6D Dynamic Camera Relocalization from Single Reference Image

Wei Feng, Fei-Peng Tian*, Qian Zhang, Jizhou Sun
School of Computer Science and Technology, Tianjin University, Tianjin, China
Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China
{wfeng, tianfeipeng, qianz, jzsun}@tju.edu.cn

Abstract

Dynamic relocalization of 6D camera pose from single reference image is a costly and challenging task that requires delicate hand-eye calibration and precision positioning platform to do 3D mechanical rotation and translation. In this paper, we show that high-quality camera relocalization can be achieved in a much less expensive way. Based on inexpensive platform with unreliable absolute repositioning accuracy (ARA), we propose a hand-eye calibration free strategy to actively relocate camera into the same 6D pose that produces the input reference image, by sequentially correcting 3D relative rotation and translation. We theoretically prove that, by this strategy, both rotational and translational relative pose can be effectively reduced to zero, with bounded unknown hand-eye pose displacement. To conquer 3D rotation and translation ambiguity, this theoretical strategy is further revised to a practical relocalization algorithm with faster convergence rate and more reliability by jointly adjusting 3D relative rotation and translation. Extensive experiments validate the effectiveness and superior accuracy of the proposed approach on laboratory tests and challenging real-world applications.

1. Introduction

Camera pose registration and relocalization is an essential problem in computer vision and robotics, fundamentally supporting a number of important real-world applications, such as structure-from-motion (SfM) [22, 37, 16], 3D tracking and mapping [24, 17], monocular SLAM [2, 34], and scene change detection [7, 5].

Despite the diversity of previous successful methods, in computer vision, most recent efforts on this topic have been focused on *static* camera registration and relocalization (SCR) [22, 30, 10]. That is, for one or a group of input images, SCR studies how to align their camera poses into a unified world coordinate system that may come from a known 3D scene [10, 30, 15] or from 3D reconstruction using the input images [22, 16]. In SCR, the camera poses of input images are fixed and cannot be actively readjusted.

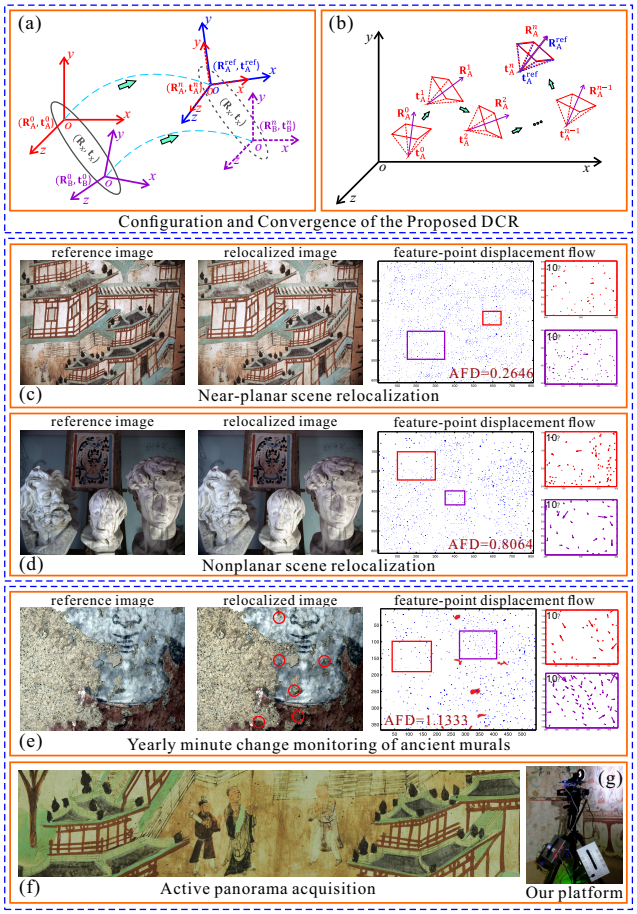

Figure 1. **Inexpensive single image DCR.** (a) Coordinate systems of eye $\langle \mathbf{R}_\mathrm{A}, \mathbf{t}_\mathrm{A} \rangle$, hand $\langle \mathbf{R}_\mathrm{B}, \mathbf{t}_\mathrm{B} \rangle$, target camera pose $\langle \mathbf{R}_\mathrm{A}^\mathrm{ref}, \mathbf{t}_\mathrm{A}^\mathrm{ref} \rangle$, and the unknown hand-eye relative pose $\langle \mathbf{R}_\mathrm{X}, \mathbf{t}_\mathrm{X} \rangle$. (b) DCR dynamic convergence process. (c)–(d) Our DCR results for near-planar and nonplanar scenes. (e) Real minute changes (0.1mm level), occurred during June 2014 and July 2015, discovered by our DCR in Cave-465 of Dunhuang Mogao Grottoes. (f) HD panoramic image directly captured by a moving camera whose trajectory is controlled by our DCR. (g) Our inexpensive DCR platform. See text for more details.

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

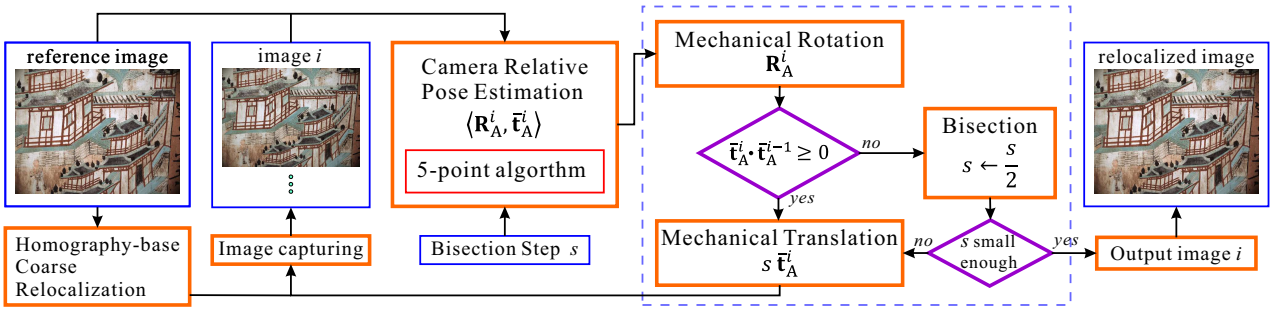

Figure 2. Working flow of the proposed 6D dynamic camera relocalization. See text for details.

In this paper, we treat camera pose relocalization as a dynamic process and study *dynamic* (or *active*) camera relocalization (DCR) from single reference image. Our work was originally motivated by a real-world problem, minute change monitoring and measurement of ancient murals for preventive conservation. This is indeed a very challenging problem due to three major reasons.

1. **High accuracy requirement.** We want to find and measure very fine changes occurred on murals with complex image content and deterioration patterns.
2. **Wild environment applicability.** Both the equipment and algorithm should work well in unrestricted environments, where both the unpleasant weather and the need of frequent disassembling/reassembling equipments to realize portability for different caves/spots can jeopardize hardware accuracy.
3. **Long time interval.** Relics usually change very slowly, thus we must precisely relocalize 6D camera pose to capture the possible minute changes.

Note, there is NO mature solution satisfying all the above three requirements. For instance, single image DCR could be directly solved as a robotic visual servoing problem [3, 21, 39, 26]. However, high precision robotics are not applicable to wild environment and cannot be frequently disassembled and reassembled, and its accuracy highly relies on hand-eye calibration [29, 14, 28]. Rephotography [1] cannot do this either, due to its much lower relocalization accuracy and the limited ability to physically handle only 3D relative translation. In contrast, as shown in Fig. 1 (c)–(e), the proposed approach guarantees to produce high-quality 6D DCR for both near-planar and nonplanar scenes. Fig. 1 also shows that our DCR has successfully discovered 0.1mm real mural changes occurred in Dunhuang Mogao Grottoes, which truly provides ice-breaking results for this important real-world problem and has great potentials in other areas, such as online status monitoring of high speed train.

Our major contribution is three-fold: 1) hand-eye calibration free, which enables frequent equipments disassembling/reassembling without losing precision; 2) applicable to common platforms and wild environments; 3) theoretically guaranteed convergence and feasible boundary condition. To our best knowledge, it is the first solid hand-eye calibration free 6D DCR method in CV and robotics.

