MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Fusion-Based Image Correlations Framework For Strain Measurement

Shi, Laixi; Liu, Dehong; Umeda, Masaki; Hana, Norihiko

TR2021-012 June 06, 2021

Abstract

We address the problem of enabling two-dimensional digital image correlation (DIC) for strain measurement on large three-dimensional objects with curved surfaces. It is challenging to acquire full-field qualified images of the surface required by DIC due to geometric distortion and the narrow visual field of the surface that a single image can cover. To overcome this issue, we propose an end-to-end DIC framework incorporating the image fusion principle to achieve full-field strain measurement over the curved surface. With a sequence of blurry images as inputs, we first recover sharp images using blind deconvolution, then project recovered sharp images to the curved surface using camera poses estimated by our proposed perspective-n-point (PnP) method called RRWLM. Images on the curved surface are stitched and then unfolded for strain analysis using DIC. Numerical experiments are conducted to validate our framework using RRWLM with comparisons to existing methods

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

FUSION-BASED DIGITAL IMAGE CORRELATION FRAMEWORK FOR STRAIN MEASUREMENT

Laixi Shi[1], *Dehong Liu*[2], *Masaki Umeda*[3], *and Norihiko Hana*[3]

[1] Carnegie Mellon University, PA, USA
[2] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA
[3] Mitsubishi Electric, Japan

ABSTRACT

We address the problem of enabling two-dimensional digital image correlation (DIC) for strain measurement on large three-dimensional objects with curved surfaces. It is challenging to acquire full-field qualified images of the surface required by DIC due to geometric distortion and the narrow visual field of the surface that a single image can cover. To overcome this issue, we propose an end-to-end DIC framework incorporating the image fusion principle to achieve full-field strain measurement over the curved surface. With a sequence of blurry images as inputs, we first recover sharp images using blind deconvolution, then project recovered sharp images to the curved surface using camera poses estimated by our proposed perspective-n-point (PnP) method called **RRWLM**. Images on the curved surface are stitched and then unfolded for strain analysis using DIC. Numerical experiments are conducted to validate our framework using **RRWLM** with comparisons to existing methods.

Index Terms— blind deconvolution, digital image correlation (DIC), image fusion and stitching, perspective-n-point (PnP).

1. INTRODUCTION

Strain measurement of materials subjected to loadings or mechanical damages is an essential task in various industrial applications [1–3]. For strain measurement, aside from the widely used pointwise strain gauge technique, digital image correlation (DIC) as a non-contact and a non-interferometric optical technique attracts a lot of attentions for its capability of providing full-field strain distribution of the surface using a simple experimental setup [4–7]. DIC is performed by comparing the digital gray intensity images of the surface before and after deformation, taking derivative of pixel displacement as a measure of strain at the pixel.

In some of the applications, it is of great interest to perform full-field two-dimensional (2D) DIC analysis [8] on the curved surface of large 3D objects. DIC has strict requirements on images taken before and after distortion for accurate pixel displacement, such as image resolution, image registration, and compensation of camera lens distortion, *etc*., since the displacements under strain are generally subtle (typically sub-pixel level) for most industrial materials [9]. Therefore, the requirements in target scenarios lead to two limitations for existing 2D DIC analysis. First, the DIC method is usually limited to 2D planar object surfaces rather than 3D curved surfaces [4, 5]. Second, the DIC method is usually restricted to small surfaces due to the extremely high pixel resolution requirement of images for DIC analysis.

In this work, we propose an end-to-end fusion-based DIC framework to enable strain measurement along the 3D object curved surface in large size using a single camera. We first use a moving camera over the 3D large surface to acquire a sequence of 2D blurry images of the surface texture. With these blurry observations, we recover the corresponding sharp images using blind deconvolution and project the pixels in them to the 3D surface using camera poses estimated by our proposed robust perspective-n-Point (PnP) method for image fusion. The stitched 3D surface images before and after deformation are unfolded to two 2D fused ones respectively, converting the 3D strain measurement into a 2D one for further DIC analysis. Since the displacements are subtle as mentioned before, their derivatives and corresponding strains are extremely sensitive to the fused image quality. Thus, the most daunting challenge in the pipeline is the stringent accuracy requirement (at least sub-pixel level) of the image fusion method for valid strain measurement, which requirement limits the application of existing PnP methods in our experiments. To address these limitations and challenges of 2D DIC analysis for target scenarios, our main contributions are as follows:

1. We propose an end-to-end DIC framework incorporating image fusion to the strain measurement pipeline. It extends the range of DIC-based strain measurement applications to the curved surface of 3D objects in large size.
2. We propose a two-stage method based on PnP method and bundle adjustment principle for image fusion. Our method outperforms state-of-arts and achieves applicable image fusion accuracy for strain measurement by DIC analysis.

1.1. Related Work

A lot of efforts have been made on 3D DIC methods based on a binocular stereo vision or a multi-camera system surrounding involving precise calibration and image stitching, which are difficult to operate in various scenarios [10–13]. This work utilizes the images captured by a single ordinary moving camera [6,7] rather than a well-calibrated multi-camera system. Furthermore, the proposed framework incorporates image fusion and camera pose estimation to automatically stitch a large number of images of the full-field curved surface under test.

