Fine-Grained Change Detection of Misaligned Scenes with Varied Illuminations

Wei Feng[1]*, Fei-Peng Tian[1,2], Qian Zhang[1], Nan Zhang[1], Liang Wan[2], Jizhou Sun[1]
[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] School of Computer Software, Tianjin University, Tianjin, China
{wfeng,tianfeipeng,qianz,nzh_cs,lwan,jzsun}@tju.edu.cn

Abstract

Detecting fine-grained subtle changes among a scene is critically important in practice. Previous change detection methods, focusing on detecting large-scale significant changes, cannot do this well. This paper proposes a feasible end-to-end approach to this challenging problem. We start from active camera relocation that quickly relocates camera to nearly the same pose and position of the last time observation. To guarantee detection sensitivity and accuracy of minute changes, in an observation, we capture a group of images under multiple illuminations, which need only to be roughly aligned to the last time lighting conditions. Given two times observations, we formulate fine-grained change detection as a joint optimization problem of three related factors, i.e., normal-aware lighting difference, camera geometry correction flow, and real scene change mask. We solve the three factors in a coarse-to-fine manner and achieve reliable change decision by rank minimization. We build three real-world datasets to benchmark fine-grained change detection of misaligned scenes under varied multiple lighting conditions. Extensive experiments show the superior performance of our approach over state-of-the-art change detection methods and its ability to distinguish real scene changes from false ones caused by lighting variations.

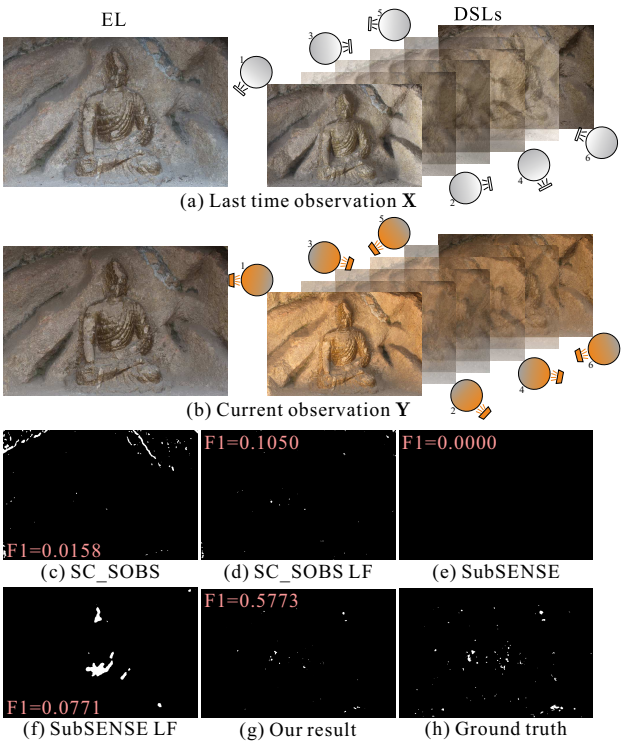

Figure 1. A real case of fine-grained change detection of a small Buddha sculpture in the Summer Palace. See text for details.

1. Introduction

Change detection is a fundamental low-level vision problem and is broadly useful in many real-world vision applications, like video surveillance, tracking, segmentation, and remote sensing, as a critical preprocessing step [20, 22, 5, 6, 7, 12]. The major challenge of change detection is to separate real scene changes from false changes caused by different imaging conditions, e.g., suddenly varied lightings and camera movements. To tackle this problem, most state-of-the-art change detection methods assume real scene changes occur in a relatively large-scale and are salient enough to transcend detailed changes caused by illumination or camera variations [10, 16, 17, 25, 26, 21].

However, some recent applications, such as high-value object monitoring, need to accurately discover and locate fine-grained subtle changes within a scene. For instance, an essential problem in preventive conservation of cultural heritage is to detect and measure the tiny changes of cultural relics from two sets of observations within long time intervals [23, 27].[1] Other examples include biomedical di-

*is the corresponding author. Tel: (+86)-22-27406538.

[1] As most important relics are properly protected, their states are usually very stable, thus their monitoring interval can be as long as a year.

agnosis and allimportant buildings (e.g., dam) monitoring, which all requires reliable detection of fine-grained changes of a scene under not strictly registered cameras and varied illuminations. Clearly, for such scenarios, state-of-the-art change detection methods cannot work well.

Fig. 1 shows a real case of status monitoring of a small golden Buddha sculpture in the Summer Palace. Fig. 1(a) and (b) are the first- and second-time observations taken at Dec. 7, 2013 and 2014, respectively. To capture enough details of the scene, we collected 7 images under 1 environment lighting (EL) and 6 directional side lightings (DSLs). The colors and directions of DSLs are also labeled along with the DSL images. We can see the two observations have different ELs, and 6 misaligned DSLs (in lighting color, size and direction). As shown in the ground truth Fig. 1(h), some fine-grained changes had occurred during one year time, e.g., a tiny part of gold lacquer in the Buddha's right arm had dropped. Fig. 1(c)-(f) are the results of two state-of-the-art change detection methods, SC_SOBS [17, 16] and SuBSENSE [22], with (labeled as LF) or without the proposed lighting and camera geometry corrections.[2] It is evident that state-of-the-art change detection methods cannot reliably detect fine-grained changes of a scene, especially under unregistered camera geometries and varied illuminations. Both their pixel-level F1-measures are lower than 0.1. Although in Fig. 1(d) and (f), they can obtain better results when the lighting and camera geometry differences are corrected (with F1-measure 0.1050 for SC_SOBS and 0.0771 for SuBSENSE), their detection accuracies to fine-grained changes are still unacceptable. In contrast, the proposed approach can satisfactorily locate real scene changes with pixel-level F1-measure 0.5773, see Fig. 1(g).

The major obstacle of previous successful change detection methods to detecting fine-grained changes of a scene lies in two aspects. First, their basic assumption about real scene changes being both spatially and photometrically significant makes them difficult to discover minute and subtle changes. Second, most of them are supposed to handle video data in near real-time speed, thus they generally cannot tolerate severe illumination and camera variations. However, in fine-grained change detection for high-value object status monitoring, accuracy and robustness to both lighting and camera geometry variations are far more important than real-time response.

