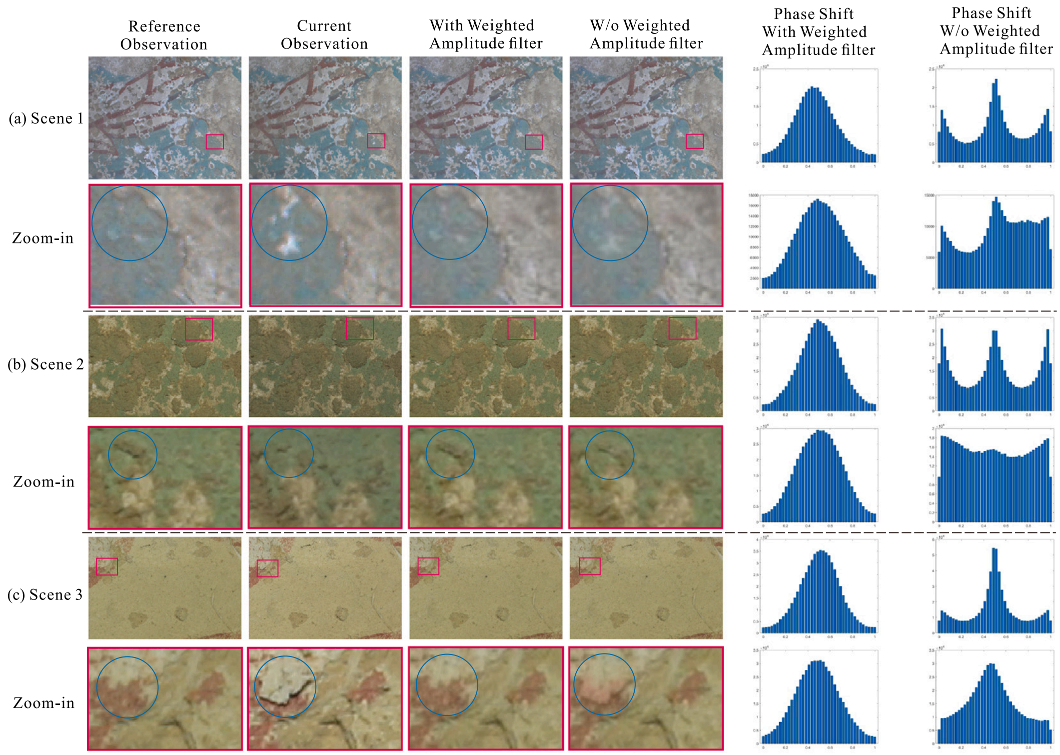


Fig. 6. Registration results with/without weighted amplitude filter and the histogram of phase shift with/without weighted amplitude filter.

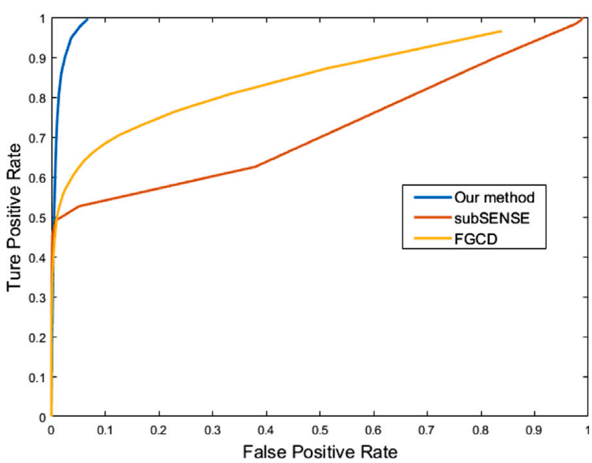


Fig. 7. ROC curves of pb-FGCD, FGCD and subSENCE on datasets HCD100.

Table 4
Results on Statues dataset D_s. The best results are shown in bold. More explanation is refereed to in text.

