3D Res (2017) 8:32
DOI 10.1007/s13319-017-0143-x

3DR EXPRESS

Non-rigid Reconstruction of Casting Process with Temperature Feature

Jinhua Lin · Yanjie Wang · Xin Li · Ying Wang · Lu Wang

Received: 15 June 2017 / Revised: 8 August 2017 / Accepted: 13 August 2017 / Published online: 21 August 2017

Abstract Off-line reconstruction of rigid scene has made a great progress in the past decade. However, the on-line reconstruction of non-rigid scene is still a very challenging task. The casting process is a non-rigid reconstruction problem, it is a high-dynamic molding process lacking of geometric features. In order to reconstruct the casting process robustly, an on-line fusion strategy is proposed for dynamic reconstruction of casting process. Firstly, the geometric and flowing feature of casting are parameterized in manner of TSDF (truncated signed distance field) which is a volumetric block, parameterized casting guarantees real-time tracking and optimal deformation of casting process. Secondly, data structure of the volume grid is extended to have temperature value, the temperature interpolation function is build to generate the temperature of each voxel. This data structure allows for dynamic tracking of temperature of casting during deformation stages. Then, the sparse RGB features is extracted from casting scene to search correspondence between geometric representation and depth constraint. The extracted color data guarantees robust tracking of flowing motion of casting. Finally, the optimal deformation of the target space is transformed into a nonlinear regular variational optimization problem. This optimization step achieves smooth and optimal deformation of casting process. The experimental results show that the proposed method can reconstruct the casting process robustly and reduce drift in the process of non-rigid reconstruction of casting.

Keywords Non-rigid · Online reconstruction · TSDF

1 Introduction

In recent years, with the introduction of depth sensors, researchers began to study high-speed algorithm of 3D reconstruction. Newcombe et al. [1] proposed an effective method to reconstruct 3D scene in real-time, but his method is usually interfered by noise. Researchers have improved the underlying data structure and the depth fusion feature of 3D reconstruction algorithm, a variety of reconstruction methods [2–4] are presented to provide high quality reconstruction [5].

These methods are able to reconstruct static scene effectively, but these methods can not reconstruct

J. Lin (✉) · X. Li · Y. Wang · L. Wang
Changchun University of Technology,
Changchun 130012, China
e-mail: ljh3832@163.com

Y. Wang
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130000, China

non-rigid scene. However, the non-rigid reconstruction is widely used in industrial field. At present, there are many geometric-tracking methods, either multi-view camera or single RGB-D camera is used for reconstruction in these methods [6]. These methods need initial template for tracking, then the template must be updated according to geometric details. This object-specific template limits the applicability to static scene, it is often difficult to reconstruct dynamic scene in practice. To this end, researchers present the template-free method [7], but the real-time performance of reconstruction can not be guaranteed.

Recently, the Dynamic Fusion method proposed by Newcombe et al. can reconstruct the implicit surface and deal with the reconstruction and tracking problem of the hard joint model at real time [8]. The algorithm achieves detail reconstruction of non-rigid scene. However it is also limited by reconstruction time. Because the color information is omitted, the tracking strategy can not effectively track the tangential motion. And the depth only strategy is also easy to generate cumulative error. To solve this problem, this paper uses sparse RGB feature to improve the tracking robustness.

Our algorithm captures discrete and subtle shape deformation by dynamically tracking the scene, rather than using rough deformations. ARAP (as-rigid-as-possible surface modeling) [9] is applied to the spatial region of the embedded surface, and it is combined with the automatically generated volumetric block to extract the geometric complexity. Volumetric block is an effective multi-resolution deformation control field, it is build to solve the nonlinear optimization problems. Finally, the robust alignment is achieved by the correspondence of sparse SIFT, reducing the drift, and realizing the stable tracking of the motion of casting process. The main innovations of the non-rigid reconstruction method proposed in this paper are as follows:

1. A dense surface volumetric representation method is presented to parameterize the geometric surface and the motion of casting.
2. The globally sparse SIFT correspondence is build to achieve robust loop-closed-tracking.
3. The optimization strategy of space deformation is designed to solve the real-time tracking problem in the non-rigid alignment process.