This paper studies this challenging problem and proposes a practical low-cost solution. We start from current observation data $\mathbf{Y}$ collection by active camera relocation that quickly relocates current camera to nearly the same camera geometry (i.e., pose and position) of the last time observation $\mathbf{X}$. To guarantee the detection sensitivity and accuracy of fine-grained subtle changes, we capture a group of images under $K+1$ illuminations, including 1 EL and K varied DSLs, to form one time observation.[3] For practical and low-cost purpose, the EL and multiple DSLs only need to be roughly aligned to the last time lightings, see Fig. 1(a)-(b). We formulate fine-grained change detection from $\mathbf{X}$ and $\mathbf{Y}$ as a joint optimization problem of three related factors: normal-aware lighting difference $\mathbf{L}$, camera geometry correction flow $\mathbf{F}$, and real scene change map $\mathbf{C}$. We solve the three factors in a coarse-to-fine manner and achieve reliable change decision by rank minimization. To benchmark the performance of different methods, we build three real-world datasets for fine-grained change detection under not strictly registered cameras and misaligned multiple lighting conditions. Experiments validate the superiority of our approach over state-of-the-art methods.

2. Related Work

Change detection. Our work is closely related to traditional change detection that usually acts as an important preprocessing step for many high-level vision applications, such as video surveillance [16], urban environment monitoring [25, 26, 21], remote sensing [30, 9], and automatic driving [4]. A recent survey [20] systematically summarizes state-of-the-art processing steps and decision rules in change detection. CDNet provides a realistic large-scale benchmark video dataset [10] and has maintained an active rank list of change detection algorithms. On CDNet, background modeling is one of the most successful strategy, based on which many recent algorithms are proposed, such as SOBS [16], SC_SOBS [17], SuBSENSE [22]. Other notable recent developments include 3D voxel based change detection [19, 3] and city-scale structural change detection using multiple panoramic images together with 3D depth data [25, 26, 21]. To tackle illumination variations and camera movements, state-of-the-art methods focus on detecting spatially and photometrically salient changes and implicitly (or explicitly) treat small subtle changes as noises [20]. However, this basic assumption highly limits their ability to fine-grained change detection that may occur at smaller scale and be photometrically less significant (see Fig. 1). In this paper, we show that low-rank analysis [28, 15, 13] can be used to decompose sparse changes from multiple images.

Color constancy. In change detection, fast color constancy is widely used to correct lighting variations. Readers can refer to [8] for a survey of state-of-the-art color constancy methods. Due to the real-time speed requirement, most change detection methods can only afford to use sim-

[2] We choose SC_SOBS [17] and SuBSENSE [22] as baselines because they are top-ranking methods on conventional change detection benchmark dataset [10] and are source code available.

[3] Another possible way is to capture multiple images from different viewpoints, based on which to detect changes in 3D structure and appearance [19, 3]. But, explicit/implicit 3D model based change detection is either too restrictive to handle varied illuminations and camera geometry [19] or not reliable enough to discover pixel-level minute changes [3].

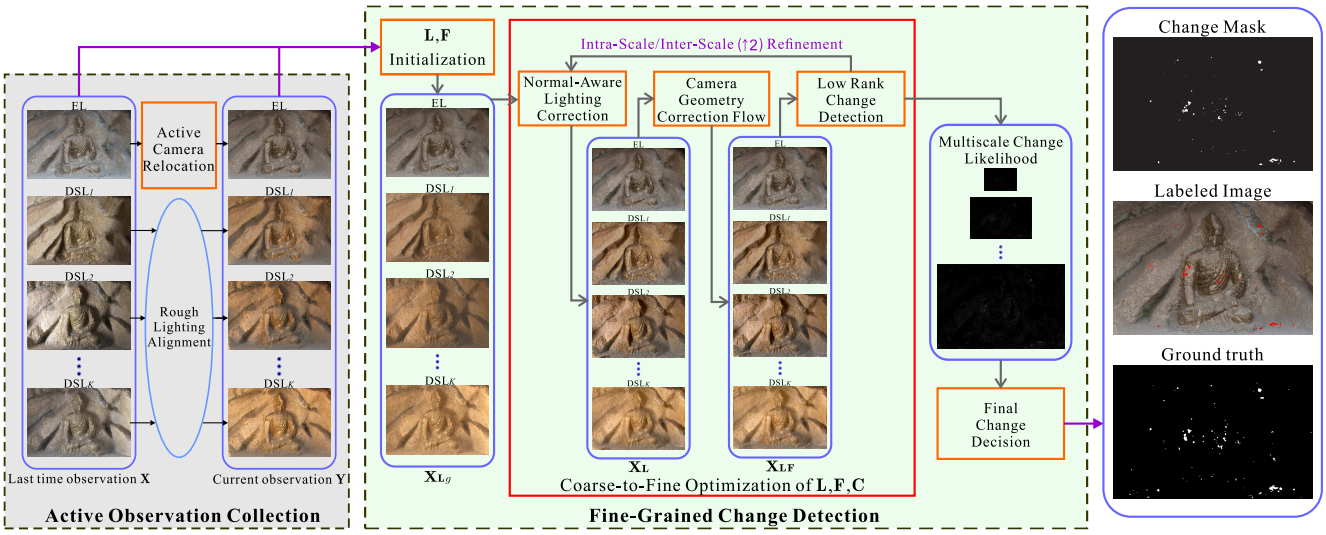

Figure 2. Working flow of the proposed fine-grained change detection approach. See text for details.

ple static color constancy processing, which restrains their ability to tolerate frequent and severe illumination variances. Besides, intrinsic image can be used to correct different lightings of a same scene [31, 18, 2]. Recently, intrinsic decomposition has been further extended to multiple components, including shape, normal, illumination and reflectance, from single or multiple images [1, 11]. However, these recent developments either require sophisticated optimization or multiple images capturing under dense lighting conditions, which make them not directly applicable to low-cost end-to-end fine-grained change detection.

Geometry correction. Geometry correction is another indispensable part in robust change detection. Typical methods include similarity, affine, or projective transformations for rigid scenes [20]. For non-rigid and dynamic scenes, optic flow can be used to correct misaligned camera geometries [24, 29]. Our approach extends SIFT flow [14] by considering multi-lighting constraints.

3. Fine-Grained Change Detection

Our major objective is to accurately discover and locate fine-grained changes occurring within high-value scenes, e.g., the cultural relic in Fig. 1. The working flow of our approach is shown in Fig. 2, which provides an end-to-end solution to fine-grained change detection. To guarantee detection sensitivity and accuracy, we start from active camera relocation to collect reliable observation data with slightly misaligned camera geometries and varied multiple illuminations. We then iteratively optimize normal-aware lighting difference $\mathbf{L}$, camera geometry correction flow $\mathbf{F}$ and real scene change map $\mathbf{C}$ in a coarse-to-fine manner.