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
SubSENCE A	0.02	0.83	0.01	0.88	0.12	0.17	12.13
SubSENCE M	0.01	**0.98**	0.00	0.66	0.34	**0.02**	34.28
SubSENCE LFA	0.27	0.28	0.34	0.99	0.01	0.72	1.57
SubSENCE LFM	0.12	0.77	0.07	0.95	00.05	0.23	5.37
FGCD(D&T)	0.53	0.78	0.43	**1.00**	**0.00**	0.22	0.28
FGCD(SVM)	0.51	0.86	0.39	**1.00**	**0.00**	0.14	0.29
NPSG	0.42	0.59	0.45	**1.00**	**0.00**	0.41	0.28
pb-FGCD	**0.70**	0.74	**0.70**	**1.00**	**0.00**	0.26	**0.08**

We further compare the average ROC curves between pb-FGCD, FGCD and subSENCE as shown in Fig. 7. It is obvious that our method outperforms FGCD method with large margins.

Table 5
Results on Statues dataset D_p. The best results are shown in bold. More explanation is refereed to in text.

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
SubSENCE A	0.02	0.55	0.02	0.60	0.40	0.45	39.41
SubSENCE M	0.02	**0.93**	0.01	0.20	0.80	**0.07**	78.85
SubSENCE LFA	0.08	0.04	0.05	0.99	0.01	0.96	1.81
SubSENCE LFM	0.07	0.22	0.04	0.95	0.05	0.78	5.85
FGCD(D&T)	0.34	0.28	0.52	**1.00**	**0.00**	0.72	**0.92**
FGCD(SVM)	0.51	0.53	0.47	0.99	0.01	0.47	1.02
NPSG	0.22	0.16	0.33	1.00	0.00	0.84	0.73
pb-FGCD	**0.52**	0.48	**0.58**	**1.00**	**0.00**	0.52	0.94

Table 6
Results on Statues dataset D_b. The best results are shown in bold. More explanation is refereed to in text.

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
SubSENCE A	0.24	0.50	0.31	0.72	0.28	0.50	28.32
SubSENCE M	0.23	**0.67**	0.19	0.66	0.34	**0.33**	34.01
SubSENCE LFA	0.06	0.03	0.26	**1.00**	**0.00**	0.97	1.43
SubSENCE LFM	0.28	0.21	0.50	0.99	0.01	0.79	1.62
FGCD(D&T)	0.45	0.40	**0.56**	**1.00**	**0.00**	0.60	1.23
FGCD(SVM)	**0.53**	0.62	0.48	0.99	0.01	0.38	1.41
NPSG	0.23	0.55	0.16	0.97	0.03	0.45	3.89
pb-FGCD	0.51	0.56	0.47	0.93	**0.00**	0.44	**1.22**

For D_p, D_m and D_b dataset, we also include possible variants of SubSENCE, NPSG and FGCD. SubSENCE reports improved results on D_b than D_s and D_p. NPSG shows a better results on D_p than D_b and D_s. For the two variants of FGCD, the one using SVM gets better recall and PWC values, while the one using D&T reports better Pr, Sp, FRR and FNR values. In comparison, our method ph-FGCD outperforms FGCD by a large margin on D_s, gets better results than FGCD on D_p, and achieves competitive results on D_b. Our method obtains an overall better performance on D_p, D_m and D_b dataset compared with other methods.

Table 7
The time complexity of different methods.

Method	Time complexity	Running time (s)
PCA-K-Means (Celik, 2009)	$O(HW)$	3.9
SubSENCE (St-Charles et al., 2014)	$O(HW)$	98.5
FGCD (Feng et al., 2015)	$O(H^2W^2)$	58.7
NPSG (Sun et al., 2021)	$O(H^2W^2)$	162.1
pb-FGCD	$O(HWlog(HW))$	53.4

Table 8
Change detection result with and without weighted amplitude filtering on HCD100, D_p, D_m and D_b. The best results are shown in bold.