2. Related Work

Static camera relocalization (SCR). Many fundamental computer vision applications relate to camera pose registration and relocalization, such as sparse or dense 3D reconstruction [15, 17, 2, 22, 24, 27], monocular SLAM and camera tracking from RGB [39, 21, 34] or RGBD images [23, 8, 30, 10]. Basically, they share a common SCR problem. That is, they all want to align camera poses of input one or multiple images within a unified world coordinate system, which can be defined by a (partially) known 3D scene (e.g., monocular SLAM [2, 34] and active scene scanning [23, 35]) or reconstructed via SfM pipeline [15, 16].

Fast and reliable camera pose estimation is critical to SCR. State-of-the-art method is 5-point algorithm [25], due to its generality and superior accuracy [31]. Given the matched feature points set [20] extracted from input images, 5-point algorithm can faithfully generate their relative rotation $\mathbf{R}$ and relative translation direction $\bar{\mathbf{t}}$.[1] Besides, ESM is also used in camera pose estimation [24, 27].

Visual servoing. Robotic visual servoing (VS), aiming to control the pose of robot end-effector (*hand*) via visual feedback (*eye*) [26, 39, 21], is closely related to our work. In restricted environment, e.g., objects with markers or planar moving assumption [21, 39], well-calibrated VS could be used to dynamically relocalize camera pose. However, as aforementioned, VS is not a cheap and reliable DCR worker, whose accuracy highly relies on hand-eye calibration or other restricted conditions.

Rephotography. Computational rephotography aims to recapture a photograph from the same viewpoint of a historical photograph. Previous work mainly focuses on the study of history [32], monitoring of natural environment, such as glacier melting [9] and geological erosion [11], ecological research [33]. These work highly relies on manual judgment, that makes it quite hard for operation and its accuracy is not high. Recently, several computational rephotogra-

[1] Unlike relative rotation $\mathbf{R}$, only the direction $\bar{\mathbf{t}} = \frac{\mathbf{t}}{\|\mathbf{t}\|}$ of relative translation $\mathbf{t}$ can be determined purely from images.

phy (CRP) methods [1, 18, 33] have been proposed and improved the efficiency and accuracy. West *et al*. [33] present a linear blending based rephotography method for the purpose of monitoring urban tree canopy. Bae *et al*. [1] propose an interactive and computational tool to guide users to reach the desired viewpoint, which can achieve about $0.5m$ physical relocalization accuracy. However, these existing methods rely on many user interactions and lack the accuracy to support minute change detection of high-value scenes.

Hand-eye calibration. Hand-eye calibration estimates relative pose between robot hand and eye coordinate systems, by solving $\mathbf{AX} = \mathbf{XB}$, where $\mathbf{X}$ is the homogeneous representation of the relative pose $\langle \mathbf{R}_{\mathrm{X}}, \mathbf{t}_{\mathrm{X}} \rangle$ between the eye coordinate system $\langle \mathbf{R}_{\mathrm{A}}, \mathbf{t}_{\mathrm{A}} \rangle$ and the hand system $\langle \mathbf{R}_{\mathrm{B}}, \mathbf{t}_{\mathrm{B}} \rangle$ [29, 14, 28]. Generally, accurate hand-eye calibration involves nonlinear optimization in $\mathrm{SO}(3)$ and $\mathbb{R}^3$ [12, 13]. State-of-the-art methods apply branch-and-bound strategy to search global optimum using labeled images with object markers in restricted environment [29, 14]. In contrast, this paper gives a hand-eye calibration free approach to DCR and provides theoretical upper bound of hand-eye pose displacement, under which relative rotational and translational pose can be certainly decreased to zero.

3. 6D Dynamic Camera Relocalization

Notation. A 3D rotation matrix $\mathbf{R} \in \mathrm{SO}(3)$ can also be expressed by axis-angle representation $\langle \theta, \bar{\mathbf{e}} \rangle$ and quaternion $(\cos\frac{\theta}{2}, \bar{\mathbf{e}} \sin\frac{\theta}{2})$, where unit vector $\bar{\mathbf{e}}$ is the invariant Euler axis of $\mathbf{R}$ and θ is the rotation magnitude about $\bar{\mathbf{e}}$. In this paper, we use $\mathbf{R} \simeq \langle \theta, \bar{\mathbf{e}} \rangle \simeq (\cos\frac{\theta}{2}, \bar{\mathbf{e}} \sin\frac{\theta}{2})$ to indicate the equivalence of three kinds of representations for a 3D rotation. Let $\langle \mathbf{R}_{\mathrm{X}}, \mathbf{t}_{\mathrm{X}} \rangle$ be the constant relative pose between eye coordinate system $\langle \mathbf{R}_{\mathrm{A}}, \mathbf{t}_{\mathrm{A}} \rangle$ and hand system $\langle \mathbf{R}_{\mathrm{B}}, \mathbf{t}_{\mathrm{B}} \rangle$, where $\mathbf{R}$ denotes the orientation and $\mathbf{t}$ is the position of a coordinate system w.r.t. a fixed world coordinate system. As shown in Fig. 1, the input reference image defines the target camera pose $\langle \mathbf{R}_{\mathrm{A}}^{\mathrm{ref}}, \mathbf{t}_{\mathrm{A}}^{\mathrm{ref}} \rangle$. Our problem is to dynamically relocalize current eye system $\langle \mathbf{R}_{\mathrm{A}}, \mathbf{t}_{\mathrm{A}} \rangle$ to the target pose $\langle \mathbf{R}_{\mathrm{A}}^{\mathrm{ref}}, \mathbf{t}_{\mathrm{A}}^{\mathrm{ref}} \rangle$ by properly moving hand, with unknown $\langle \mathbf{R}_{\mathrm{X}}, \mathbf{t}_{\mathrm{X}} \rangle$. We use i to indicate iteration number, thus $\langle \mathbf{R}_{\mathrm{A}}^i, \mathbf{t}_{\mathrm{A}}^i \rangle$ and $\langle \mathbf{R}_{\mathrm{B}}^i, \mathbf{t}_{\mathrm{B}}^i \rangle$ represent the current camera and hand poses after i-iterations movement, respectively.

3.1. Problem formulation

We want to realize hand-eye calibration free DCR on common low-cost positioning platform. Compared to expensive high-precision platforms whose *absolute repositioning accuracy* (ARA) and *repetitive repositioning accuracy* (RRA) are both reliable, for proper low-cost platforms, their ARA is unreliable but RRA can be trusted.[2] Hence, we should make best use of repetitive moving strategy. Besides, our DCR model needs to overcome two realistic difficulties:

1. **Difficulty 1:** Unknown hand-eye calibration,
2. **Difficulty 2:** We can only obtain the direction of relative camera translation $\bar{\mathbf{t}}$ from images [25, 31].

We start from the *ideal* relations between the hand and eye coordinate systems during the DCR process. Let $\mathbf{p} \in \mathbb{R}^3$ be an arbitrary 3D point in world coordinate system, $\mathbf{c}$ and $\mathbf{h}$ denote its new coordinate in the reference camera system and corresponding target hand system, respectively. Similarly, $\mathbf{c}_i$ and $\mathbf{h}_i$ denote the coordinate of point $\mathbf{p}$ in current camera and hand system after i-iterations adjustment. Since hand-eye relation is fixed and 5-point algorithm can produce reliable camera pose, ideally, we have

$$\begin{aligned} \mathbf{c} &= \mathbf{R}_{\mathrm{A}}^i \mathbf{c}_i + \mathbf{t}_{\mathrm{A}}^i, \\ \mathbf{h} &= \mathbf{R}_{\mathrm{B}}^i \mathbf{h}_i + \mathbf{t}_{\mathrm{B}}^i, \\ \mathbf{c} &= \mathbf{R}_{\mathrm{X}} \mathbf{h} + \mathbf{t}_{\mathrm{X}}, \\ \mathbf{c}_i &= \mathbf{R}_{\mathrm{X}} \mathbf{h}_i + \mathbf{t}_{\mathrm{X}}. \end{aligned} \tag{1}$$

Since we want hand-eye calibration free DCR (difficulty 1), we need to "guess" hand-eye relative pose, denoted by $\langle \tilde{\mathbf{R}}_{\mathrm{X}}, \tilde{\mathbf{t}}_{\mathrm{X}} \rangle$. Considering camera pose adjustment is achieved by moving hand, in practice, we have

$$\begin{aligned} \mathbf{c}_{i+1} &= \tilde{\mathbf{R}}_{\mathrm{A}}^i \mathbf{c}_i + \tilde{\mathbf{t}}_{\mathrm{A}}^i, \\ \mathbf{h}_{i+1} &= \tilde{\mathbf{R}}_{\mathrm{B}}^i \mathbf{h}_i + \tilde{\mathbf{t}}_{\mathrm{B}}^i, \end{aligned} \tag{2}$$

where $\langle \tilde{\mathbf{R}}_{\mathrm{A}}^i, \tilde{\mathbf{t}}_{\mathrm{A}}^i \rangle$ and $\langle \tilde{\mathbf{R}}_{\mathrm{B}}^i, \tilde{\mathbf{t}}_{\mathrm{B}}^i \rangle$ are deviated pose adjustment, caused by inaccurate guess of $\langle \tilde{\mathbf{R}}_{\mathrm{X}}, \tilde{\mathbf{t}}_{\mathrm{X}} \rangle$.