This work extends the range of applications based on image fusion and stitching [14] to strain measurement in mechanical engineering. The proposed framework decouples the image fusion problem into a sequence of well-known PnP problems, which have been widely explored by using both non-iterative and iterative methods [15–20]. Some are with extra outlier rejection [15, 18] or incorporate the observation uncertainty information [21]. The proposed image fusion method combining the bundle adjustment principle and an iterative PnP method which outperforms existing PnP methods and achieves applicable fusion accuracy.

Laixi Shi performed this work as an intern at MERL. The authors would like to thank Dr. Yanting Ma for fruitful discussions.

2. FUSION-BASED DIC FRAMEWORK

2.1. Image Acquisition Model and Problem Formulation

Without loss of generality, we consider the strain measurement of a cylinder surface which is of interest in many applications. For image acquisition, a moving camera captures a sequence of images $\{\boldsymbol{Y}_i\}_{i=1}^{p}$ for the cylindrical surface texture $\boldsymbol{U}_b$ before deformation, and $\{\boldsymbol{Y}_i'\}_{i=1}^{q}$ for $\boldsymbol{U}_f$ after deformation, as illustrated in Fig.1. Each sequence consists of p (or q) images in order which overlapping with their neighbors. In the following sections, we mainly show the model and analysis for the sequence $\{\boldsymbol{Y}_i\}_{i=1}^{p}$ without loss of generality.

Since out-of-focus blur is a common image degradation phenomenon, we consider a six degree of freedom (6-DOF) pinhole camera model [14] with a camera lens' point spread function (PSF) (blur kernel) $\boldsymbol{K} \in \mathbb{R}^{(2r_g+1)\times(2r_g+1)}$, which is assumed to be a truncated Gaussian kernel [22]:

$$\boldsymbol{K}(x,y) = \begin{cases} \frac{1}{C_1}\exp(\frac{-(x^2+y^2)}{2\sigma^2}) & \sqrt{x^2+y^2} \leqslant r_g \\ 0 & \sqrt{x^2+y^2} > r_g, \end{cases} \tag{1}$$

where r_g is the radius, C_1 is the normalization term to ensure the energy of the PSF $\sum_{x,y}\frac{1}{C_1}\exp(\frac{-(x^2+y^2)}{2\sigma^2}) = 1$. Then the captured images $\{\boldsymbol{Y}_i\}_{i=1}^{p}$ can be modeled as [23, 24]:

$$\boldsymbol{Y}_i = \boldsymbol{K} \circledast \boldsymbol{X}_i, \quad i = 1, 2, \cdots, p, \tag{2}$$

where $\circledast$ denotes the convolution operation, $\boldsymbol{X}_i \in \mathbb{R}^{m\times n}$ is the i-th sharp camera focal plane image in the sequence $\{\boldsymbol{X}_i\}_{i=1}^{p}$.

Each pixel $\boldsymbol{x} = [x, y]^\top$ in $\boldsymbol{X}_i$ is projected from a pixel $\boldsymbol{u} = [x_u, y_u, z_u]^\top$ on the 3D surface according to:

$$\begin{bmatrix}\boldsymbol{x}\\1\end{bmatrix} = \frac{1}{v}\boldsymbol{P}_s\begin{bmatrix}\boldsymbol{R} & \boldsymbol{T}\\0 & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{u}\\1\end{bmatrix} = \frac{1}{v}\begin{bmatrix}f & 0 & 0 & 0\\0 & f & 0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{R} & \boldsymbol{T}\\0 & 1\end{bmatrix}\begin{bmatrix}x_u\\y_u\\z_u\\1\end{bmatrix}, \tag{3}$$

where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and $\boldsymbol{T} \in \mathbb{R}^{3}$ are the rotation matrix and the translation vector respectively, depending on the camera pose of $\boldsymbol{X}_i$, v is a pixel-dependent scalar projecting the pixel to the focal plane, and $\boldsymbol{P}_s$ is the perspective matrix of the camera.

Note that each image $\boldsymbol{Y}_i$ ($\boldsymbol{Y}_i'$) in the sequence covers a narrow field of the cylinder surface $\boldsymbol{U}_b$ ($\boldsymbol{U}_f$), whose camera pose is not known exactly except for the first image due to random perturbations of the camera. With the prior knowledge of the cylinder geometry, our goal is to recover the full-field unfolded images of the curved surface based on $\{\boldsymbol{Y}_i\}_{i=1}^{p}$ and $\{\boldsymbol{Y}_i'\}_{i=1}^{q}$ such that the strain on the cylindrical surface can be analyzed using 2D DIC. In the following subsections, we will introduce our proposed framework including image deblurring, image fusion, and DIC, as illustrated in Fig.1.