Method	Dataset	F1	Re	Pr	Sp	FPR	FNR	PWC
W/o filter	HCD100	0.21	0.14	**0.74**	1.00	0.00	0.86	0.63
With filter	HCD100	**0.51**	**0.54**	0.54	**1.00**	**0.00**	**0.46**	**0.63**
W/o filter	D_s	0.42	0.72	0.39	0.99	0.01	0.28	1.29
With filter	D_s	**0.70**	**0.74**	**0.70**	**1.00**	**0.00**	**0.26**	**0.08**
W/o filter	D_p	0.21	0.25	0.19	0.98	0.02	0.75	3.87
With filte	D_p	**0.52**	**0.48**	**0.58**	**1.00**	**0.00**	**0.52**	**0.94**
W/o filter	D_b	0.28	0.46	0.23	**0.95**	0.05	0.54	5.40
With filter	D_b	**0.51**	**0.56**	**0.47**	0.93	**0.00**	**0.44**	**1.22**

4.4.3. Running time

We report the running time for fine-grained change detection between a pair of 1280*720 images. All measurements were performed on a standard desktop (Intel Core i7, 16 GB memory).

Our method spends an average time of 53.4 s, Most of our time is taken in the process of registration. Since our method tackles geometry difference and lighting difference in the spatial-frequency domain, the main computation cost is Fourier transform which has a fast algorithm FFT, and the computational complexity of FFT is $O(HWlog(HW))$.

The time complexity and running time of comparison methods are shown in Table 7. Time complexity is the amount of the time taken by an algorithm to run, as a function of the length of the input. Since this time complexity is generally difficult to compute exactly, and the running time for small inputs is usually not consequential, one commonly focuses on the behavior of the complexity when the input size increases. The time complexity of SubSENCE (St-Charles et al., 2014) is lower than other methods. However, it is the second-worst method according to running time taking 98.5 s. As SubSENCE (St-Charles et al., 2014) forms the background model of each scene by a set of 50 background samples to improve the robustness to lighting variations. According to Cormen et al. (2022), the lower order terms and the constant multiplier of the highest order term are ignored for calculating time complexity. The constant multiplier '50' is ignored when calculating time complexity. For computing the running time, the multiplier '50' has a quite large impact.

4.5. Ablation study

To evaluate the effectiveness of the key components of our method, we do ablation studies in the following.

The effectiveness of weighted amplitude filter. We first conduct ablation study of weighted amplitude filter on dataset HCD100. Results are reported in Table 8. Without the weighted amplitude filter, we obtain F1-measure of 0.21, recall value of 0.14, precision of 0.74. By using the weighted amplitude filter, our approach can boost the change detection performance greatly, by reporting F1-measure of 0.51 (143% higher), recall value of 0.54. We can also observe that weighted amplitude filtering achieves consistent improvement on D_p, D_m and D_b dataset.

It is because that weighted amplitude filter can preserve the changing area when we warp the reference observation to the current observation.

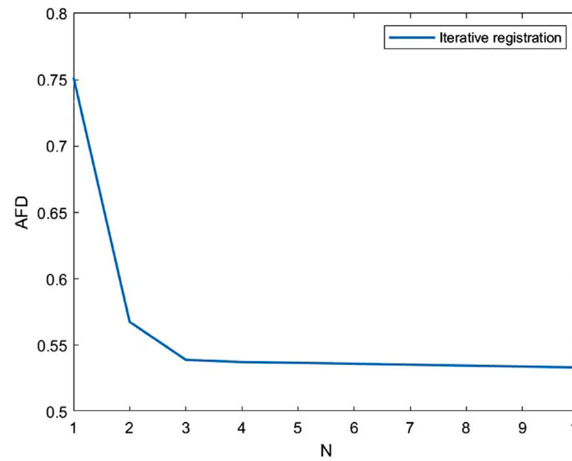


Fig. 8. Quantitative comparison of iterative registration.

Table 9
Change detection results with and without structure information on HCD100. The best results are shown in bold.

Method	F1	Re	Pr	Sp	FPR	FNR	PWC
W/o	0.10	0.36	0.13	0.87	0.13	0.64	13.07
With	**0.51**	**0.54**	**0.54**	**1**	**0**	**0.46**	**0.63**

The effectiveness of iterative alignment strategy. Fig. 8 shows the results with respect to the different numbers of iterations of registration. As shown in the figure, iterative alignment achieves consistent performance gains.