Considering difficulty 2, we cannot obtain real $\mathbf{t}_{\mathrm{A}}^i$ but its orientation $\bar{\mathbf{t}}_{\mathrm{A}}^i$ from two consecutively captured images. Its length should also be "guessed" by s_i, thus yielding another approximation

$$\hat{\mathbf{t}}_{\mathrm{A}}^i = s_i \bar{\mathbf{t}}_{\mathrm{A}}^i. \tag{3}$$

Combining Eqs. (1)–(3) leads to

$$\begin{aligned} \mathbf{c} &= \mathbf{R}_{\mathrm{A}}^{i+1} \mathbf{c}_{i+1} + \mathbf{t}_{\mathrm{A}}^{i+1}, \\ \mathbf{R}_{\mathrm{A}}^{i+1} &= \mathbf{R}_{\mathrm{A}}^i \mathbf{R}_{\mathrm{X}}^* {\mathbf{R}_{\mathrm{A}}^i}^{-1} {\mathbf{R}_{\mathrm{X}}^*}^{-1}, \\ \mathbf{t}_{\mathrm{A}}^{i+1} &= \mathbf{t}_{\mathrm{A}}^i + \triangle \mathbf{t}_{\mathrm{A}}^i, \\ \triangle \mathbf{t}_{\mathrm{A}}^i &= \mathbf{R}_{\mathrm{A}}^i \mathbf{t}_{\mathrm{X}} - \mathbf{R}_{\mathrm{A}}^i \mathbf{R}_{\mathrm{X}}^* \tilde{\mathbf{t}}_{\mathrm{X}} - \mathbf{R}_{\mathrm{A}}^i \mathbf{R}_{\mathrm{X}}^* {\mathbf{R}_{\mathrm{A}}^i}^{-1} \mathbf{q}, \\ \mathbf{q} &= \hat{\mathbf{t}}_{\mathrm{A}}^i - \tilde{\mathbf{t}}_{\mathrm{X}} + {\mathbf{R}_{\mathrm{X}}^*}^{-1} \mathbf{t}_{\mathrm{X}}, \end{aligned} \tag{4}$$

where $\mathbf{R}_{\mathrm{X}}^* = \mathbf{R}_{\mathrm{X}} \tilde{\mathbf{R}}_{\mathrm{X}}^{-1}$. Eq. (4) is a general hand-eye calibration free DCR model that shows the recurrence relation of camera pose between two adjacent adjustments. A feasible DCR process should guarantee to reduce the relative 3D rotational and translation pose displacement to zero rapidly. The major obstacle comes from the inaccurate guesses $\langle \tilde{\mathbf{R}}_{\mathrm{X}}, \tilde{\mathbf{t}}_{\mathrm{X}} \rangle$ and $\hat{\mathbf{t}}_{\mathrm{A}}^i$, due to Difficulty 1 and 2. It is clear to see that, if the two guesses are correct, Eq. (4) trivially collapses to a one-step relocalization.

[2] An industrial-grade 6D miniature hexapod could worth more than 50,000 USD and requires clean working environment. The repositioning system in our DCR platform worths less than 30,000 RMB.

3.2. An easy-to-understand strategy

Here we give an easy-to-understand strategy by first relocalizing 3D relative rotational pose to convergence via iteratively adjustments using Eq. (5), then reducing 3D relative translation to zero using the bisection-try method described in Theorem 2. The following two theorems theoretically guarantee that the influence of bounded hand-eye relative pose displacement $\langle \mathbf{R}_\mathrm{X}, \mathbf{t}_\mathrm{X} \rangle$ can be simply ignored in DCR. Specifically, we can just guess $\tilde{\mathbf{R}}_\mathrm{X} = \mathbf{I}$ and $\tilde{\mathbf{t}}_\mathrm{X} = \mathbf{0}$. Accordingly, Eq. (4) can be simplified to

$$\mathbf{R}_\mathrm{A}^{i+1} = \mathbf{R}_\mathrm{A}^{i} \mathbf{R}_\mathrm{X} {\mathbf{R}_\mathrm{A}^{i}}^{-1} {\mathbf{R}_\mathrm{X}}^{-1}, \tag{5}$$

$$\mathbf{t}_\mathrm{A}^{i+1} = \mathbf{t}_\mathrm{A}^{i} - \mathbf{R}_\mathrm{A}^{i} \mathbf{R}_\mathrm{X} {\mathbf{R}_\mathrm{A}^{i}}^{-1} (\hat{\mathbf{t}}_\mathrm{A}^{i} + {\mathbf{R}_\mathrm{X}}^{-1} \mathbf{t}_\mathrm{X}) + \mathbf{R}_\mathrm{A}^{i} \mathbf{t}_\mathrm{X}. \tag{6}$$

Theorem 1 ($\mathbf{R}_\mathrm{A}$ convergence)**.** *By the rotation adjustment strategy defined by Eq. (5), if $\theta_\mathrm{X} \leq \frac{\pi}{3}$, then $\theta^{i+1} \leq \theta^i$ and $\lim_{i\to\infty} \theta^i = 0$. θ^i and $\bar{\mathbf{e}}^i$ are the angle and axis of $\mathbf{R}_\mathrm{A}^i$. θ_X and $\bar{\mathbf{e}}_\mathrm{X}$ are the angle and axis of $\mathbf{R}_\mathrm{X}$.*

Proof. Using quaternion representation, we have

$$\begin{aligned} \mathbf{R}_\mathrm{A}^{i+1} &\simeq (\cos\tfrac{\theta^{i+1}}{2}, \bar{\mathbf{e}}^{i+1} \sin\tfrac{\theta^{i+1}}{2}), \\ \mathbf{R}_\mathrm{A}^{i} &\simeq (\cos\tfrac{\theta^{i}}{2}, \bar{\mathbf{e}}^{i} \sin\tfrac{\theta^{i}}{2}), \\ \mathbf{R}_\mathrm{X} &\simeq (\cos\tfrac{\theta_\mathrm{X}}{2}, \bar{\mathbf{e}}_\mathrm{X} \sin\tfrac{\theta_\mathrm{X}}{2}), \\ {\mathbf{R}_\mathrm{A}^{i}}^{-1} &\simeq (\cos\tfrac{\theta^{i}}{2}, -\bar{\mathbf{e}}^{i} \sin\tfrac{\theta^{i}}{2}), \\ {\mathbf{R}_\mathrm{X}}^{-1} &\simeq (\cos\tfrac{\theta_\mathrm{X}}{2}, -\bar{\mathbf{e}}_\mathrm{X} \sin\tfrac{\theta_\mathrm{X}}{2}). \end{aligned} \tag{7}$$