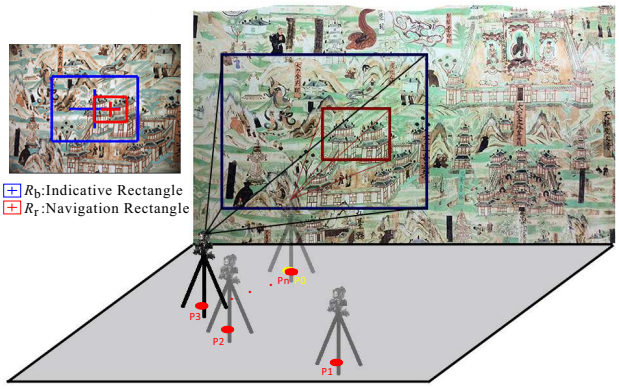

Figure 3. Illustration of active camera relocation.

3.1. Active camera relocation

Our first step is *active camera relocation*. Given last time observation matrix $\mathbf{X} = [\mathbf{x}_{\mathrm{EL}}, \mathbf{x}_{\mathrm{DSL}_1}, \cdots, \mathbf{x}_{\mathrm{DSL}_K}]$ that is composed of 1 environment lighting (EL) image $\mathbf{x}_{\mathrm{EL}}$ and K directional side lighting (DSL) images $\mathbf{x}_{\mathrm{DSL}_k}$ ($1 \leq k \leq K$), we first relocate current camera to nearly the same pose and position of the last time observation. We then collect current observation matrix $\mathbf{Y} = [\mathbf{y}_{\mathrm{EL}}, \mathbf{y}_{\mathrm{DSL}_1}, \cdots, \mathbf{y}_{\mathrm{DSL}_K}]$ by roughly aligning current multiple DSLs to last time ones, e.g., to setup DSLs in a clockwise order around the object of interest to cast shadow in roughly similar directions. Note, in observation matrices $\mathbf{X}$ and $\mathbf{Y}$, the captured images are arranged as column vectors. Instead of directly using expensive 3D scanner data, in an observation, we use multiple images under varied illuminations to capture the rich 3D structural information by a practical and low-cost way.

As illustrated by Fig. 3, our camera relocation strategy is pretty simple. We first initialize current camera to a reasonable position to cover large enough area that embraces the real target region. Let $\mathbf{I}_c$ indicate present image taken at current camera pose and position. During the relocation process, we actively maintain a blue indicative rectangle R_b lying the center of $\mathbf{I}_c$ and a red navigation rectangle R_r that encodes the relative geometry difference between current camera pose and position to the target state. Specifically, $R_r = \mathbf{H}R_b$, where $\mathbf{H}$ is the homography matrix calculated from $\mathbf{I}_c$ and $\mathbf{x}_{\mathrm{EL}}$. Guided by the indicative and navigation rectangles R_b and R_r, we just need to dynamically adjust camera's pose and position to make R_r coincide with R_b.

3.2. L and F initialization

Given current observation $\mathbf{Y}$, we first conduct *global linear color constancy* to the last time observation $\mathbf{X}$. Specifically, for images $\mathbf{x}^i$ and $\mathbf{y}^i$ ($i = 0$ representing EL images, $1 \le i \le K$ indicating DSL images), we derive a global linear photometric correction matrix $\hat{\mathbf{A}}^i$ (3×3) and a bias vector $\hat{\mathbf{b}}^i$ (3×1) by

$$[\hat{\mathbf{A}}^i, \hat{\mathbf{b}}^i] = \arg\min_{\mathbf{A}^i, \mathbf{b}^i} \|\mathbf{A}^i \tilde{\mathbf{x}}^i + \mathbf{b}^i - \tilde{\mathbf{y}}^i\|_F^2, \tag{1}$$

where $\tilde{\mathbf{x}}^i$ and $\tilde{\mathbf{y}}^i$ denote the RGB colors of matched SIFT feature points in $\mathbf{x}^i$ and $\mathbf{y}^i$, respectively. Note, Eq. (1) is indeed a quasi canonical illumination model [8] and can be efficiently solved in closed form.

After global photometric correction, we can approximately register last time observation $\mathbf{X}$ to current observation $\mathbf{Y}$ using multi-lights flow estimation presented in next subsection, thus yielding $\mathbf{X}_{\mathbf{F}}$.

3.3. Normal-aware lighting correction

We then conduct more accurate *normal-aware lighting correction*. Our method is based on the Lambertian reflectance model, i.e., $\mathbf{I}_p = \int \langle \mathbf{n}_p, \omega \rangle \rho_p L(\omega) d\omega$, where $\mathbf{I}_p$, $\mathbf{n}_p$ and ρ_p represent the color, normal and albedo of pixel p, respectively, and $L(\omega)$ is the spherical lighting function. Hence, we can add a "virtual" light $L^{\mathrm{v}}(\cdot)$ to $\mathbf{X}_{\mathbf{F}}$ to correct its lighting differences to current observation $\mathbf{Y}$,

$$\begin{aligned} \mathbf{x}_{\mathbf{LF}p} &= \int \langle \mathbf{n}_p, \omega \rangle \rho_p (L^{\mathrm{x}}(\omega) + L^{\mathrm{v}}(\omega)) d\omega, \\ &= \mathbf{x}_{\mathbf{F}p} + \mathbf{L}_p^{\mathrm{v}} = \mathbf{y}_p, \end{aligned} \tag{2}$$

where $\mathbf{x}_{\mathbf{LF}p}$ is the corrected color of pixel p and $\mathbf{L}_p^{\mathrm{v}}$ is the color increment caused by the virtual light. Therefore, to equalize the illumination differences between $\mathbf{X}$ and $\mathbf{Y}$, we need to minimize the following objective function:

$$\begin{aligned} \mathbf{L}^i = \arg\min_{\mathbf{L}^{\mathrm{v}}} & \textstyle\sum_p (\mathbf{x}^i_{\mathbf{F}p} + \mathbf{L}^{\mathrm{v}}_p - \mathbf{y}^i_p)^2 \exp(-\frac{\mathbf{C}_p}{\sigma}), \\ & + \alpha \textstyle\sum_{p \sim q} w_{pq} (\mathbf{L}^{\mathrm{v}}_p - \mathbf{L}^{\mathrm{v}}_q)^2, \end{aligned} \tag{3}$$