2 Related Work

The static reconstruction method based on RGB-D scanning is divided into point-based representation and grid-based representation method [10]. Following the development of depth sensors, the implicit surface are normally used for reconstruction [11]. The implicit surface can effectively regularize the noise of low-resolution data. Based on this surface representation, there has been a way to reconstruct geometric details in real-time. For example, the static reconstruction method proposed by Newcombe et al. in 2011 captures the truncated signed distance field (TSDF) to achieve dense reconstruction of the scene at sensor rate [12]. Rusinkiewicz et al. proposed iterative closest point (ICP) algorithm that can perform model-to-frame tracking in real-time [13]. Steinbruecker et al. proposed a multi-resolution reconstruction algorithm that extends surface representation to larger volumetric reconstruction [14]. In recent years, on-line reconstruction methods are mostly based on sparse feature tracking, motion structure and multi-view combination. They are integrated into a signed distance field to achieve the on-line reconstruction of geometric surface [15–17].

These static reconstruction methods perform well for solid scenes, they still can not be adapted to non-rigid reconstruction. In order to do dynamic reconstruction, several methods are proposed to track non-rigid deformation in real-time. For example, Collet et al. [18] proposed the use of multiple RGB or depth inputs to achieve non-rigid tracking of certain types of objects. Guo et al. [19] also proposed a template-based approach for tracking deformed mesh. Garg et al. [20] proposed a non-rigid self-motion method that captures the deformation geometry of RGB video, whereas the reconstruction results are rough. Thus, non-rigid reconstruction methods compensate the deficiency of traditional static reconstruction method. For example, Li et al. [21] proposed a non-rigid method to track deformation of objects, the objects is scanned by RGB-D camera, but an initial step is needed to integrate the static templates before reconstruction. To this end, template-free method is proposed for tracking of moving scene in real-time [22]. Most of these tracking methods are based on empirical analysis of captured motion. Therefore, in addition to the time problem, the drift and over-smooth are also a long input sequence problems. DF method uses RGB-D

input to reconstruct non-rigid scene. The method uses sparse deformed graph to achieve the optimization of the non-rigid motion.

The works of Agudo et al. [23–27] and others [28] provide online solutions for the dynamic domain, using exclusively a sequence of RGB images. It is worth noting that the Non-Rigid reconstruction methods proposed by Agudo et al. can run even in real time at frame rate of 30 Hz.

Most of these methods reconstruct non-rigid scene such as people or objects, they can not reconstruct the process of casting. In this paper, TSDF is embedded into the deformation of casting, color features are used for matching correspondence between frames, these features are also used as global anchors to reconstruct casting.

3 System Outline

We use an RGB-D sensor running at 30 Hz frame rate to capture the frame sequence of casting. The color and depth image are captured in resolution of 640×480. The truncated signed distance field (TSDF) is used for parameterizing the geometry and deformation feature of casting [29] (see Sect. 4). The geometric and color information of casting are stored in its undeformed mesh. In order to define the rigid transformation of each voxel, the deformation field is stored at the same resolution as TSDF. In each input frame, the deformation field is continually updated and the new RGB-D image is integrated into an undeformed shape. The reconstruction pipeline of our method is shown in Fig. 1. The polygon mesh of shape P is first generated, which is the isometric surface of the current deformation field. Then, the correspondence between P and depth graph is matched robustly based on the sparse RGB feature and the dense depth data (see Sect. 5). The space deformation is adjusted to match the best correspondence. The update of the deformation field is repeated in iterative closest point (ICP) manner. Next, the depth and color data of each captured frame are integrated into the TSDF. In each reconstruction step, a data parallel iterative strategy is used for solving the high dimensional nonlinear optimization problem of space deformation (see Sect. 6). Finally, the non-rigid reconstruction of casting is done in real-time, the comparison is made between the advanced methods (see Sect. 7).