2.2. Image Deblurring

The goal of this module is to recover sharp focal plane images $\{\boldsymbol{X}_i\}_{i=1}^{p}$ and the unknown blur kernel $\boldsymbol{K}$ simultaneously from the blurry observations $\{\boldsymbol{Y}_i\}_{i=1}^{p}$ in (2). To this end, we formulate the blind deconvolution problem as

$$\min_{\boldsymbol{K},\{\boldsymbol{X}_i\}} \sum_{i=1}^{p}\left(\frac{\beta}{2}\left\|\boldsymbol{Y}_i - \boldsymbol{K}\circledast\boldsymbol{X}_i\right\|_F^2 + \sum_{j=1}^{m\cdot n}\left\|\boldsymbol{D}_j\boldsymbol{X}_i\right\|_2\right) + \boldsymbol{I}_{\mathcal{G}}(\boldsymbol{K}), \tag{4}$$

where $\|\cdot\|_F$ represents the Frobenius norm of a matrix, $\boldsymbol{I}_{\mathcal{G}}(\cdot)$ is the indicator function to ensure $\boldsymbol{K}$ is a truncated Gaussian kernel, $\boldsymbol{D}_j$ represents the derivative of $\boldsymbol{X}_i$ at pixel j in both x and y directions, and β is a weight depending on the noise level of the image $\boldsymbol{Y}_i$. In the objective function, the first term is a data fidelity term; and the second term is total variation (TV) [25] which is a widely used regularization term to preserve sharpness of the image. Equation (4) is solved by alternating minimization with respect to $\boldsymbol{K}$ and $\{\boldsymbol{X}_i\}_{i=1}^{p}$. Especially, we update $\{\boldsymbol{X}_i\}_{i=1}^{p}$ [26] utilizing circular convolution with the periodic boundary assumption on $\{\boldsymbol{X}_i\}_{i=1}^{p}$ for fast computation by FFT [4].

To obtain a great initialization $\boldsymbol{K}_0$ of the blur kernel, we use Wiener filter [4] by minimizing the normalized sparsity measure [23] in the possible region of σ as

$$\boldsymbol{K}_0 = \operatorname*{argmin}_{\boldsymbol{K}} \sum_{i=1}^{L} \frac{\left\|\nabla_x\overline{\boldsymbol{X}}_i(\boldsymbol{K},\boldsymbol{Y}_i)\right\|_1}{\left\|\nabla_x\overline{\boldsymbol{X}}_i(\boldsymbol{K},\boldsymbol{Y}_i)\right\|_2} + \frac{\left\|\nabla_y\overline{\boldsymbol{X}}_i(\boldsymbol{K},\boldsymbol{Y}_i)\right\|_1}{\left\|\nabla_y\overline{\boldsymbol{X}}_i(\boldsymbol{K},\boldsymbol{Y}_i)\right\|_2}, \tag{5}$$

where $\overline{\boldsymbol{X}}_i(\boldsymbol{K},\boldsymbol{Y}_i) = Wiener(\boldsymbol{K},\boldsymbol{Y}_i)$ is the filtered image of $\boldsymbol{Y}_i$ with kernel $\boldsymbol{K}$, ∇_x and ∇_y denote the derivatives in x and y directions respectively, and L is the number of images used.

2.3. Image Fusion

In this module, we reconstruct the super-resolution texture over the 3D object curved surface using the deblurred sequence of images $\{\hat{\boldsymbol{X}}_i\}_{i=1}^{p}$ for DIC analysis.

2.3.1. *Camera pose estimation*

Without loss of generality, we consider the problem of estimating the camera pose of a target deblurred image $\hat{\boldsymbol{X}}_i$ by registering it with an overlapping reference image $\hat{\boldsymbol{X}}_j$ of which the camera pose is known.

Firstly, we acquire the well-known SIFT [27] feature point sets $\boldsymbol{\Omega}_i^{\text{SIFT}} = \{\boldsymbol{x}_i\}$ in the target image $\hat{\boldsymbol{X}}_i$ and $\boldsymbol{\Omega}_j^{\text{SIFT}} = \{\boldsymbol{x}_j\}$ in the reference $\hat{\boldsymbol{X}}_j$. Then we seek a set of matched feature points $\mathcal{A}_{(j,i)} = \{(\boldsymbol{x}_j^m, \boldsymbol{x}_i^m) | \boldsymbol{x}_j^m \in \boldsymbol{\Omega}_j^{\text{SIFT}}, \boldsymbol{x}_i^m \in \boldsymbol{\Omega}_i^{\text{SIFT}}, m = 1, 2, ...\}$ satisfying [28]

$$\left\|\boldsymbol{a}(\boldsymbol{x}_i^m) - \boldsymbol{a}(\boldsymbol{x}_j^m)\right\|_2 \leqslant C_2 \cdot \min_{\boldsymbol{x}\in\boldsymbol{\Omega}_j^{\text{SIFT}}\backslash\boldsymbol{x}_j^m} \left\|\boldsymbol{a}(\boldsymbol{x}_i^m) - \boldsymbol{a}(\boldsymbol{x})\right\|_2, \tag{6}$$

where $\boldsymbol{a}(\boldsymbol{x})$ denotes the SIFT feature vector at the pixel $\boldsymbol{x}$, $\boldsymbol{\Omega}_j^{\text{SIFT}}\backslash\boldsymbol{x}_j^m$ is the set of $\boldsymbol{\Omega}_j^{\text{SIFT}}$ excluding $\boldsymbol{x}_j^m$, and $0 < C_2 \leqslant 1$ is a constant chosen to remove feature outliers, typically $C_2 = 0.7$.