where $\mathbf{L}^i$ is the spatially variant illumination difference for the i-th lighting of $\mathbf{X}$ and $\mathbf{Y}$. Note, as shown in Eq. (2), $\mathbf{L}^i_p$ is related to the normal of pixel p. In Eq. (3), the objective function is composed of two parts. The first part encourages the photometric consistency between $\mathbf{x}^i_{\mathbf{F}}$ and $\mathbf{y}^i$. $\exp(-\frac{\mathbf{C}_p}{\sigma})$ is the change switch term, where $0 \le \mathbf{C}_p \le 1$ is the change likelihood of pixel p. If $\mathbf{C}_p$ is close to 1, the first term is disabled. The second term encourages spatial smoothness of the virtual light, where w_{pq} represents the normal similarity of pixel p and q, $p \sim q$ indicates pixels p and q are neighbors. In our experiments, we measure w_{pq} as shading closeness of p and q, which can be conveniently obtained by dividing per-pixel color with the approximate reflectance, i.e., chromaticity [31]. α and σ are parameters that control the influence of second term and pixel-level change likelihood $\mathbf{C}_p$, respectively. Note, Eq. (3) defines a sparse unconstrained quadratic minimization problem that can also be efficiently solved in closed form.

3.4. Camera geometry correction

With normal-aware lighting difference $\mathbf{L}^i$, we can obtain a new lighting corrected last time observation, denoted as $\mathbf{X}_{\mathbf{L}}$. From this, we can further refine the camera geometry correction flow $\mathbf{F}$ by taking all illuminations of $\mathbf{X}_{\mathbf{L}}$ and $\mathbf{Y}$ into account. Specifically, we extend the SIFT flow framework [14] by revising its energy into the following form:

$$\begin{aligned} E(\mathbf{F}) \quad &= \sum_{i,p} \|\mathbf{x}^i_{\mathbf{L}}(p + \mathbf{F}_p) - \mathbf{y}^i(p)\|_1 \exp(-\frac{\mathbf{C}_p}{\sigma}) \\ &+ \beta \textstyle\sum_p \|\mathbf{F}_p\|_2^2 \\ &+ \textstyle\sum_{p \sim q} \min(\gamma \|\mathbf{F}_p - \mathbf{F}_q\|_1, d), \end{aligned} \tag{4}$$

where $\exp(-\frac{\mathbf{C}_p}{\sigma})$ is the change switch term with the same role in Eq. (3). Note, the new energy Eq. (4) can also be effectively minimized using the two-layer BP algorithm [14].

Since we focus on rigid scenes, we find that, in most cases, $\mathbf{F}$ can be faithfully initialized by the homography matrix $\mathbf{H}$ obtained by the stage of active camera relocation.

3.5. Low-rank change detection

Let $\mathbf{X}_{\mathbf{LF}}$ denote the lighting and camera geometry corrected last time observation images. For each particular illumination condition i, we have $\mathbf{O}^i = [\mathbf{X}^i_{\mathbf{LF}}, \mathbf{Y}^i]$ by stacking $\mathbf{X}^i_{\mathbf{LF}}$ and $\mathbf{Y}^i$ into a two-columns matrix. We further stack $\mathbf{O}^i$ together to get $\mathbf{O}$. Since all corresponding lighting conditions and camera geometries are aligned by $\mathbf{L}$ and $\mathbf{F}$, we have the following program to detect fine-grained changes

$$\begin{aligned} \arg\min_{\mathbf{Z}, \mathbf{E}} & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_1 + \kappa \|\mathbf{T}\mathbf{E}\|_F^2, \\ \text{s.t.} \quad & \mathbf{O} = \mathbf{Z} + \mathbf{E}, \end{aligned} \tag{5}$$

where $\mathbf{O} \in \mathcal{R}^{3N \times 2(K+1)}$ is the spatially and photometrically aligned two times observation images, N is the number of pixel of an image, $K + 1$ is the number of illuminations, $\mathbf{Z}$ represents the unchanged parts and $\mathbf{E}$ indicates the sparse changes between $\mathbf{X}_{\mathbf{LF}}$ and $\mathbf{Y}$. Matrix

$\mathbf{T} = \mathrm{diag}(\mathbf{A}, \mathbf{A}, \mathbf{A})$ encodes the pixel-level neighboring relations, i.e., for two neighboring pixels p and q, $\mathbf{A}_{pp} = \mathbf{A}_{qq} = 1$ and $\mathbf{A}_{pq} = \mathbf{A}_{qp} = -1$. The third term of Eq. (5) encourages spatially smoothness of the change components. By introducing an auxiliary variable $\mathbf{J}$, the problem can be transformed to

$$\begin{aligned} \arg\min_{\mathbf{Z},\mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda\|\mathbf{J}\|_1 + \kappa\|\mathbf{T}\mathbf{E}\|_F^2, \\ & + \Phi(\mathbf{Y}_1, \mathbf{O} - \mathbf{Z} - \mathbf{E}) \\ & + \Phi(\mathbf{Y}_2, \mathbf{J} - \mathbf{E}), \end{aligned} \tag{6}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are augmented Lagrange multipliers, $\Phi(\mathbf{Y}, \mathbf{Z}) = \frac{\mu}{2}\|\mathbf{Y}\|_F^2 + \langle \mathbf{Y}, \mathbf{Z} \rangle$ is augmented Lagrange constraint function [15]. We solve Eq. (6) by iteratively optimizing $\mathbf{Z}$, $\mathbf{E}$ and $\mathbf{J}$ using the Augmented Lagrange Multiplier (ALM) algorithm [13, 15]. From $\mathbf{E}$, we derive the change likelihood map $\mathbf{C}$ by averaging all change components derived from multiple lighting conditions.