From Eqs. (5) and (7), we have

$$\begin{aligned} \mathbf{R}_\mathrm{A}^{i} \mathbf{R}_\mathrm{X} \simeq (&\cos\tfrac{\theta^i}{2}\cos\tfrac{\theta_\mathrm{X}}{2} - \sin\tfrac{\theta^i}{2}\sin\tfrac{\theta_\mathrm{X}}{2}\langle\bar{\mathbf{e}}^i, \bar{\mathbf{e}}_\mathrm{X}\rangle, \\ &\cos\tfrac{\theta^i}{2}\sin\tfrac{\theta_\mathrm{X}}{2}\bar{\mathbf{e}}_\mathrm{X} + \sin\tfrac{\theta^i}{2}\cos\tfrac{\theta_\mathrm{X}}{2}\bar{\mathbf{e}}^i \\ &+ \sin\tfrac{\theta^i}{2}\sin\tfrac{\theta_\mathrm{X}}{2}\,\bar{\mathbf{e}}^i \times \bar{\mathbf{e}}_\mathrm{X}), \end{aligned} \tag{8}$$

$$\begin{aligned} {\mathbf{R}_\mathrm{A}^{i}}^{-1} {\mathbf{R}_\mathrm{X}}^{-1} \simeq (&\cos\tfrac{\theta^i}{2}\cos\tfrac{\theta_\mathrm{X}}{2} - \sin\tfrac{\theta^i}{2}\sin\tfrac{\theta_\mathrm{X}}{2}\langle\bar{\mathbf{e}}^i, \bar{\mathbf{e}}_\mathrm{X}\rangle, \\ &-\cos\tfrac{\theta^i}{2}\sin\tfrac{\theta_\mathrm{X}}{2}\bar{\mathbf{e}}_\mathrm{X} - \sin\tfrac{\theta^i}{2}\cos\tfrac{\theta_\mathrm{X}}{2}\bar{\mathbf{e}}^i \\ &+ \sin\tfrac{\theta^i}{2}\sin\tfrac{\theta_\mathrm{X}}{2}\,\bar{\mathbf{e}}^i \times \bar{\mathbf{e}}_\mathrm{X}), \end{aligned} \tag{9}$$

where $\langle\bar{\mathbf{e}}^i, \bar{\mathbf{e}}_\mathrm{X}\rangle$ and $\bar{\mathbf{e}}^i \times \bar{\mathbf{e}}_\mathrm{X}$ denote the inner and cross product of $\bar{\mathbf{e}}^i$ and $\bar{\mathbf{e}}_\mathrm{X}$, respectively. Combing Eqs. (5), (8) and (9) yields

$$\begin{aligned} \cos\tfrac{\theta^{i+1}}{2} &= \cos^2\tfrac{\theta^i}{2} + (1 - 2\sin^2\tfrac{\theta_\mathrm{X}}{2})\sin^2\tfrac{\theta^i}{2} \\ &\quad + \sin^2\tfrac{\theta^i}{2}\sin^2\tfrac{\theta_\mathrm{X}}{2}\langle\bar{\mathbf{e}}^i, \bar{\mathbf{e}}_\mathrm{X}\rangle^2 \\ &\geq \cos^2\tfrac{\theta^i}{2} + \cos\theta_\mathrm{X}\sin^2\tfrac{\theta^i}{2}. \end{aligned} \tag{10}$$

Note, by the right-hand rule, $\theta^i \in [0, \pi]$ and $\theta_\mathrm{X} \in [0, \pi]$.

Define $f(\theta^i) = \cos^2\frac{\theta^i}{2} + \cos\theta_\mathrm{X}\sin^2\frac{\theta^i}{2} - \cos\frac{\theta^i}{2}$ that is an even function. We need to prove $f(\theta^i) \geq 0$ for $\theta^i \in [0, \pi]$. Since $f'(\theta^i) = \sin\frac{\theta^i}{2}[\cos\frac{\theta^i}{2}(\cos\theta_\mathrm{X} - 1) + \frac{1}{2}]$. Clearly, when $\theta_\mathrm{X} \leq \frac{\pi}{3}$, $f'(\theta^i) \geq 0$. Hence, $f(\theta^i)$ is non-decreasing in $[0, \pi]$. Meanwhile, due to $f(0) = 0$ and the fact that $f(\theta^i)$ is an even function, we have $f(\theta^i) \geq 0$ in $[-\pi, \pi]$, thus $\cos\frac{\theta^{i+1}}{2} \geq \cos\frac{\theta^i}{2}$ and $\theta^{i+1} \leq \theta^i$.

Moreover, note that $f'(\theta^i) \geq 0$ with the equality only occurs at $\theta^i = 0$. Therefore, θ^i will consistently be reduced to zero, i.e., $\lim_{i\to\infty} \mathbf{R}_\mathrm{A}^i = \mathbf{I}$ or $\lim_{i\to\infty} \theta^i = 0$ equivalently. □

Theorem 2 ($\mathbf{t}_\mathrm{A}$ bisection-try convergence)**.** *When eye relative rotation $\mathbf{R}_\mathrm{A}^i$ converges to $\mathbf{I}$, with current step size s_i, we measure $\langle \bar{\mathbf{t}}_\mathrm{A}^{i+1}, \bar{\mathbf{t}}_\mathrm{A}^i \rangle$. If $\langle \bar{\mathbf{t}}_\mathrm{A}^{i+1}, \bar{\mathbf{t}}_\mathrm{A}^i \rangle \geq 0$, we guess the length of $\mathbf{t}_\mathrm{A}^{i+1}$ by $s_{i+1} = s_i$, otherwise we do bisection $s_{i+1} := \frac{s_i}{2}$ and guess $\hat{\mathbf{t}}_\mathrm{A}^{i+1} = s_{i+1}\bar{\mathbf{t}}_\mathrm{A}^{i+1}$. Using this bisection-try method, just before the bisection, we strictly have $\|\mathbf{t}_\mathrm{A}^{i+1}\| \leq s_{i+1}$. As s_i is monotonically reduced to zero, so is $\|\mathbf{t}_\mathrm{A}^{i+1}\|$.*

Proof. If the relative rotation has converged to $\mathbf{I}$, Eq. (6) can be reduced to

$$\mathbf{t}_\mathrm{A}^{i+1} = \mathbf{t}_\mathrm{A}^{i} - \triangle^i = \mathbf{t}_\mathrm{A}^{i} - \mathbf{R}_\mathrm{X}\hat{\mathbf{t}}_\mathrm{A}^{i} = \mathbf{t}_\mathrm{A}^{i} - \mathbf{R}_\mathrm{X} s_i \bar{\mathbf{t}}_\mathrm{A}^{i}, \tag{11}$$

where s_i is the bisection step size before i-th adjustment. By Rodrigues' rotation formula, we have

$$\begin{aligned} \mathbf{R}_\mathrm{X}\mathbf{t}_\mathrm{A}^{i} &= \cos\theta_\mathrm{X}\mathbf{t}_\mathrm{A}^{i} + \sin\theta_\mathrm{X}(\bar{\mathbf{e}}_\mathrm{X} \times \mathbf{t}_\mathrm{A}^{i}) \\ &\quad + (1 - \cos\theta_\mathrm{X})\langle\bar{\mathbf{e}}_\mathrm{X}, \mathbf{t}_\mathrm{A}^{i}\rangle\bar{\mathbf{e}}_\mathrm{X}. \end{aligned} \tag{12}$$

Hence, $\langle \mathbf{R}_\mathrm{X}\mathbf{t}_\mathrm{A}^{i}, \mathbf{t}_\mathrm{A}^{i} \rangle \geq 0$, because

$$\begin{aligned} \langle \mathbf{R}_\mathrm{X}\mathbf{t}_\mathrm{A}^{i}, \mathbf{t}_\mathrm{A}^{i} \rangle &= \cos\theta_\mathrm{X}\|\mathbf{t}_\mathrm{A}^{i}\|^2 + 0 \\ &\quad + (1 - \cos\theta_\mathrm{X})\langle\bar{\mathbf{e}}_\mathrm{X}, \mathbf{t}_\mathrm{A}^{i}\rangle^2 \geq 0. \end{aligned} \tag{13}$$

That is, $\angle(\mathbf{t}_\mathrm{A}^{i}, \mathbf{R}_\mathrm{X}\mathbf{t}_\mathrm{A}^{i}) \leq \frac{\pi}{2}$, where $\angle(\mathbf{a}, \mathbf{b})$ denotes the angle between vectors $\mathbf{a}$ and $\mathbf{b}$. Since $\angle(\triangle^i, \mathbf{R}_\mathrm{X}\mathbf{t}_\mathrm{A}^{i}) = 0$, $\angle(\mathbf{t}_\mathrm{A}^{i}, \triangle^i) \leq \frac{\pi}{2}$ and $0 \leq \cos\angle(\mathbf{t}_\mathrm{A}^{i}, \triangle^i) \leq 1$.