We project each feature point $\boldsymbol{x}_j^m$ in $\mathcal{A}_{(j,i)}$ to the 3D surface and get the corresponding set of points $\{\boldsymbol{u}_j^m = (x_{u_j}^m, y_{u_j}^m, z_{u_j}^m)\}$, using (3) with the pose of $\hat{\boldsymbol{X}}_j$ and the object geometry. Then the camera pose estimation problem becomes the widely known PnP problem to estimate the camera pose using the point set $\mathcal{M}_{(j,i)} = \{(\boldsymbol{u}_j^m, \boldsymbol{x}_i^m)\}$ [16].

PnP problem can usually be formulated as a nonlinear sum of least squares problem [17]. Considering that $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$ holds in $\boldsymbol{R} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3]^\top$, we use $\boldsymbol{h} = [\boldsymbol{r}_1^\top, \boldsymbol{r}_2^\top, \boldsymbol{T}^\top]^\top \in \mathbb{R}^9$ to denote unknown parameters of the camera pose. Then the camera pose $\boldsymbol{h}_i$ associated with $\hat{\boldsymbol{X}}_i$ can be achieved by seeking $\boldsymbol{h}$ that [15, 29]:

$$\begin{aligned} \min_{\boldsymbol{h}} g(\boldsymbol{h}|\mathcal{M}_{(j,i)}) = \sum_{(\boldsymbol{u}_j^m, \boldsymbol{x}_i^m)\in\mathcal{M}_{(j,i)}} w_m \left\|\hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m, \boldsymbol{h}) - \boldsymbol{x}_i^m\right\|_2^2, \\ \text{s.t. } \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}, \end{aligned} \tag{7}$$

where $\hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m, \boldsymbol{h})$ is the projection result from the 3D point $\boldsymbol{u}_j^m$ to the camera focal plane of $\hat{\boldsymbol{X}}_i$ with respect to the camera pose $\boldsymbol{h}$ using

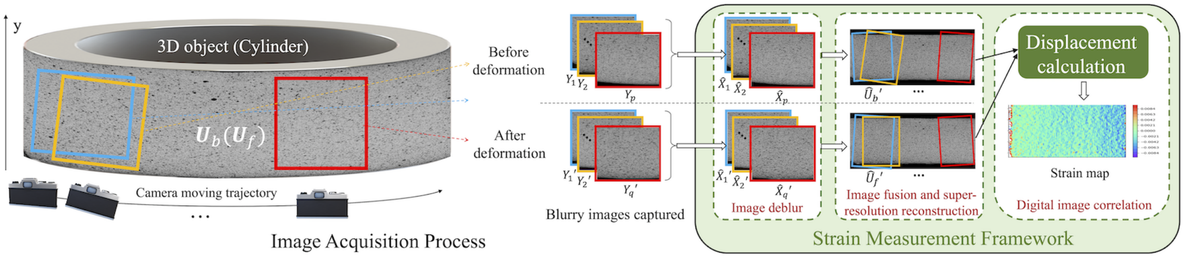


Fig. 1. The pipeline of the image acquisition and the strain measurement framework.

(3), $\boldsymbol{R}$ is determined by $\boldsymbol{h}$ as above, and $w_m = \frac{1}{\left\|\hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m,\boldsymbol{h})-\boldsymbol{x}_i^m\right\|_2^\alpha}$ represents the inverse of the measurement error for the m-th feature pair, for $m = 1, \cdots, |\mathcal{M}_{(j,i)}|$, and typically $\alpha = 0.5$.

To solve this problem, we utilize the widely used Levenberg-Marquardt algorithm (LM) [29] in conjunction with the projection operator $\mathcal{P}(\cdot)$ to keep the orthonormality of the rotation matrix $\boldsymbol{R}$. Given the present estimation $\boldsymbol{h}^{(t)}$, one step update $\boldsymbol{h}^{(t+1)} = \boldsymbol{h}^{(t)} + \Delta\boldsymbol{h}$ for (7) by LM can be seen as the interpolation of the greedy descent and Gauss-Newton update with

$$\Delta\boldsymbol{h} = \left(\boldsymbol{H} + \lambda\,\mathrm{diag}(\boldsymbol{H})\right)^{-1}\boldsymbol{b}, \tag{8}$$

where $\boldsymbol{H} = \sum_{\mathcal{M}_{(j,i)}} w_m \frac{\nabla\hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m,\boldsymbol{h})}{\nabla\boldsymbol{h}}\left[\frac{\nabla\hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m,\boldsymbol{h})}{\nabla\boldsymbol{h}}\right]^\top$ is the Hessian matrix, $\boldsymbol{b} = \sum_{\mathcal{M}_{(j,i)}} w_m \left[\frac{\nabla\hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m,\boldsymbol{h})}{\nabla\boldsymbol{h}}\right]^\top \left[\boldsymbol{x}_i^m - \hat{\boldsymbol{x}}_i(\boldsymbol{u}_j^m,\boldsymbol{h})\right]$, and λ is a parameter varying with iterations to determine the interpolation level accordingly.