3.6. Coarse-to-fine optimization and final decision

As shown in Fig. 2, we conduct the optimization of $\mathbf{L}$, $\mathbf{F}$ and $\mathbf{C}$ in $3 \sim 5$ rounds to converge. Moreover, for the purpose of acceleration and reliable multiscale change detection, we implement our iterative optimization within a coarse-to-fine scheme. Specifically, the camera geometry correction flow $\mathbf{F}$ and change likelihood map $\mathbf{C}$ at level $l-1$ are propagated to level l through upsampling to generate the lighting correction difference $\mathbf{L}$ at finer scale. Empirically, changes detected in coarse level have lower false alarm, while changes derived from finer levels tend to have lower false negative. Hence, we obtain the overall change likelihood $\mathbf{C}^{\mathrm{all}}$ by averaging the change likelihood maps of all scales. As shown in our experiments, the final change decision can be obtained by simply thresholding $\mathbf{C}^{\mathrm{all}}$. To get reliable change decision, for each pixel, we further use a local 7×7 window to collect its change likelihood feature and make the final change decision via a linear SVM, which actually acts as rotation-involved thresholding.

Table 1. Average performance on the Summer Palace dataset $\mathbf{D_p}$.

Method	F1	Re	Pr	Sp	FRR	FNR	PWC
SC_SOBS A	0.03	0.90	0.01	0.09	0.91	0.10	89.80
SC_SOBS M	0.02	**0.97**	0.01	0.03	0.97	**0.03**	96.21
SC_SOBS LFA	0.11	0.34	0.06	0.95	0.05	0.66	5.62
SC_SOBS LFM	0.16	0.20	0.14	0.99	0.01	0.80	1.92
SubSENSE A	0.02	0.55	0.02	0.60	0.40	0.45	39.41
SubSENSE M	0.02	0.93	0.01	0.20	0.80	0.07	78.85
SubSENSE LFA	0.08	0.04	0.05	0.99	0.01	0.96	1.81
SubSENSE LFM	0.07	0.22	0.04	0.95	0.05	0.78	5.85
Ours (D&T)	0.34	0.28	**0.52**	**1.00**	**0.00**	0.72	**0.92**
Ours (SVM)	**0.51**	0.53	0.47	0.99	0.01	0.47	1.02

4. Experimental Results

4.1. Datasets

We have built 3 real-world datasets to benchmark fine-grained change detection of misaligned scenes under varied illuminations.[4] The first dataset ($\mathbf{D_p}$) contains two outdoor scenes in the Summer Palace. Each scene was observed twice with one year time interval. Since the scenes are outdoor and observations are taken with very long time interval, we can see the lighting conditions are highly different. The camera geometry (position and pose) was relocated by active camera relocation introduced in section 3.1. For each time data collection, we captured 7 images under 7 different illuminations, including one environment lighting (EL) and six directional side lightings (DSLs). The six DSLs were roughly aligned to their corresponding directions of the last-time observation, but the tone and intensity were quite different. Two groups of images for different scenes are named $\mathbf{D_p}$-1 and $\mathbf{D_p}$-2, respectively. The second dataset ($\mathbf{D_b}$) include 10 groups of images of laboratory testing blocks for the ageing simulation tests of mural deteriorations. Similar to dataset $\mathbf{D_p}$, all testing blocks were observed at 7 different illuminations (1 EL, 6 DSLs). Since $\mathbf{D_b}$ was taken in lab, the illuminations (both direction and intensity), camera pose and position were well controlled. As a result, the illumination and camera geometry variances are less apparent than dataset $\mathbf{D_p}$. But, the changes occurred very fast at fine-grained scale. These images are grouped as $\mathbf{D_b}$-1 to $\mathbf{D_b}$-10 according to different mural briquettes. The third dataset ($\mathbf{D_s}$) are collected from 4 different statues. Similar to $\mathbf{D_p}$ and $\mathbf{D_b}$, each scene in dataset $\mathbf{D_s}$ was observed at 7 different illuminations (1 EL, 6 DSLs). For each statue, artificial

Table 2. Average results on mural briquettes dataset $\mathbf{D_b}$.

Method	F1	Re	Pr	Sp	FRR	FNR	PWC
SC_SOBS A	0.03	0.31	0.02	0.69	0.31	0.69	31.52
SC_SOBS M	0.03	0.36	0.02	0.65	0.35	0.64	35.07
SC_SOBS LFA	0.02	0.02	0.03	0.99	0.01	0.98	1.99
SC_SOBS LFM	0.09	0.08	0.14	0.99	0.01	0.92	2.01
SubSENSE A	0.24	0.50	0.31	0.72	0.28	0.50	28.32
SubSENSE M	0.23	**0.67**	0.19	0.66	0.34	**0.33**	34.01
SubSENSE LFA	0.06	0.03	0.26	**1.00**	**0.00**	0.97	1.43
SubSENSE LFM	0.28	0.21	0.50	0.99	0.01	0.79	1.62
Ours (D&T)	0.45	0.40	**0.56**	**1.00**	**0.00**	0.60	**1.23**
Ours (SVM)	**0.53**	0.62	0.48	0.99	0.01	0.38	1.41

Table 3. Average results on statue dataset $\mathbf{D_s}$.

Method	F1	Re	Pr	Sp	FRR	FNR	PWC
SC_SOBS A	0.01	0.59	0.00	0.78	0.22	0.41	22.29
SC_SOBS M	0.01	0.66	0.00	0.73	0.27	0.34	27.41
SC_SOBS LFA	0.19	0.34	0.14	**1.00**	**0.00**	0.66	0.31
SC_SOBS LFM	0.27	0.44	0.19	**1.00**	**0.00**	0.56	**0.24**
SubSENSE A	0.02	0.83	0.01	0.88	0.12	0.17	12.13
SubSENSE M	0.01	**0.98**	0.00	0.66	0.34	**0.02**	34.28
SubSENSE LFA	0.27	0.28	0.34	0.99	0.01	0.72	1.57
SubSENSE LFM	0.12	0.77	0.07	0.95	0.05	0.23	5.37
Ours (D&T)	**0.53**	0.78	**0.43**	**1.00**	**0.00**	0.22	0.28
Ours (SVM)	0.51	0.86	0.39	**1.00**	**0.00**	0.14	0.29

Table 4. Average F1-measure with different coarse-to-fine scales.

Dataset	1 Level	2 Levels	3 Levels
$\mathbf{D_p}$	0.3595	0.4476	**0.5134**
$\mathbf{D_b}$	0.3907	0.4897	**0.5254**
$\mathbf{D_s}$	0.4236	**0.6737**	0.5118

[4] Both the datasets and the code of our approach are available online: http://cs.tju.edu.cn/szdw/jsfjs/fengwei/papers/fcd_ICCV2015/fcd_iccv15.htm.