The AFD score tends to be stable for 4-iterations, and hence we adopt 4 iterations in all our experiments.

The effectiveness of structure information. Here we carry out more detailed explorations on our component design without the structure information on HCD100. It can be seen in Table 9 that without structure information (i.e. using Eq. (11) only to estimate changes in the spatial domain) introduces a big decrease to the change detection results.

4.6. Discussion

4.6.1. Parameter determination

Our method has four parts of hyperparameters. (1) The first part is the phase-based registration parameters, the scales ω=4 and orientations θ=8 used to define the complex steerable pyramid (Meyer et al., 2015; Wadhwa et al., 2013);

(2) The second part is parameters in the weighted amplitude filtering and iterative alignment, including the displacement factor γ, the Gaussian kernel parameter ρ and the iteration number N. We set γ=0.5 for achieving the reasonable displacement manipulated by the phase and preservation of details. As for Gaussian kernel parameter ρ, Fig. 9 left shows the F1-measure for ρ ranging from 0 to 9. Since F1-measure tends to be stable for $\rho = 6$, we set $\rho = 6$ in our experiments. As for iteration number N=4, Fig. 8 shows the results of the different numbers of iterations of registration. Iterative alignment achieves consistent performance gains, and the AFD score tends to be stable for 4-iterations.

(3) The third part is the structure information extraction parameter, i.e. the spatial threshold ζ. We follow the work of Sampat et al. (2009), Zhang et al. (2011). (4) The fourth part is the change detection part parameter, i.e. the structure information threshold δ. To determine an optimal threshold δ for structure information, we compute the evolution of the F1-measure of our method shown in Fig. 9 right. The evolution is computed in a subset that contains 10 scenes in HCD100. According to the figure, we set δ=0.3 in Eqs. (11).

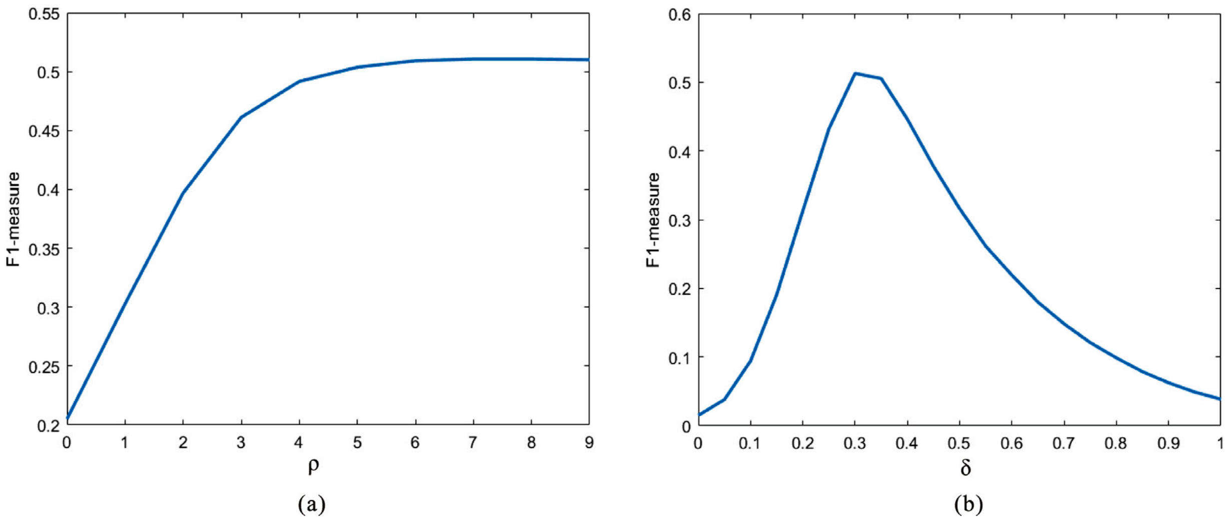


Fig. 9. (a) F1-measure for ρ ranging from 0 to 9. (b) F1-measure for δ ranging from 0 to 1.