The projection operator $\mathcal{P}(\boldsymbol{h})$ is defined to orthonormalize $\boldsymbol{r}_1, \boldsymbol{r}_2$. We revise the method used in [30] which approximately apportions half of the error to $\boldsymbol{r}_1'$ and $\boldsymbol{r}_2'$ as

$$\begin{bmatrix}\boldsymbol{r}_1'\\ \boldsymbol{r}_2'\end{bmatrix} := \begin{bmatrix}\boldsymbol{r}_1 - \frac{\boldsymbol{r}_2^\top\boldsymbol{r}_1}{2}\cdot\boldsymbol{r}_2\\ \boldsymbol{r}_2 - \frac{\boldsymbol{r}_1^\top\boldsymbol{r}_2}{2}\cdot\frac{\boldsymbol{r}_1'}{\left\|\boldsymbol{r}_1'\right\|_2^2}\end{bmatrix}, \tag{9}$$

with outputs $\boldsymbol{r}_1, \boldsymbol{r}_2$ being orthonormalized $\boldsymbol{r}_1'/\left\|\boldsymbol{r}_1'\right\|_2, \boldsymbol{r}_2'/\left\|\boldsymbol{r}_2'\right\|_2$.

For each image $\hat{\boldsymbol{X}}_i$ in the sequence $\{\hat{\boldsymbol{X}}_i\}_{i=2}^p$, using the previous image $\hat{\boldsymbol{X}}_{i-1}$ as the reference image, we estimate its camera pose $\boldsymbol{h}_i$ by iteratively update the camera pose using (8) with matching feature set $\mathcal{M}_{(i-1,i)}$ followed by the projection operation $\mathcal{P}(\cdot)$ and an evaluation step.

2.3.2. *Camera pose refinement and image fusion*

Motivated by the bundle adjustment principle [14], we propose to further refine camera pose estimations to take advantage of more useful matching feature pairs. With this observation, for the i-th image $\hat{\boldsymbol{X}}_i$, we search feature pairs in all the previous images and form the index set $\mathcal{L}_i = \{l | l < i, \hat{\boldsymbol{X}}_l \bigcap \hat{\boldsymbol{X}}_i \neq \boldsymbol{0}\}$ of images overlapping with $\hat{\boldsymbol{X}}_i$. Using the same condition in (6) for the feature point matching between the target image $\hat{\boldsymbol{X}}_i$ and each image with index in the set $\mathcal{L}_i$, we obtain the union of matching feature sets $\bigcup_{j\in\mathcal{L}_i}\mathcal{A}_{(j,i)}$. Initialized with the estimated camera poses $\{\hat{\boldsymbol{h}}_i\}_{i=2}^p$ from Sec.2.3.1, the proposed method **RRWLM** alternatively updates one pose while keeping other poses fixed, as summarized in Algorithm 1.

Finally, with accurately estimated camera poses for the sequence of images $\{\hat{\boldsymbol{X}}_i\}_{i=1}^p$ (or $\{\hat{\boldsymbol{X}}_i'\}_{i=1}^q$ after deformation), we project all the pixels in these images back to the 3D surface, utilize the linear interpolation to achieve the super-resolution surface texture $\hat{\boldsymbol{U}}_b$ ($\hat{\boldsymbol{U}}_f$) [14], and unfold it to the final 2D image $\hat{\boldsymbol{U}}_b'$ ($\hat{\boldsymbol{U}}_f'$).

Algorithm 1: Refined robust weighted LM (**RRWLM**)

```
Input: Matching feature set $\{\bigcup_{j\in\mathcal{L}_i}\mathcal{A}_{(j,i)}\}_{i=2}^p$, initial
       poses $\{\boldsymbol{h}_i = \hat{\boldsymbol{h}}_i\}_{i=1}^p$, and parameters
       $N, M, \epsilon, L_1, L_2, \epsilon_1, \epsilon_2$;
for $m \leftarrow 0$ to $M$ do
    for $i \leftarrow 2$ to $p$ do
        Construct $\bigcup_{j\in\mathcal{L}_i}\mathcal{M}_{(j,i)} = \{(\boldsymbol{u}_j^m, \boldsymbol{x}_i^m)\}$ from
          $\bigcup_{j\in\mathcal{L}_i}\mathcal{A}_{(j,i)}$ using $\{\boldsymbol{h}_j\}$;
        Initial $\lambda$ and $\boldsymbol{h}_i^{(0)} = \boldsymbol{h}_i$;
        for $t \leftarrow 0$ to $N-1$ do
            $\Delta\boldsymbol{h} \leftarrow (\boldsymbol{H} + \lambda\,\mathrm{diag}(\boldsymbol{H}))^{-1}\boldsymbol{b}$ ;
            $\boldsymbol{h}_o \leftarrow \mathcal{P}(\boldsymbol{h}_i^{(t)} + \Delta\boldsymbol{h})$ ;
            Evaluate $\Delta\boldsymbol{h}$:    $\rho(\boldsymbol{h}_o) =$
              $\frac{g(\boldsymbol{h}_i^{(t)}|\bigcup_{j\in\mathcal{L}_i}\mathcal{M}_{(j,i)}) - g(\boldsymbol{h}_o|\bigcup_{j\in\mathcal{L}_i}\mathcal{M}_{(j,i)})}{\Delta\boldsymbol{h}^\top\cdot[\lambda\,\mathrm{diag}(\boldsymbol{H})\Delta\boldsymbol{h}+\boldsymbol{b}]}$;
            if $\rho(\boldsymbol{h}_o) \geqslant \epsilon$ ($\boldsymbol{h}_o$ is good) then
                $\boldsymbol{h}_i^{(t+1)} \leftarrow \boldsymbol{h}_o$    $\lambda = \max[\lambda/L_1, \epsilon_1]$ ;
            else
                $\boldsymbol{h}_i^{(t+1)} \leftarrow \boldsymbol{h}_i^{(t)}$    $\lambda = \min[\lambda\cdot L_2, \epsilon_2]$
            end
        end
        Update $\boldsymbol{h}_i \leftarrow \boldsymbol{h}_i^N$;
    end
end
Output: $\{\boldsymbol{h}_i\}_{i=1}^p$.
```