Now, let us check the i-th iteration i when s bisection happens. We have $\angle(\mathbf{t}_\mathrm{A}^{i+1}, \mathbf{t}_\mathrm{A}^{i}) \geq \frac{\pi}{2}$. This means

$$\begin{aligned} \langle \mathbf{t}_\mathrm{A}^{i} - \triangle^i, \mathbf{t}_\mathrm{A}^{i} \rangle &= \|\mathbf{t}_\mathrm{A}^{i}\|^2 - \langle\triangle^i, \mathbf{t}_\mathrm{A}^{i}\rangle \\ &= \|\mathbf{t}_\mathrm{A}^{i}\|^2 - \|\mathbf{t}_\mathrm{A}^{i}\| s_i \cos\angle(\mathbf{t}_\mathrm{A}^{i}, \triangle^i) \\ &\leq 0. \end{aligned} \tag{14}$$

Hence, we have $\|\mathbf{t}_\mathrm{A}^{i}\|^2 \leq \|\mathbf{t}_\mathrm{A}^{i}\| s_i \cos\angle(\mathbf{t}_\mathrm{A}^{i}, \triangle^i) \leq \|\mathbf{t}_\mathrm{A}^{i}\| s_i$, which leads to $\|\mathbf{t}_\mathrm{A}^{i}\| \leq s_i$. □

Theorem 2 shows that using the bisection-try method, $\hat{\mathbf{t}}_\mathrm{A}^{i} = s_i\bar{\mathbf{t}}_\mathrm{A}^{i}$ can gradually approach to $\mathbf{t}_\mathrm{A}^{i}$. In this case, Eq. (6) and (11) can be further simplified to

$$\mathbf{t}_\mathrm{A}^{i+1} = \mathbf{t}_\mathrm{A}^{i} - \mathbf{R}_\mathrm{X}\mathbf{t}_\mathrm{A}^{i} = (\mathbf{I} - \mathbf{R}_\mathrm{X})\mathbf{t}_\mathrm{A}^{i}. \tag{15}$$

Therefore, we have $\|\mathbf{t}_\mathrm{A}^{i+1}\| \leq \|\mathbf{t}_\mathrm{A}^{i}\|$, if and only if (iff) the magnitudes of eigenvalues of $\mathbf{I} - \mathbf{R}_\mathrm{X}$ are not greater than 1. Clearly, since the eigenvalues of $\mathbf{R}_\mathrm{X}$ are 1 and $e^{\pm j\theta_\mathrm{X}}$, the eigenvalues of $\mathbf{I} - \mathbf{R}_\mathrm{X}$ are 0 and $1 - \cos\theta_\mathrm{X} \pm j\sin\theta_\mathrm{X}$, whose magnitudes are not greater than 1 iff $\cos\theta_\mathrm{X} \geq \frac{1}{2}$. That is, we need $\theta_\mathrm{X} \leq \frac{\pi}{3}$ to guarantee $\|\mathbf{t}_\mathrm{A}^{i+1}\| \leq \|\mathbf{t}_\mathrm{A}^{i}\|$ in Eq. (15).

Theorem 1 and 2 theoretically prove that this easy-to-understand strategy can lead to convergent DCR. Moreover, the bisection strategy to approach relative camera translation takes advantage of the RRA of inexpensive positioning platforms, thus its real repetitive accuracy is reliable.

3.3. The algorithm and implementation details

As shown in Fig. 1(a), there exist pair ambiguities between particular axes rotation and translation, e.g., pitch and height movement. That is, with the existence of translational displacement, during the first rotational relocalization stage, the theoretical DCR strategy may be not able to converge well, because in this case the 5-point algorithm is quite possible to generate unreliable large rotation estimation that is actually caused by translation. To overcome this problem, based on Theorem 1 and 2, we propose a practical DCR algorithm that jointly adjusts both 3D rotation and translation in the process. Detailed working flow of the proposed DCR algorithm is shown in Fig. 2 and Algorithm 1. Besides, separately relocalizing rotational and translational relative pose highly relies on the mechanical independence of rotation and translation platforms. In practice, some translation axes movement may cause extra rotational displacement, even the first stage of relative rotational pose relocalization does very well, subsequent translation axes movement may certainly jeopardize the final DCR accuracy. As verified by our extensive experiments, the rotation-translation joint relocalization algorithm constantly outperforms sequential adjustment strategy in both convergence rate and accuracy.

Algorithm 1 Practical 6D dynamic camera relocalization

Input: I^{ref}, initial and stopping moving step s_0 and $s_{\min}$.

1: Initialization: initialize platform to zero position, $s = s_0$, $\langle \tilde{\mathbf{R}}_{\text{X}} = \mathbf{I}, \tilde{\mathbf{t}}_{\text{X}} = \mathbf{0} \rangle$, $i = 1$, $\bar{\mathbf{t}}_{\text{A}}^0 = \mathbf{0}$;
2: Homography-based coarse camera relocalization to reduce error in the range of platform [7];
3: **while** $s > s_{\min}$ **do**
4: Capture current image I_i and compute $\langle \mathbf{R}_{\text{A}}^i, \bar{\mathbf{t}}_{\text{A}}^i \rangle$;
5: Rotate platform by $\tilde{\mathbf{R}}_{\text{B}}^i = \tilde{\mathbf{R}}_{\text{X}}^{-1} \mathbf{R}_{\text{A}}^i \tilde{\mathbf{R}}_{\text{X}} = \mathbf{R}_{\text{A}}^i$;
6: **if** $\bar{\mathbf{t}}_{\text{A}}^i \cdot \bar{\mathbf{t}}_{\text{A}}^{i-1} < 0$ **then**
7: Bisection $s = \frac{s}{2}$;
8: **end if**
9: $\hat{\mathbf{t}}_{\text{A}}^i = s\bar{\mathbf{t}}_{\text{A}}^i$;
10: Translate platform by $\tilde{\mathbf{t}}_{\text{B}}^i = \hat{\mathbf{t}}_{\text{A}}^i$;
11: i++;
12: **end while**

Output: DCR result I_{i-1}.

Specifically, since precision positioning platforms usually cannot have large moving range, we first use the homography-based CR [7] to roughly relocate camera to make its relative pose within the range of mechanical platforms. We then repeatedly capturing current image and calculate its 6D pose displacement $\langle \mathbf{R}_{\text{A}}^i, \bar{\mathbf{t}}_{\text{A}}^i \rangle$. We directly rotate $\mathbf{R}_{\text{A}}^i$ and check the direction consistency of $\bar{\mathbf{t}}_{\text{A}}^i$ and $\bar{\mathbf{t}}_{\text{A}}^{i-1}$. If they are not in the same direction and s is not small enough, this means last time move has passed the objective position, i.e., s is too large and need to be bisected $s \leftarrow \frac{s}{2}$ first and then to move $s\bar{\mathbf{t}}_{\text{A}}^i$; otherwise, we directly move $s\bar{\mathbf{t}}_{\text{A}}^i$. In practice, we initialize s_0 as $\frac{1}{5}$ full translation range of the platform. Larger or smaller s_0, if feasible, will only impact the convergence rate. Besides, to counteract inevitable mechanical gaps, for a single axis movement along a particular direction with length L, we first go $1.1L$ along that direction and then return $0.1L$ inversely.

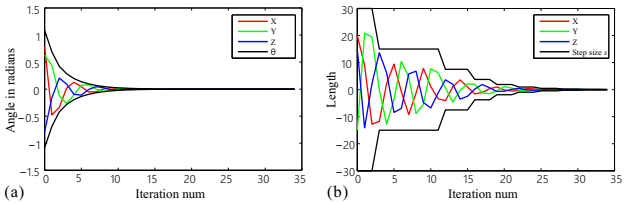

Figure 3. **Convergence of $\mathbf{R}_{\text{A}}$ and $\mathbf{t}_{\text{A}}$.** (a) and (b) represent the 3D rotation and translation convergence, respectively. Red, green and blue lines represent X, Y and Z axis respectively, and black lines indicate the angle θ and step size s in (a) and (b).

4. Experimental Results

Baselines. We compare our DCR with *manual* relocalization, *homography*-based relocalization [7] and computational rephotography [1, 33]. Note, manual relocalization just uses human visual judgment to manually steer camera pose. For better performance, we do manual relocalization very carefully in our experiments. The homography-based relocalization uses reference image and current image to generate a homography matrix that produces two navigation rectangles, guiding users to correct relative camera pose [7].