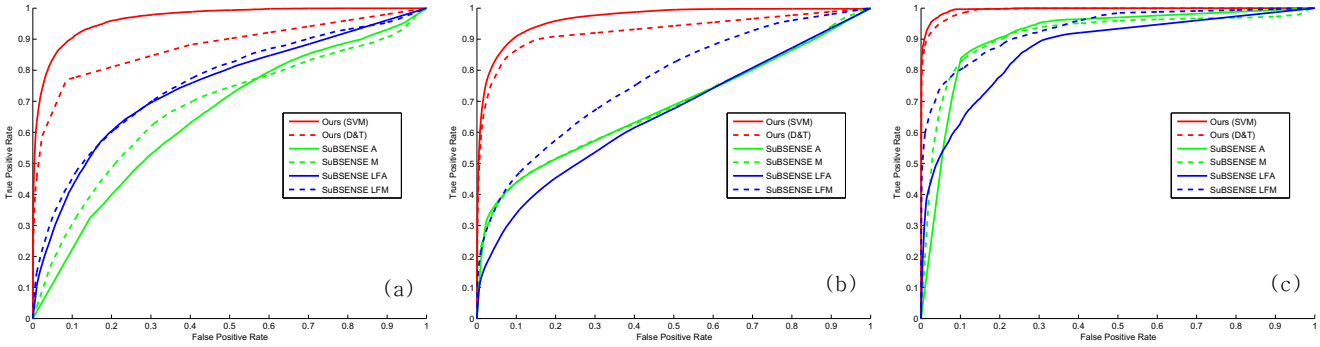

Figure 4. ROC curves of datasets $\mathbf{D_p}$ (a), $\mathbf{D_b}$ (b), and $\mathbf{D_s}$ (c).

little damage of the statues were purposely created to cause the fine-grained changes between two times observations. They are named $\mathbf{D_s}$-1 to $\mathbf{D_s}$-4, respectively.

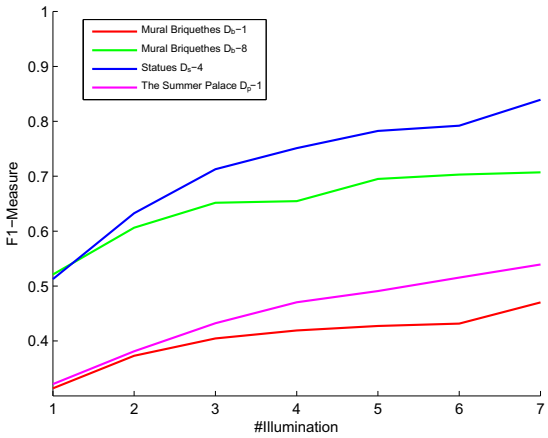

Figure 5. Fine-grained change detection on multiple illuminations.

4.2. Setup

We choose two state-of-the-art change detection methods as our baselines, i.e., SC_SOBS [17] and SuBSENSE [22], both of which report very good results in the benchmark CDNet Challenge 2012 and 2014 [10]. Since we have multiple illuminations images in one-time observation, to fairly evaluate the baseline methods, we tried two fusion strategies: 1) feeding baseline methods with a two-frame video composed of the average illumination images, and getting one change decision as result (indicated by A at the end of method name, e.g., SC_SOBS A); 2) feeding baseline methods with 7 two-frame videos, corresponding to 7 different illuminations, and use the average change detection map as final change decision (denoted by M at the end of method name, e.g., SuBSENSE M). Considering the existence of lighting and camera geometry variations in our datasets, we further conduct lighting correction and optical flow compensation using the proposed method on the image groups before sending them to the baseline methods (marked with LF, e.g., SC_SOBS LF). Therefore, we totally have 8 baseline methods in our evaluation.

To verify the effect of linear SVM in our final change decision, we compared our method using linear SVM (marked by SVM) and simple "difference + thresholding" as the final change decision strategy (marked by D&T). Specifically, the simple "difference + thresholding" uses change likelihood map $\mathbf{C}^{\text{all}}$ as input and thresholds it to get the final change decision. In our experiments, we chose $\mathbf{D_p}$-1, $\mathbf{D_b}$-1, and $\mathbf{D_s}$-4 to form a training set and trained a single linear SVM model for all datasets. When training the simple "difference + thresholding" strategy, we used the same training set, but trained different thresholds at each datasets respectively, according to the best F1-measure.

Our method has three types of parameters. The first part is the lighting correction parameters, α and σ, as shown in Eq. (3). Since in different datasets the lighting differences are different, the lighting correction parameters α and σ are dataset-related. For dataset $\mathbf{D_p}$, $\sigma = 0.033$ and $\alpha = 1$; for $\mathbf{D_b}$, $\sigma = 0.0125$ and α is between 3 and 10; for $\mathbf{D_s}$, $\sigma = 0.0125$ and $\alpha = 10$. The second part is camera geometry correction parameters. We just used the default parameters of single image SIFT Flow [14]. The third part is change detection parameters λ and κ in Eq. (5). λ is fixed as 0.006 at the coarsest level and scaled by 0.5 at each finer level. We used 3 levels pyramid in our experiments. κ is fixed as 0.01 for all datasets. For methods SC_SOBS and SuBSENSE, we have tried a series of reasonable parameters, and used the best one (measured by F1-measure) in our evaluation.

Following [10], we compared 7 metrics for quantitative evaluation, including F1-measure (F1), recall (Re), precision (Pr), specificity (Sp), false positive rate (FRR), false negative rate (FNR) and percentage of wrong classifications (PWC). In the next, we comprehensively compare the proposed approach to state-of-the-art baselines in terms of the 7 metrics on three benchmark datasets.