4 Deformation Field with Temperature Feature

We reconstruct non-rigid scenes by combining surface reconstruction and motion tracking. In this paper, the truncated signed distance function (TSDF) is used for representing the initial undeformed model, and the deformation field is used for tracking motion. First, the model is initialized through regularization of volumetric grid, the points of grid are saved in index table. Each grid point stores six properties, the first three properties are truncated signed distance of $D_i \in \mathbf{R}$, color values $C_i \in [0,\ 255]^3$ and confidence weights $W_i \in \mathbf{R}$. The function values of D at the zero point represent undeformed meshes, which are referred to as canonical gestures. The essence of our deformation is the newly acquired depth data is continuously fused into this canonical frame and the D is updated according to the confidence weight to achieve non-rigid deformation. The last three properties store the information of the current space deformation. For each grid point i, the position of the deformed point is stored in t_i, the current rotation angle is stored in R_i (3 Euler angles), and the temperature value of the point is stored in T_i. The improved data structure with the temperature property is shown in Fig. 2.

In this paper, the temperature interpolation function is proposed to calculate the temperature field of casting, the temperature field refers to the change of the temperature with time or space. The finite element method [30, 31] is used for establishing the temperature function (T) of the voxels of casting. The four vertexes of voxel are numbered according to right hand criteria in clockwise direction, see Fig. 2. The calculation of the temperature field during the casting process is defined as follows:

$$T = N_i T_i + N_j T_j + N_k T_k + N_l T_l \tag{1}$$

in which, N_i, N_j, N_k, N_l is the shape factor, which defined as follows:

$$\begin{cases} N_i = \dfrac{1}{6V}(a_i + b_i x + c_i y + d_i z) \\ N_j = \dfrac{1}{6V}(a_j + b_j x + c_j y + d_j z) \\ N_k = \dfrac{1}{6V}(a_k + b_k x + c_k y + d_k z) \\ N_l = \dfrac{1}{6V}(a_l + b_l x + c_l y + d_l z) \end{cases} \tag{2}$$