Criterion. To evaluate the performance of camera relocalization, we present feature-point displacement flow (FDF), a sparse field of feature-point displacement vectors, and average feature-point displacement (AFD) to quantitatively measure the relocalization accuracy:

$$\text{AFD}(\mathbf{P}_{\text{ref}}, \mathbf{P}_{\text{cur}}) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{P}_{\text{ref}}^i - \mathbf{P}_{\text{cur}}^i\|_2 \qquad (16)$$

where $\mathbf{P}_{\text{ref}}$ and $\mathbf{P}_{\text{cur}}$ are the matched feature-point coordinates in reference image and current relocalized image, respectively, n is the number of matches. In our experiments, we use SIFT feature detector, descriptor and robust matching with RANSAC correction.

4.1. Convergence and accuracy

Laboratory tests. Given the initialization of $\mathbf{R}_{\text{A}}$, $\mathbf{t}_{\text{A}}$ and fixed known $\mathbf{R}_{\text{X}}$, $\mathbf{t}_{\text{X}}$, we can simulate the proposed camera relocalization process in computer. Fig. 3 demonstrates the rotation and translation convergence process of our DCR. The initial setup is as follows: initial angle of $\mathbf{R}_{\text{A}}$, $\mathbf{t}_{\text{A}}$, angle of $\mathbf{R}_{\text{X}}$ and $\mathbf{t}_{\text{X}}$ are ($\frac{\pi}{4}$,$\frac{\pi}{5}$,-$\frac{\pi}{4}$), (20,-15,15), ($\frac{\pi}{8}$,$\frac{\pi}{8}$,$\frac{\pi}{8}$) and (15,10,10), respectively.

Fig. 3 is a typical DCR convergence process, which clearly shows that $\mathbf{t}_{\mathrm{A}}$ is bounded by the step size s during the whole DCR procedure. With the decrease of step size s, $\mathbf{t}_{\mathrm{A}}$ is accordingly descending. In contrast, the angle θ of $\mathbf{R}_{\mathrm{A}}$ is monotonically decreasing to zero. Fig. 3 indeed validates the theoretical convergence of our DCR.

Table 1. Average number of iterations and AFD of 10-times DCR tests of theoretical strategy and the proposed algorithm.

Theoretical strategy		The algorithm	
#iterations	AFD	#iterations	AFD
14.4	13.37	10.4	0.85

Convergence rate comparison of two strategies. Here, we provide detailed comparison of the convergence rate and accuracy of the proposed two DCR strategies. Table 1 compares the AFD and iteration number in 10 independent DCR tests using the two strategies. In Fig. 4, we visually compare the DCR accuracy of the two strategies. Note, the selected scene in Fig. 4 has minimum AFD value in 10-times tests using the theoretical strategy.

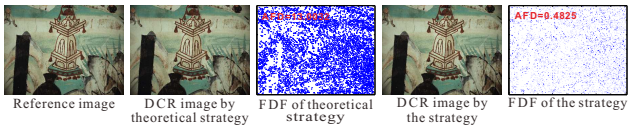

Figure 4. A visual comparison of DCR using theoretical strategy and the proposed algorithm on a near-planar scene.

From Table 1 and Fig. 4, we can clearly see that the proposed algorithm consistently converges faster than the theoretically strategy, and produces much better accuracy. This is because separately relocalizing 3D rotational and translational relative pose highly relies on the mechanical independence of rotation and translation platforms. In practice, some translation axes movement may cause extra rotational displacement, even after the first stage of relative rotational pose relocalization, which may inevitably jeopardize the final DCR accuracy.

DCR accuracy comparison. Physically measuring the relocalization error is an important and direct criterion to evaluate relocalization accuracy. However, ground truth is hard to obtain in practice. To this end, we establish scene-related FDF/AFD rulers by densely sampling the platform positions and orientations in 6 DoFs, with two adjacent samplings having only 0.1mm positional displacement or 0.01 degree rotation shift. Via such rulers, for a particular scene, we can quantitatively evaluate physically meaningful 6D DCR accuracy with radar chart.

Fig. 5 shows the physical DCR accuracy measurement on two different scenes, including one near-planar scene and one nonplanar scene. Table 2 gives the detailed statistics. From Fig. 5 and Table 2, we can clearly see that our DCR algorithm has reached subpixel accuracy and always

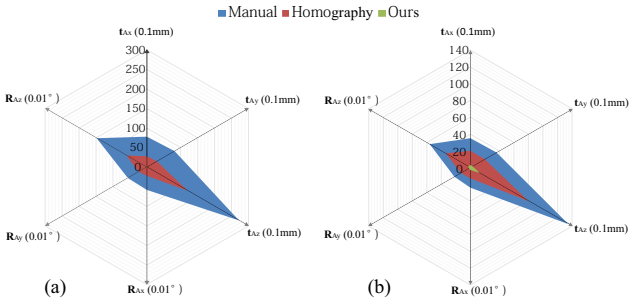

Figure 5. DCR accuracy measurement for two scenes, (a) near-planar scene, (b) nonplanar scene.

Table 2. DCR accuracy statistics for two scenes. Unit of measurement is 1 degree for rotation and 1mm for translation.

Near-planar	$\mathbf{R}_{\mathrm{A}x}$	$\mathbf{R}_{\mathrm{A}y}$	$\mathbf{R}_{\mathrm{A}z}$	$\mathbf{t}_{\mathrm{A}x}$	$\mathbf{t}_{\mathrm{A}y}$	$\mathbf{t}_{\mathrm{A}z}$	AFD
manual	0.58	0.56	1.48	7.8	8.2	27.2	20.18
homogrphy	0.22	0.24	0.6	2.8	3	12	8.82
ours	0.01	0.02	0.01	0.1	0.1	0.2	0.26
Nonplanar	$\mathbf{R}_{\mathrm{A}x}$	$\mathbf{R}_{\mathrm{A}y}$	$\mathbf{R}_{\mathrm{A}z}$	$\mathbf{t}_{\mathrm{A}x}$	$\mathbf{t}_{\mathrm{A}y}$	$\mathbf{t}_{\mathrm{A}z}$	AFD
manual	0.24	0.22	0.56	3.5	3.6	13.5	6.98
homogrphy	0.13	0.13	0.34	2	2	8	6.01
ours	0.03	0.02	0.03	0.3	0.3	1.3	0.81

outperforms the baseline manual and homography-based relocalization methods significantly. We also observe that, in all tested scenes, the physical deviations in Z axis for both rotation and translation are constantly larger than other axes. This mainly attributes to the particular focal length and FoV of the camera. That is, in our camera configuration, it requires relatively larger movement and rotation in Z axis to cause comparable pixel displacement in image plane than the translation and rotation in other axes.

Table 3. Comparison of physical camera relocalization error.

	CRP [1]	Ours	Ratio
Translation error (m)	1.135	**0.00038**	2987
Rotation error (○)	NA	0.02	NA

We also compare the proposed DCR with a state-of-the-art computational rephotography (CRP) tool, rephoto [33]. Fig. 6 shows the comparative results in two outdoor scenes. Table 3 gives a comparison of physical camera relocalization error of our DCR and another CRP method [1]. It is clear that, compared to CRP, the proposed DCR have much higher physical relocalization accuracy.