4.3. Quantitative comparison

The average quantitative performance of different methods on three benchmark datasets are shown in Table 1, 2

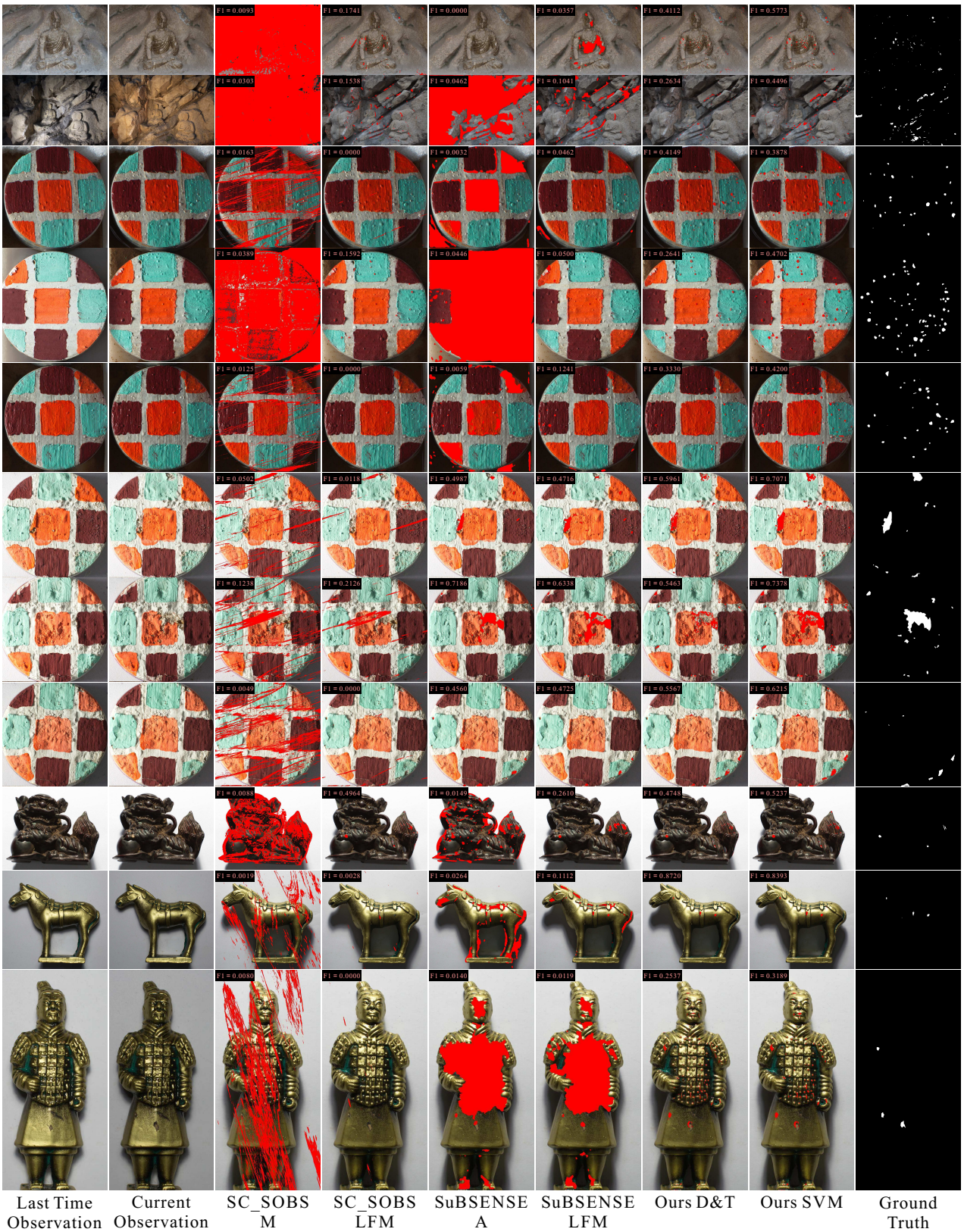

Figure 6. Visual comparison of fine-grained change detection.

and 3. We can clearly see that our method achieves highest F1-measure and low PWC error simultaneously. We also note that using linear SVM as the final change decision strategy produces apparently higher F1-measure than using D&T in datasets $\mathbf{D_p}$, $\mathbf{D_b}$, and is comparable in dataset $\mathbf{D_s}$. Note, in our evaluation, we use pixel-level F1-measure that is a very strict criterion. As show in the last three columns of Fig. 6, change detection with pixel-level F1-measure higher than 0.4 is visually pretty close to the ground truth. By taking illumination and camera geometry compensation into account, the extended baseline methods (denoted as LF) can generally produce higher precision in most cases, and reduce FRR, PWC values significantly.

We compare the average ROC curves in Fig. 4. Since we need per-pixel soft change likelihood (e.g., the change decision likelihood of linear SVM) and SC_SOBS only produces hard decision value, we cannot generate ROC curves for SC_SOBS and its extensions. Hence, in Fig. 4, we only show the ROC curves of our method (SVM and D&T), SuBSENSE and its extended versions. For SuBSENSE, we use $\#\{\mathrm{dist}(\mathbf{I_t}(x), \mathbf{B_n}(x)) < R, \forall n\}$ as the change decision score [22]. From Fig. 4 (a), (b) and (c), we can also see fact that our method clearly outperforms state-of-the-art SuBSENSE algorithm and its extensions.

The average speed of handling 720×515 images by different methods is as following: SC_SOBS A 2.14s, SC_SOBS M 14.5s, SC_SOBS LFA 595.4s, SC_SOBS LFM 607.3s, SubSENSE A 6.52s, SubSENSE M 46.36s, SubSENSE LFA 601.3s, SubSENSE LFM 639.5s, Ours (D&T) 593.6s, Ours (SVM) 594.3s. As we argued in section 1, for fine-grained change detection, accuracy and robustness to both lighting and camera geometry variations are far more important than real-time response. Most of our time cost is in the coarse-to-fine optimization. For LF extended methods, time used for lighting correction and camera geometry correction are included.

4.4. More results

To verify the effectiveness of coarse-to-fine optimization, Table 4 compares the average F1-measure of increasing number of levels used in our method on three datasets. We can clearly see that using 2 and 3 levels gives much better results than single level, and most results of using 3 levels are better than using 2 levels. The benefits of multiscale change detection are two-folds: 1) coarser level change detection tends to produce lower false alarm, while finer level tends to generate lower false negative, so multiscale detection may have higher overall performance; 2) using multiscale change detection is faster than finest scale detection with same number of optimization rounds.

To testify the role of multiple illuminations in fine-grained change detection via capturing rich fine-scale structural information of a scene, we compare the change detection performance of our approach by using increasing number of illuminations in Fig. 5. In this experiment, the illumination number ranges from 1 to 7. For each illumination number, we randomly select 7 testing cases to do fine-grained changes and calculate the average of F1-measure. We chose four images from the three datasets, $\mathbf{D_p}$-1, $\mathbf{D_b}$-1, $\mathbf{D_b}$-8 and $\mathbf{D_s}$-4, as the testing data to carry on this experiment. Fig. 5 compares the fine-grained change detection results with increasing number of illuminations for the four testing images, respectively. From Fig. 5, we can clearly see the benefit of multiple lightings to the performance. This is because multiple illuminations provide more information about the 3D structure of a scene. Hence, using multiple lightings is certainly helpful to do better fine-grained change detection. Generally, the more the number of illuminations, the better the performance.