2.4. DIC

From previous modules, we obtain the reference $\hat{\boldsymbol{U}}_b'$ and the deformed image $\hat{\boldsymbol{U}}_f'$ of large visual fields of the 3D surface from two sequences of images $\{\boldsymbol{Y}_i\}_{i=1}^p$ and $\{\boldsymbol{Y}_i'\}_{i=1}^q$ of narrow visual fields as inputs, respectively. The basic principle of DIC is tracking of the chosen points between two images recorded before and after deformation for displacement. The sub-level displacement can be computed by tracking pixels in the sparse grid defined on the reference image, thanks to feature tracking methods [31]. Under the assumption that the displacement is small in most engineering applications, our DIC module enables the computation of strain measurement by displacement in different smooth levels based on the programming in [31].

3. NUMERICAL EXPERIMENTS

3.1. Experimental Settings

For the 3D surface under test, two sequences of images are captured, before and after deformation respectively, by a moving camera as illustrated in Fig.1, where the region outside the cylinder is assumed to be black. The 3D cylinder is of radius $r = 500$mm and of height $H = 80$mm. The camera moving trajectory approximately lies in a co-axial virtual cylinder surface of radius $r_2 = 540$mm.

For super-resolution reconstruction of the surface texture, the camera moves in a snake scan pattern, taking 5 images as it moves along the axial direction and then moving forward in the tangential direction for the next 5 images along the axial direction, and so on. We collect a total of $p = q = 160$ images of size $m \times n = 500 \times 600$ for each sequence. Both sequences cover the same area, about 60 degree of the cylinder surface with slightly different camera starting positions before and after deformation, which can be directly extended to the 360° surface.

3.2. Implementation and Evaluation

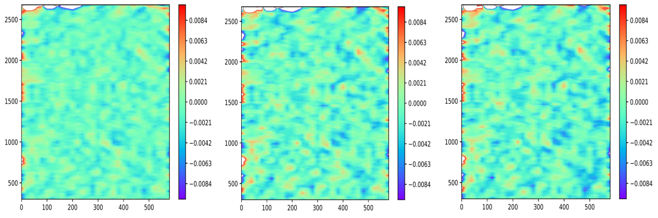
To examine our proposed framework and the essential PnP method for image fusion, we consider 5 baseline methods consisting of a classical iterative method **LHM** [17], four state-of-art non-iterative methods **EPnP + GN** [16], **OPnP + LM** [18], **ASPnP** [20], and **REPPnP** [19] rejecting outliers. For comparison, we denote the non-refined estimation process in Sec 2.3.1 as robust weighted LM (**RWLM**) and the refined robust weighted LM as **RRWLM** in Alg.1. All the baseline methods use the same matching feature set as **RWLM**. Both **LHM** and **RWLM** use their own camera pose estimation of the previous image as initialization for estimation of the present image. **RRWLM** runs with $\{\mathcal{L}_i = \{l < i, |l - i| \leqslant 30\}\}_{i=2}^{p}$, $M = 20$, and other parameters set as [29]. To evaluate the accuracy of the camera pose estimation $\{\hat{\boldsymbol{R}}, \hat{\boldsymbol{T}}\}$, we compute the rotation and translation error together with the ground truth $\{\boldsymbol{R}, \boldsymbol{T}\}$ as $\left\|\left[\hat{\boldsymbol{R}} - \boldsymbol{R}, \hat{\boldsymbol{T}} - \boldsymbol{T}\right]\right\|_2$ and widely used PSNR for the image stitching results $\hat{\boldsymbol{U}}_b'$ and $\hat{\boldsymbol{U}}_f'$.

Firstly, using only the first 10 images of each sequence of images, i.e., $\{\hat{\boldsymbol{X}}_i\}_{i=1}^{10}$ and $\{\hat{\boldsymbol{X}}_i'\}_{i=1}^{10}$ for the reference and deformed texture, we show the average of camera pose estimation errors and the average PSNR of the stitched surface texture images $\hat{\boldsymbol{U}}_b'$ and $\hat{\boldsymbol{U}}_f'$ with comparison to the best 3 baseline methods in Table 1. The strain analysis results by DIC are presented in Fig.2. We observe that the proposed methods have competitive accuracy compared to existing methods when the number of images for fusion is relatively small.