4.2. Active panorama acquisition

With the proposed DCR, we can realize active acquisition of high-quality panorama. Fig. 7 illustrates the detailed working flow of active HD panorama acquisition by the proposed DCR. Intuitively, we repeatedly relocalize the "left" part of current camera FoV to the same 6D pose of the reference image, which is the "right" part of the last-time captured image. With high-quality camera calibration and relocalization, this process guarantees to produce

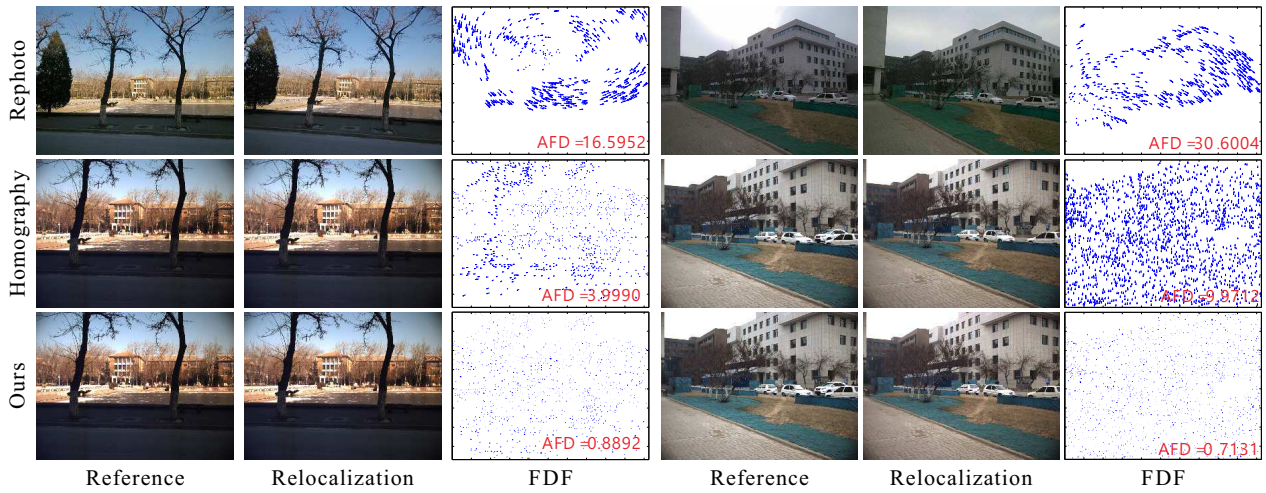

Figure 6. Accuracy of our DCR with 2 state-of-the-art competitors: Rephoto [33] and homography-based coarse camera relocalization [7].

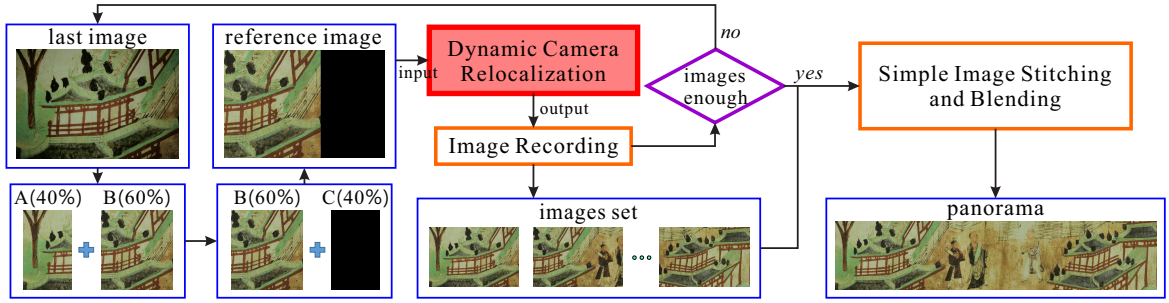

Figure 7. Working flow of DCR-based active panorama acquisition. Detailed DCR process in red block is shown in Fig. 2.

seamless panoramic images, by direct image stitching, with theoretically "unlimited" resolution, as such panorama can be actively captured as long and big as possible. Note, to guarantee the stitching quality, we suggest the ratio of common part of two successive images is not less than 30%. As shown in Fig. 8, compared to state-of-the-art image stitching method, like APAP [36], the proposed DCR-based active panorama acquisition is able to generate warping and artifact free panoramic image, even for highly-structured scenes taken at very close distance. Fig. 1 (f) shows a HD panorama of ancient mural acquired by our approach.

4.3. Minute change monitoring of ancient murals

Another promising application of the proposed DCR is long-time-interval minute change monitoring of ancient Mogao murals, which, as discussed in Sec. 1, is an open real-world problem. Although ancient Mogao murals have been seriously protected, they still suffer from many types of deteriorations caused by various environmental and human influences. As a result, their status is constantly changing in a very low speed. It is critically important to provide feasible imaging method, through which fine-grained changes become visually apparent thus timely detection and accurate measurement of such minute changes can trigger and support multi-disciplinary protective preservation. However, current fact is no suitable method and equipment can be used for accurate minute change imaging within the unrestricted environment in Dunhuang Mogao Grottoes.

Using our inexpensive DCR platform shown in Fig. 1(g), we selected 46 monitoring spots (very small regions) from 11 real Dunhuang caves and have conducted twice DCR-based image capturing in June 2014 and July 2015, respectively. The proposed approach and platform have successfully discovered 0.1mm-level minute changes (measured by close-range photogrammetry) in 31 monitoring spots (67%). This is truly a breakthrough, considering practical state-of-the-art way is simply naked-eye observation by experts that can only support object-level change detection during 100-year period. Some realistic yearly DCR results and corresponding changes derived by image-difference-aided human labeling are shown in Fig. 9. Our yearly monitoring data also find that caves open for tourists have 14 times faster deterioration speed than closed caves. This, for the first time, provides real data evidence about the influence of tourism to cultural heritages during 1 year period.

5. Conclusion

In this paper, we have proposed an inexpensive 6D dynamic camera relocalization approach. We theoretically

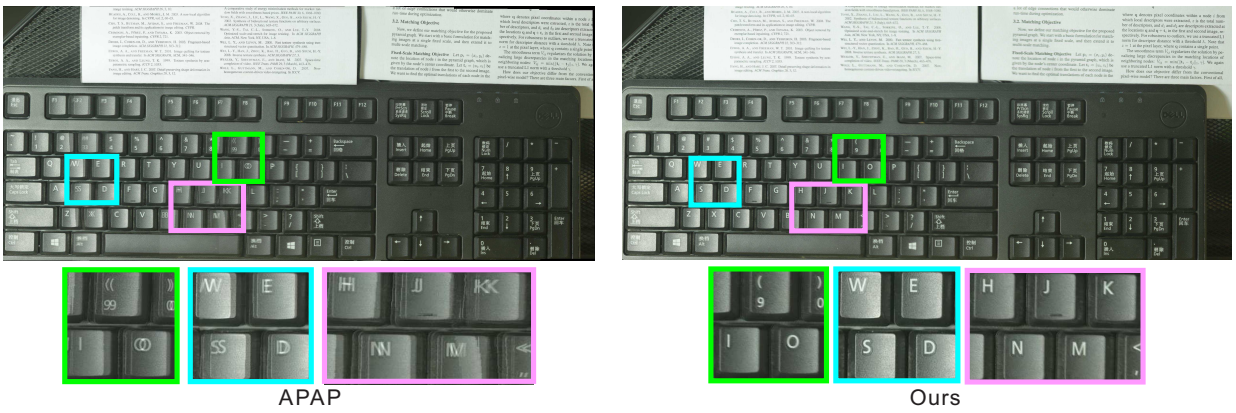

Figure 8. Comparison of panorama acquisition using state-of-the-art image stitching method APAP [36] and the proposed approach.