4.6.2. Analysis of weighted amplitude filter

Here we illustrate the phase shift in terms of the histogram to demonstrate why weighted amplitude filter works. We show 3 scenes in Fig. 6. Each scene has the reference observation, the current observation, the registration result with/without using the weighted amplitude filter, the histogram of their phase shift with respect to the current observation, and the zoom-in regions. Note that the complex steerable pyramid contains multiple frequency bands, and here we randomly select two bands to compute the histograms. Intuitively, for the well-aligned images, the phase shift will be close as nearby pixels have similar shifts.

We can observe in the rightmost column in each scene that there are many values close to 0 and 1, which is introduced by the change of pixel values due to unalignment. When we use the weighted amplitude filter to process the phase variation, it can be seen in the fifth volume that the phase shift now becomes more concentrated in the middle range. In other words, by using the weighted amplitude filter, we let more pixels get a more accurate phase shift. As a result, we can align the image pair while preserving the changing area.

5. Conclusion

In this paper, we proposed a novel method for fine-grained change detection based on the extensions of the phase-based method, to compensate for geometry difference and lighting difference. Relying on the complex steerable pyramid, we introduce the weighted amplitude filter and iterative alignment strategy for accurate registration, and estimate changes by fusing both spatial domain and frequency domain estimations. In addition, we propose a real-world dataset for different scenes under varying lighting conditions. We make a comparison with the state-of-the-art fine-grained change detection methods. Our approach increases the F1-measure by more than 190% over the state-of-the-art method, and our method is much faster.

CRediT authorship contribution statement

Xuzhi Wang: Methodology, Writing, Experiment conduction, Visualization. **Liang Wan:** Methodology, Writing, Supervision. **Di Lin:** Revision. **Wei Feng:** Methodology, Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in parts by the National Key Research and Development Program of China (2020YFC1522703), and the National Natural Science Foundation of China (62072334).

Appendix

In this section, we give an intuition for phase-based image registration. An example is shown in Fig. 10(a), in which a 1D image intensity profile undergoes global translation over time. Then, we explain the limitation of our method as demonstrated in 10 (b).

Let one-dimensional sinusoidal function defined as:

$$y_1 = Asin(\omega x - \phi_1), \tag{18}$$

where A is the amplitude, ω is the angular frequency and *phi* is the phase. A translation of this function can be described by modifying the phase, yielding a second sinusoidal function:

$$y_2 = Asin(\omega x - \phi_2). \tag{19}$$

We can align y_2 to y_1 by modifying the phase variations between ϕ_1 and ϕ_2, i.e.

$$\phi_{var} = \phi_1 - \phi_2, \tag{20}$$

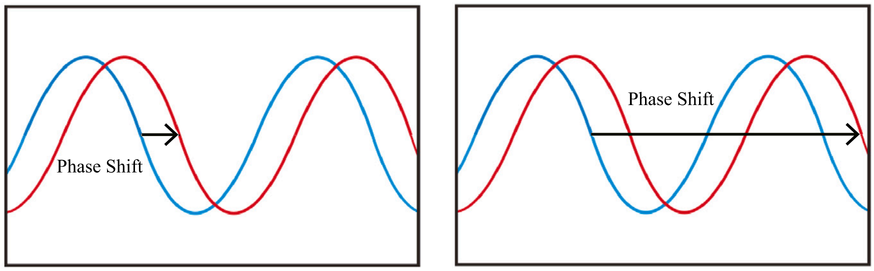
$$y_2 = Asin(\omega x - \phi_2 - \phi_{var}). \tag{21}$$

Now, considering the general functions $f(x)$ translated by a displacement function $\sigma(x)$, this can be represented in the Fourier domain as a sum of complex sinusoids over all frequencies ω:

$$f(x + \sigma(t)) = \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega(x+\sigma(t))}, \tag{22}$$

in which each band corresponds to a single frequency ω.