Method	Average pose error		PSNR of $\hat{\boldsymbol{U}}_b'$ and $\hat{\boldsymbol{U}}_f'$	
	$\{\hat{\boldsymbol{X}}_i\}_{i=1}^{160}$	$\{\hat{\boldsymbol{X}}_i\}_{i=1}^{10}$	$\{\hat{\boldsymbol{X}}_i\}_{i=1}^{160}$	$\{\hat{\boldsymbol{X}}_i\}_{i=1}^{10}$
LHM	40.67	0.15	11.86	30.63
EPnP+GN	34.23	0.18	11.91	30.63
OPnP+LM	29.35	0.12	12.53	30.67
RWLM	0.22	0.08	28.09	27.84
RRWLM	0.13	0.07	30.18	30.74

Table 1. The average error of camera pose estimation and the average PSNR (in dB) of the image fusion results $\hat{\boldsymbol{U}}_b'$ and $\hat{\boldsymbol{U}}_f'$, using all 160 images or only the first 10 images in each sequence.

Table 1 also displays the same quantities using all images in the sequence of size $p = q = 160$ instead. Compared to **RWLM**, the proposed method **RRWLM** improves the performance by camera pose refinement, and it also significantly outperforms the baseline methods when stitching a large number of images. The main reason for improvement is that the proposed method **RRWLM** reduces the irreversible camera pose error accumulation in the targeted scenarios.



(a) True strain (b)**RRWLM** (c) **OPnP+ LM**

Fig. 2. Comparison of strain maps of a small field over the cylinder using: (a) ideal surface images, (b) fused images by proposed **RRWLM** method, and (c) fused images of the best baseline result by **OPnP + LM**.

For illustration, the image fusion results for the reference image $\boldsymbol{U}_b'$ via the proposed **RRWLM** are shown in Fig.3 (b) with comparison to the ideal image shown in Fig.3 (a) and the best baseline method **OPnP + LM** in Fig.3 (c). As the image fusion results by existing methods are not applicable anymore for reasonable strain measurement, we only compare the strain measurement result by DIC using **RRWLM** with the ground truth in Fig.4 (we only display the strain in xx direction owing to space limit). It implies that the proposed framework achieves at least sub-pixel and applicable accuracy of image fusion results for strain measurement even if a large number of images are under fusion.

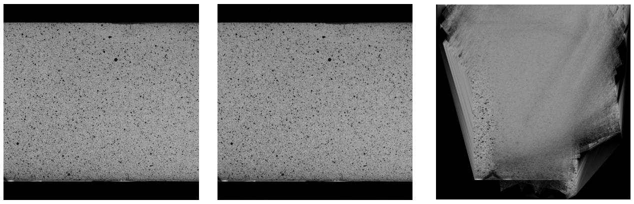
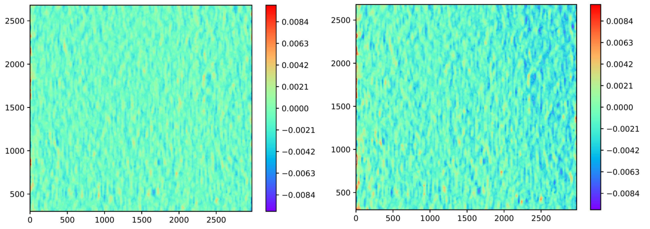

(a) $\boldsymbol{U}_b$ (b) $\hat{\boldsymbol{U}}_b'$ by **RRWLM** (c) $\hat{\boldsymbol{U}}_b'$ by **OPnP + LM**

Fig. 3. Comparison of surface images: (a) ideal image $\boldsymbol{U}_b$, (b) fused image $\hat{\boldsymbol{U}}_b'$ by **RRWLM**, and (c) fused image $\hat{\boldsymbol{U}}_b'$ by **OPnP +LM**.



(a) Strain using $\boldsymbol{U}_b$ and $\boldsymbol{U}_f$ (b) Strain via **RRWLM** fusion

Fig. 4. Comparison of strain maps of a large field over the cylinder using (a) ideal surface images $\boldsymbol{U}_b$ and $\boldsymbol{U}_f$, and (b) fused images $\hat{\boldsymbol{U}}_b'$ and $\hat{\boldsymbol{U}}_f'$ by **RRWLM**.

4. CONCLUSION

We proposed an end-to-end fusion-based DIC framework for 2D strain measurement along curved surfaces of 3D objects in large size. To address the challenges of the single image's narrow visual field of the surface, we incorporate the image fusion principle and decouple the image fusion problem into a sequence of perspective-n-point(PnP) problems. The framework based on our proposed PnP method in conjunction with bundle adjustment accurately recovers the 3D surface texture stitched by a large number of images and achieves applicable strain measurement by DIC. Numerical experiments are conducted to show its outperformance with comparisons to existing methods.

5. REFERENCES

[1] B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review," *Measurement science and technology*, vol. 20, no. 6, p. 062001, 2009.