Fig. 6 shows more change detection results for two scenes in $\mathbf{D_p}$, six mural briquettes in $\mathbf{D_b}$ and three statues in $\mathbf{D_s}$. Since there are multiple illumination images, we select images with similar DSL illumination for illustration. It is clear that SC_SOBS and SuSENSE produce rather poor fine change detection results. On the other hand, our method is able to generate satisfactory change detection results that are visually quite similar to ground truth.

5. Conclusion

In this paper, we have proposed a low-cost active approach to fine-grained change detection for high-value object state monitoring. The major contributions of this work are three-folds. First, unlike previous change detection methods that focus on detecting large-scale significant changes, our approach provides a feasible end-to-end solution to this challenging problem, which covers both active observation data collection and reliable fine-grained change detection. We particularly show how to use coarse-to-fine optimization and low-rank analysis to achieve high-quality change decisions. Second, we build three real-world datasets to benchmark this broadly interesting problem. Third, our approach can also be used to accurately reconstruct pixel-level correspondence under illumination and camera geometry variations, thus is widely applicable to many vision tasks, such as temporal 3D model reconstruction with 3D point correspondence. In near future, we plan to incorporate more physical constraints, such as scene normal and albedo priors, into our model, and to study parallel optimization algorithms to further accelerate the detection speed. We are also interested in detecting and measuring fine-grained change trends from multiple observations.

Acknowledgements We appreciate anonymous reviewers and the area chair for their valuable comments. This work is supported by the National Science and Technology Support Project (2013BAK01B01) and NSFC (61572354, 61272266).

References

[1] J. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE TPAMI*, 2014. 3

[2] H. Dai, W. Feng, L. Wan, and X. Nie. L0 co-intrinsic images decomposition. In *ICME*, 2014. 3

[3] I. Eden and D. B. Cooper. Using 3D line segments for robust and efficient change detection from multiple noisy images. In *ECCV*, 2008. 2

[4] C.-Y. Fang, S.-W. Chen, and C.-S. Fuh. Automatic change detection of driving environments in a vision-based driver assistance system. *IEEE TNN*, 14(3):646–657, 2003. 2

[5] W. Feng, J. Jia, and Z.-Q. Liu. Self-validated labeling of Markov random fields for image segmentation. *IEEE TPAMI*, 32(10):1871–1887, 2010. 1

[6] W. Feng and Z.-Q. Liu. Region-level image authentication using Bayesian structural content abstraction. *IEEE TIP*, 17(12):2413–2424, 2008. 1

[7] W. Feng, Z.-Q. Liu, L. Wan, C.-M. Pun, and J. Jiang. A spectral-multiplicity-tolerant approach to robust graph matching. *Pattern Recognition*, 46(10):2819–2829, 2013. 1

[8] A. Gijsenij, T. Gevers, and J. Van De Weijer. Computational color constancy: Survey and experiments. *IEEE TIP*, 20(9):2475–2489, 2011. 2, 4

[9] L. Giustarini, R. Hostache, P. Matgen, and G. Schumann. A change detection approach to flood mapping in urban areas using terrasar-x. *IEEE TGRS*, 51(4):2417–2430, 2013. 2

[10] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. changedetection.net: A new change detection benchmark dataset. In *CVPRW*, 2012. 1, 2, 6

[11] D. Hauagge, S. Wehrwein, K. Bala, and N. Snavely. Photometric ambient occlusion. In *CVPR*, 2013. 3

[12] L. Li, W. Feng, L. Wan, and J. Zhang. Maximum cohesive grid of superpixels for fast object localization. In *CVPR*, 2013. 1

[13] Z. Lin, A. Ganesh, J. Wright, M. Chen, L. Wu, and Y. Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *SIAM J. Optimization*, 2011. 2, 5

[14] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE TPAMI*, 33(5):978–994, 2011. 3, 4, 6

[15] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE TPAMI*, 35(1):171–184, 2013. 2, 5

[16] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE TIP*, 17(7):1168–1177, 2008. 1, 2

[17] L. Maddalena and A. Petrosino. The SOBS algorithm: what are the limits? In *CVPRW*, 2012. 1, 2, 6

[18] X. Nie, W. Feng, L. Wan, H. Dai, and C.-M. Pun. Intrinsic image decomposition by hierarchical L0 sparsity. In *ICME*, 2014. 3

[19] T. Pollard and J. L. Mundy. Change detection in a 3-d world. In *CVPR*, 2007. 2

[20] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE TIP*, 14(3):294–307, 2005. 1, 2, 3

[21] K. Sakurada, T. Okatani, and K. Deguchi. Detecting changes in 3D structure of a scene from multi-view images captured by a vehicle-mounted camera. In *CVPR*, 2013. 1, 2

[22] P. St-Charles, G. Bilodeau, and R. Bergevin. Subsense: A universal change detection method with local adaptive sensitivity. *IEEE TIP*, 24(1), 2015. 1, 2, 6, 8

[23] S. Staniforth, editor. *Historical Perspectives on Preventive Conservation*. Getty Conservation Institute, 2013. 1

[24] D. Sun, S. Roth, and M. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *IJCV*, 106(2):115–137, 2014. 3

[25] A. Taneja, L. Ballan, and M. Pollefeys. Image based detection of geometric changes in urban environments. In *ICCV*, 2011. 1, 2

[26] A. Taneja, L. Ballan, and M. Pollefeys. City-scale change detection in cadastral 3D models using images. In *CVPR*, 2013. 1, 2

[27] H. Wirilander. Preventive conservation: a key method to ensure cultural heritage's authenticity and integrity in preservation process. *E-Conservation Magazine*, 6(24), 2012. 1

[28] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. In *CVPR*, 2009. 2

[29] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE TPAMI*, 34(9):1744–1757, 2012. 3

[30] O. Yousif and Y. Ban. Improving urban change detection from multitemporal sar images using pca-nlm. *IEEE TGRS*, 51(4):2032–2041, 2013. 2

[31] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin. A closed-form solution to retinex with non-local texture constraints. *IEEE TPAMI*, 34(7):1437–1444, 2012. 3, 4