In general, motions in two observations of different times are local not global. We use the complex steerable pyramid to deal with local motions for registration. The scheme has limitation to handle large displacements. Since large displacements correspond to a phase difference more than the periodicity of the phase value, shifting can lead to a phase ambiguity. An example is shown in Fig. 10(b), where the actual displacement is larger than the computed phase. Also note that



(a) (b)

Fig. 10. Illustration of phase-based image registration.

this limitation can be alleviated by the pre-registration step. We apply the affine transformation whose parameters are calculated by matching SIFT descriptor and RANSAC (Fischer et al., 1981; Lowe, 2004) for pre-registration.

References

Barron, J., & Malik, J. (2015). Shape, illumination, and refectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(8), 1670–1687.

Bescós, J. (2004). Real-time shot change detection over online MPEG-2 video. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(4), 475–484.

Bouwmans, T., Sobral, A., Javed, S., Jung, S. K., & Zahzah, E.-H. (2017). Decomposition into low-rank plus additive matrices for background/forground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, *23*, 1–71.

Brox, T., & Malik, J. (2011). Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(3), 500–513.

Celik, T. (2009). Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters*, *6*(4), 772–776.

Chen, M., Wei, X., Yang, Q., Li, Q., Wang, G., & Yang, M.-H. (2017). Spatiotemporal GMM for background subtraction with superpixel hierarchy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1518–1525.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms*. MIT Press.

Davis, A., Bouman, K. L., Chen, J. G., Rubinstein, M., Durand, F., & Freeman, W. T. (2015). Visual vibrometry: Estimating material properties from small motions in video. In *IEEE conference on computer vision and pattern recognition*.

Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G. J., Durand, F., & Freeman, W. T. (2014). The visual microphone: passive recovery of sound from video. *ACM Transactions on Graphics*, *33*(4), 79.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazırbas, C., & Golkov, V. (2015). FlowNet: Learning optical flow with convolutional networks. In *IEEE international conference on computer vision*.

Elgammal, A., Harwood, D., & Davis, L. (2000). Non-parametric model for background subtraction. In *European conference on computer vision*.

Elgharib, M. A., Hefeeda, M., Durand, F., & Freeman, W. T. (2015). Video magnification in presence of large motions. In *IEEE conference on computer vision and pattern recognition*.

Feng, W., Tian, F.-P., Zhang, Q., Zhang, N., Wan, L., & Sun, J. (2015). Fine-grained change detection of misaligned scenes with varied illuminations. In *IEEE international conference on computer vision*.

Fischer, Bolles, M. A., & Robert, C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.

Fourure, D., Emonet, R., Fromont, E., Muselet, D., Trémeau, A., & Wolf, C. (2016). Mixed pooling neural networks for color constancy. In *International conference on image processing*.

Gargi, U., Kasturi, R., & Strayer, S. H. (2000). Performance characterization of video-shot-change detection methods. *IEEE Transactions on Circuits and Systems for Video Technology*, *10*(1), 1–13.

Gijsenij, A., Gevers, T., & van de Weijer, J. (2011). Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, *20*(9), 2475–2489.

Hauagge, D., Wehrwein, S., Bala, K., & Snavely, N. (2013). Photometric ambient occlusion. In *IEEE conference on computer vision and pattern recognition*.

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*(1–3), 185–203.

Hou, A., Zhang, Z., Sarkis, M., Bi, N., Tong, Y., & Liu, X. (2021). Towards high fidelity face relighting with realistic shadows. In *IEEE conference on computer vision and pattern recognition*.

Hui, T.-W., Tang, X., & Loy, C. C. (2018). LiteFlowNet: A lightweight convolutional neural network for optical flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., & Brox, T. (2017). FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *IEEE conference on computer vision and pattern recognition*.

Lin, Y., Tong, Y., Cao, Y., Zhou, Y., & Wang, S. (2017). Visual-attention-based background modeling for detecting infrequently moving objects. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(6), 1208–1221.

Liu, C., Yuen, J., Torralba, A., & Freeman, W. T. (2011). SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(5), 978–994.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Luan, F., Pairs, S., Shechtman, E., & Bala, K. (2017). Deep photo style transfer. In *IEEE conference on computer vision and pattern recognition*.