[2] D. Dawicke and M. Sutton, "Crack-tip-opening angle measurements and crack tunneling under stable tearing in thin sheet 2024-t3 aluminum alloy," 1993.

[3] Z. Tang, J. Liang, Z. Xiao, and C. Guo, "Large deformation measurement scheme for 3d digital image correlation method," *Optics and Lasers in Engineering*, vol. 50, no. 2, pp. 122–130, 2012.

[4] R. C. Gonzalez, R. E. Woods *et al.*, "Digital image processing," 2002.

[5] X. Shao, Z. Chen, X. Dai, and X. He, "Camera array-based digital image correlation for high-resolution strain measurement," *Review of Scientific Instruments*, vol. 89, no. 10, p. 105110, 2018.

[6] B. LeBlanc, C. Niezrecki, P. Avitabile, J. Chen, and J. Sherwood, "Damage detection and full surface characterization of a wind turbine blade using three-dimensional digital image correlation," *Structural Health Monitoring*, vol. 12, no. 5-6, pp. 430–439, 2013.

[7] B. LeBlanc, C. Niezrecki, P. Avitabile, J. Sherwood, and J. Chen, "Surface stitching of a wind turbine blade using digital image correlation," in *Topics in Modal Analysis II, Volume 6*. Springer, 2012, pp. 277–284.

[8] P. Kelly, "Solid mechanics part I: An introduction to solid mechanics," *A Creative Commons Attributions, Mountain View, CA*, vol. 94042, 2013.

[9] Y. Su, Q. Zhang, X. Xu, Z. Gao, and S. Wu, "Interpolation bias for the inverse compositional gauss–newton algorithm in digital image correlation," *Optics and Lasers in Engineering*, vol. 100, pp. 267–278, 2018.

[10] Y. Wang, P. Lava, S. Coppieters, P. Houtte, and D. Debruyne, "Application of a multi-camera stereo dic set-up to assess strain fields in an erichsen test: methodology and validation," *Strain*, vol. 49, no. 2, pp. 190–198, 2013.

[11] X. Chen, L. Yang, N. Xu, X. Xie, B. Sia, and R. Xu, "Cluster approach based multi-camera digital image correlation: Methodology and its application in large area high temperature measurement," *Optics & Laser Technology*, vol. 57, pp. 318–326, 2014.

[12] M. Malesa, M. Kujawińska, K. Malowany, and T. Lusa, "Application of multi-camera dic system for measurements of industrial structures," *Procedia Engineering*, vol. 114, pp. 453–460, 2015.

[13] C.-H. Hwang, W.-C. Wang, and Y.-H. Chen, "Camera calibration and 3d surface reconstruction for multi-camera semicircular dic system," in *International Conference on Optics in Precision Engineering and Nanotechnology (icOPEN2013)*, vol. 8769. International Society for Optics and Photonics, 2013, p. 876913.

[14] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.

[15] R. Sheffer and A. Wiesel, "PnP-Net: A hybrid perspective-n-point network," *arXiv preprint arXiv:2003.04626*, 2020.

[16] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *International journal of computer vision*, vol. 81, no. 2, p. 155, 2009.

[17] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 22, no. 6, pp. 610–622, 2000.

[18] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi, "Revisiting the PnP problem: A fast, general and optimal solution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2344–2351.

[19] L. Ferraz, X. Binefa, and F. Moreno-Noguer, "Very fast solution to the PnP problem with algebraic outlier rejection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 501–508.

[20] Y. Zheng, S. Sugimoto, and M. Okutomi, "ASPnP: An accurate and scalable solution to the perspective-n-point problem," *IEICE Trans. on Information and Systems*, vol. 96, no. 7, pp. 1525–1535, 2013.

[21] S. Urban, J. Leitloff, and S. Hinz, "MLPnP-a real-time maximum likelihood solution to the perspective-n-point problem," *arXiv preprint arXiv:1607.08112*, 2016.

[22] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852–1858, 2011.

[23] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR 2011*. IEEE, 2011, pp. 233–240.

[24] L. Shi and Y. Chi, "Manifold gradient descent solves multichannel sparse blind deconvolution provably and efficiently," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5730–5734.

[25] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[26] M. Tao, J. Yang, and B. He, "Alternating direction algorithms for total variation deconvolution in image reconstruction," *TR0918, Department of Mathematics, Nanjing University*, 2009.

[27] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[28] Y. Ma, D. Liu, H. Mansour, U. S. Kamilov, Y. Taguchi, P. T. Boufounos, and A. Vetro, "Fusion of multi-angular aerial images based on epipolar geometry and matrix completion," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 1197–1201.

[29] H. P. Gavin, "The levenberg-marquardt algorithm for nonlinear least squares curve-fitting problems," *Department of Civil and Environmental Engineering, Duke University http://people. duke. edu/˜ hpgavin/ce281/lm. pdf*, pp. 1–19, 2019.

[30] W. Premerlani and P. Bizard, "Direction cosine matrix imu: Theory," *Diy Drone: Usa*, pp. 13–15, 2009.

[31] D. ANDRE. A python suite for local digital image correlation. [Online]. Available: https://gitlab.com/damien.andre/pydic