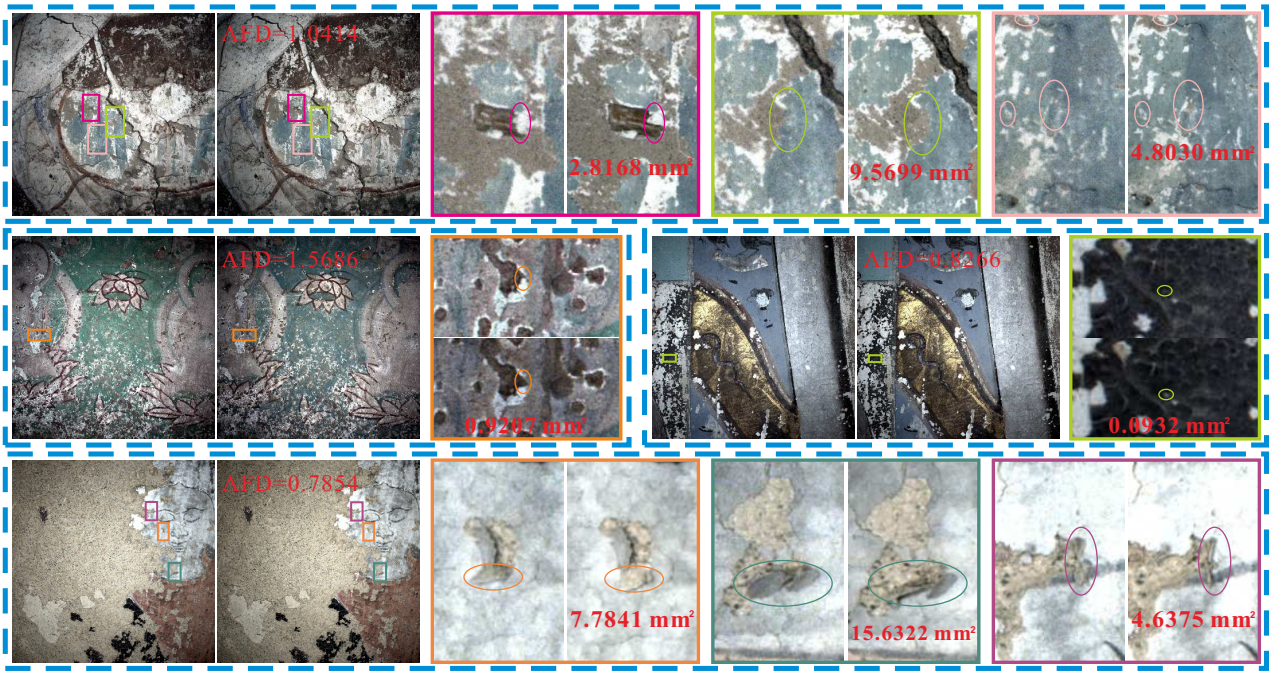

Figure 9. Real minute deteriorations of Dunhuang Mogao murals discovered by the proposed DCR from June 2014 and July 2015.

prove that both 3D rotational and translational pose displacement can be effectively reduced to zero based on low-cost RRA-reliable platform, without hand-eye calibration. We also find that, to make the proposed strategy work, the feasible upper bound of hand-eye 3D orientation displacement angle should be less than $\frac{\pi}{3}$. Extensive tests validate the effectiveness and accuracy of our approach. More importantly, we demonstrate the promising applications of our approach in solving two challenging real-world problems: 1) active acquisition of seamless high-definition panorama image; and 2) 0.1mm-level minute deterioration monitoring of very-slowly-changed ancient murals in Dunhuang Mogao Grottoes. Our work indeed shows the great potentials of combining algorithms with common hardware in active vision. In near future, we plan to further accelerate our DCR approach based on reliable superpixel segmentation and matching [4, 19, 6, 38] and apply it to solve more real-world change monitoring problems. Besides, we want to extend the proposed DCR model to faithfully relocalize hand coordinate system from multiple images. We are also interested in dynamic lighting recurrence from single image.

Acknowledgements We thank Xudong Wang, Bomin Su, Mingming Wang, Gangquan Chen, Qinglin Guo, Xiaowei Wang, Bolong Chai, Shujun Ding, Shengli Sun of Dunhuang Academy China for valuable discussions on slow-and-minute change monitoring of ancient murals. We thank Jiliang Sun, Jiawan Zhang, Qifeng Yue, Nan Zhang, Rui Huang, Yifeng Zhang, Dongrui Xiao for their contributions to our platform development and onsite data collection. This work is supported by the National Science and Technology Support Project (2013BAK01B01, 2013BAK01B05) and NSFC (61572354, 61272266).

References

[1] S. Bae, A. Agarwala, and F. Durand. Computational rephotography. *ACM TOG*, 29(3):1–15, 2010.

[2] A. J. Davison. Real-time simultaneous localization and mapping with a single camera. In *ICCV*, 2003.

[3] Y. Fang, W. E. Dixon, D. M. Dawson, and P. Chawda. Homography-based visual servo regulation of mobile robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(5):1041–1050, 2005.

[4] 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.

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

[6] 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.

[7] W. Feng, F.-P. Tian, Q. Zhang, N. Zhang, L. Wan, and J. Sun. Fine-grained change detection of misaligned scenes with varied illuminations. In *ICCV*, 2015.

[8] B. Glocker, S. Izadi, J. Shotton, and A. Criminisi. Real-time RGB-D camera relocalization. In *ISMAR*, 2013.

[9] D. Guggenheim, A. Gore, L. Bender, S. Z. Burns, L. David, and G. W. Documentary. *An inconvenient truth*. Paramount Classics Hollywood, CA, 2006.

[10] A. Guzman-Rivera, P. Kohli, B. Glocker, J. Shotton, T. Sharp, A. Fitzgibbon, and S. Izadi. Multi-output learning for camera relocalization. In *CVPR*, 2014.

[11] F. C. Hall. *Photo point monitoring handbook: Part A-field procedures*. USDA Forest Service, 2002.

[12] R. Hartley and F. Kahl. Global optimization through rotation space search. *IJCV*, 82(1):64–79, 2009.

[13] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *IJCV*, 103(3):267–305, 2013.

[14] J. Heller, M. Havlena, and T. Pajdla. A branch-and-bound algorithm for globally optimal hand-eye calibration. In *CVPR*, 2012.

[15] A. Irschara, C. Zach, J.-M. Frahm, and H. Bischof. From structure-from-motion point clouds to fast location recognition. In *CVPR*, 2009.

[16] N. Jiang, Z. Cui, and P. Tan. A global linear method for camera pose registration. In *ICCV*, 2013.

[17] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *ISMAR*, 2007.

[18] K.-T. Lee, S.-J. Luo, and B.-Y. Chen. Rephotography using image collections. In *Computer Graphics Forum*, 2011.

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

[20] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[21] G. L. Mariottini, G. Oriolo, and D. Prattichizzo. Image-based visual servoing for nonholonomic mobile robots using epipolar geometry. *IEEE Transactions on Robotics*, 23(1):87–100, 2007.

[22] R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In *CVPR*, 2010.

[23] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *ISMAR*, 2011.

[24] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison. DTAM: Dense tracking and mapping in real-time. In *ICCV*, 2011.

[25] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE TPAMI*, 26(6):756–770, 2004.

[26] D.-H. Park, J.-H. Kwon, and I.-J. Ha. Novel position-based visual servoing approach to robust global stability under field-of-view constraint. *IEEE Transactions on Industrial Electronics*, 59(12):4735–4752, 2012.

[27] V. Pradeep, C. Rhemann, S. Izadi, C. Zach, M. Bleyer, and S. Bathiche. MonoFusion: Real-time 3D reconstruction of small scenes with a single web camera. In *ISMAR*, 2013.

[28] T. Ruland, T. Pajdla, and L. Kruger. Globally optimal hand-eye calibration. In *CVPR*, 2012.

[29] Y. Seo, Y.-J. Choi, and S. W. Lee. A branch-and-bound algorithm for globally optimal calibration of a camera-and-rotation-sensor system. In *ICCV*, 2009.

[30] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, and A. Fitzgibbon. Scene coordinate regression forests for camera relocalization in RGB-D images. In *CVPR*, 2013.

[31] H. Stewenius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006.

[32] J. E. Wells II. *Western landscapes, western images: a rephotography of US Highway 89*. PhD thesis, Kansas State University, 2012.

[33] R. West, A. Halley, D. Gordon, J. O'Neil-Dunne, and R. Pless. Collaborative rephotography. In *ACM SIGGRAPH Studio Talks*, 2013.

[34] B. Williams, G. Klein, and I. Reid. Automatic relocalization and loop closing for real-time monocular SLAM. *IEEE TPAMI*, 33(9):1699–1712, 2011.

[35] K. Xu, H. Huang, H. Li, P. Long, J. Caichen, W. Sun, and B. Chen. Autoscanning for coupled scene reconstruction and proactive object analysis. In *ACM SIGGRAPH Asia*, 2015.

[36] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter. As-projective-as-possible image stitching with moving dlt. *IEEE TPAMI*, 36(7):1285–1298, 2014.

[37] G. Zhang, J. Jia, T.-T. Wong, and H. Bao. Consistent depth maps recovery from a video sequence. *IEEE TPAMI*, 31(6):974–988, 2009.

[38] S. Zhang, W. Feng, J. Zhang, and C.-M. Pun. Bag of squares: A reliable model of measuring superpixel similarity. In *ICME*, 2014.

[39] X. Zhang, Y. Fang, and X. Liu. Motion-estimation-based visual servoing of nonholonomic mobile robots. *IEEE Transactions on Robotics*, 27(6):1167–1175, 2011.