Meyer, S., Djelouah, A., McWilliams, B., Sorkine-Hornung, A., Gross, M., & Schroers, C. (2018). PhaseNet for video frame interpolation. In *IEEE conference on computer vision and pattern recognition*.

Meyer, S., Wang, O., Zimmer, H., Grosse, M., & Sorkine-Hornung, A. (2015). Phase-based frame interpolation for video. In *IEEE conference on computer vision and pattern recognition*.

Nguyen, T. P., Pham, C. C., Ha, S. V.-U., & Jeon, J. W. (2004). Change detection by training a triplet network for motion feature extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(4), 475–484.

Ranjan, A., & Black, M. J. (2017). Optical flow estimation using a spatial pyramid network. In *IEEE conference on computer vision and pattern recognition*.

Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., & Markey, M. K. (2009). Complex wavelet structural similarity: A new image similarity index. *IEEE Transactions on Image Processing*, *18*(11), 2385–2401.

St-Charles, P.-L., Bilodeau, G.-A., & Bergevin, R. (2014). SuBSENSE : a universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, *24*(1), 359–373.

Stent, S., Gherardi, R., Stenger, B., & Cipolla, R. (2016). Precise deterministic change detection for smooth surfaces. In *IEEE winter conference on applications of computer vision*.

Sun, Y., Lei, L., Li, X., Sun, H., & Kuang, G. (2021). Nolocal patch similarity based heterogeneous remote sensing change detection. *Pattern Recognition*, *109*.

Sun, D., Yang, X., Liu, M.-Y., & Kautz, J. (2018). PWC-net: CNNs for optical flow using pyramid, warping, and cost volume. In *IEEE conference on computer vision and pattern recognition*.

Takeda, S., Okami, K., Mikami, D., Isogai, M., & Kimata, H. (2018). Jerk-aware video acceleration magnification. In *IEEE conference on computer vision and pattern recognition*.

Tian, F.-P., Feng, W., Zhang, Q., Wang, X., Sun, J., Loia, V., & Liu, Z.-Q. (2018). Active camera relocalization from a single reference image without hand-eye calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(12), 2791–2806.

Urhan, O., Güllü, M. K., & Ertürk, S. (2006). Modified phase-correlation based robust hard-cut detection with application to archive film. *IEEE Transactions on Circuits and Systems for Video Technology*, *16*(6), 753–770.

Varghese, A., Gubbi, J., Ramaswamy, A., & Balamuralidhar, P. (2018). ChangeNet: A deep learning architecture for visual change detection. In *European conference on computer vision workshop*.

Wadhwa, N., Rubinstein, M., Durand, F., & Freeman, W. T. (2013). Phase-based video motion processing. *ACM Transactions on Graphics*, *32*(4), 1–10.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Wang, X., Lin, D., & Wan, L. (2022). FFNet: Frequency fusion network for semantic scene completion. In *AAAI conference on artificial intelligence*.

Weinzaepfel, P., Revaud, J., Harchaoui, Z., & Schmid, C. (2013). DeepFlow: Large displacement optical flow with deep matching. In *IEEE international conference on computer vision*.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, *31*(4), 1–8.

Yang, K.-F., Gao, S.-B., & Li, Y.-J. (2015). Efficient illuminant estimation for color constancy using grey pixels. In *IEEE conference on computer vision and pattern recognition*.

Zhang, Y., Pintea, S. L., & van Gemert, J. C. (2017). Video acceleration magnification. In *IEEE conference on computer vision and pattern recognition*.

Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, *20*(8), 2378–2386.

Zhao, F., Li, J., Zhao, J., & Feng, J. (2018). Weakly supervised phrase localization with multi-scale anchored transformer network. In *IEEE conference on computer vision and pattern recognition*.

Zhao, J., Li, J., Zhao, F., Yan, S., & Feng, J. (2017). Marginalized CNN: Learning deep invariant representations. In *British machine vision conference*.

Zhou, L., Zhao, J., Li, J., Yuan, L., & Feng, J. (2018). Object relation detection based on one-shot learning. arXiv:1807.05857.