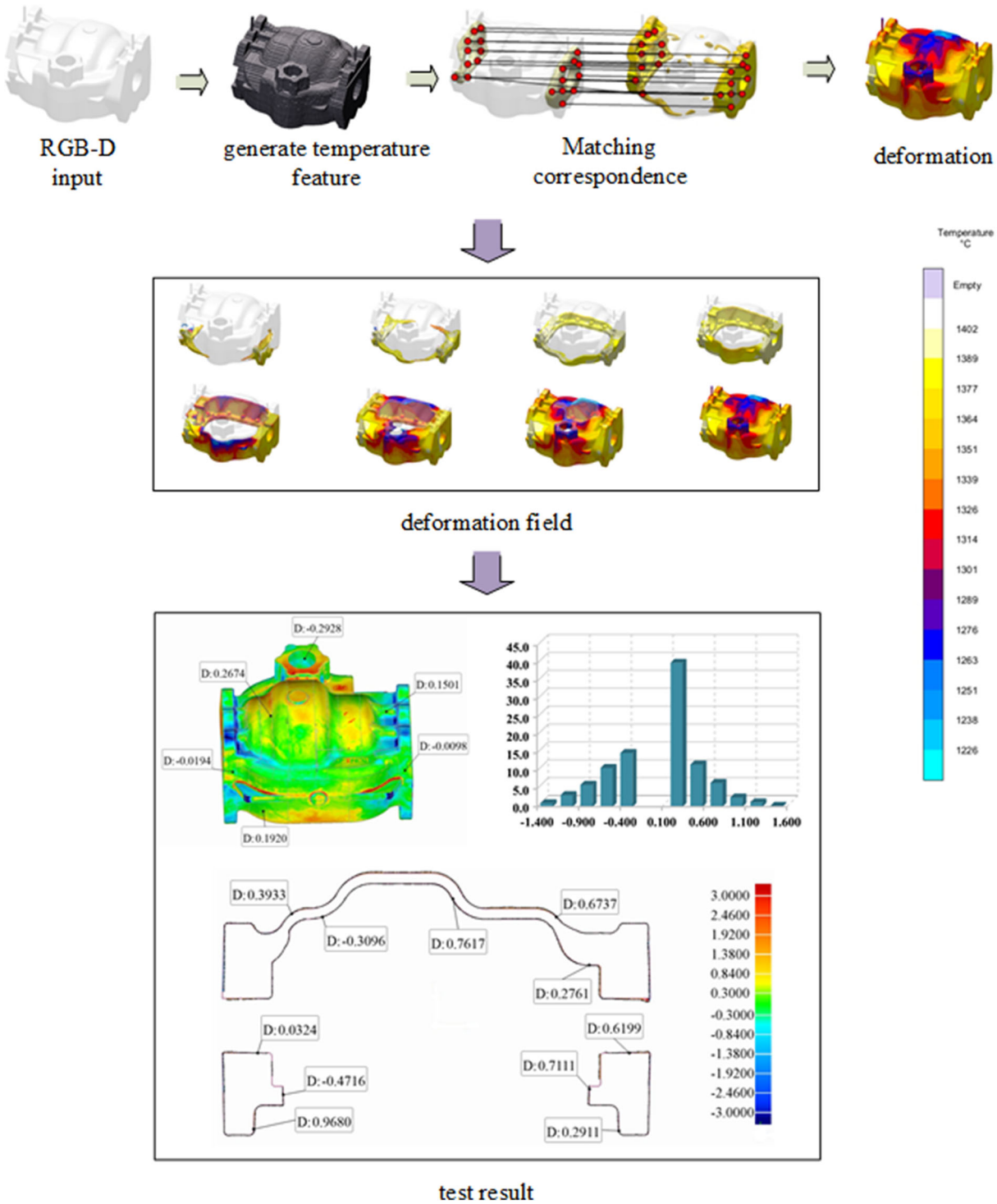


Fig. 1 Reconstruction pipeline of our method. The polygon mesh of casting is first generated, which is the isometric surface of the current deformation field. Then, the correspondence between model and depth graph is matched robustly based on the sparse RGB feature and the dense depth data

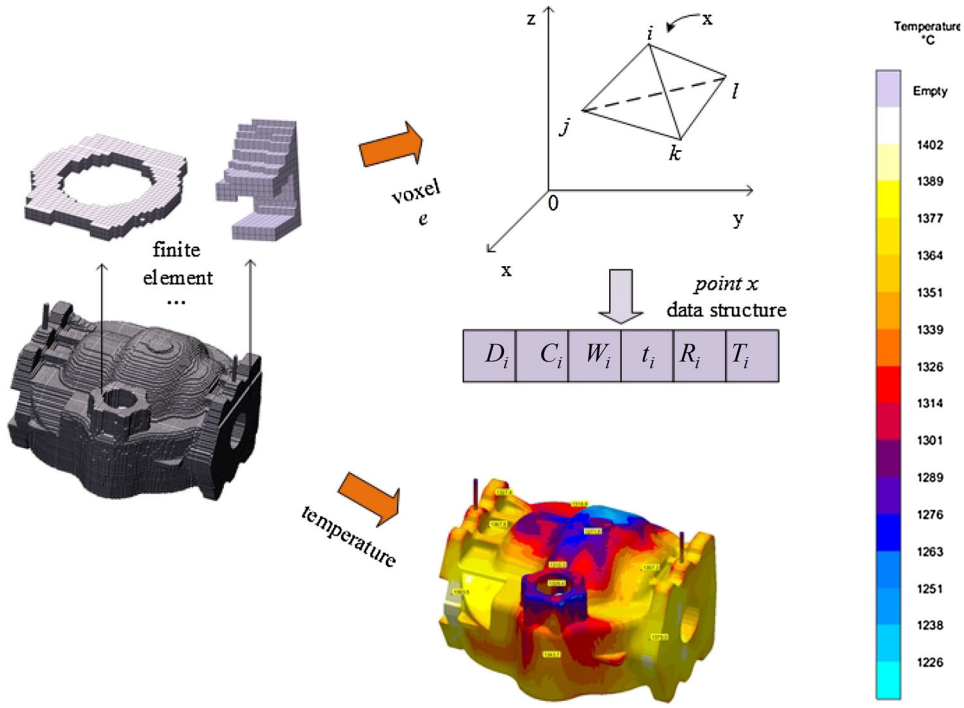


Fig. 2 Space deformation with temperature property. The four nodes of the voxel (e) are defined by the direction of the counterclockwise and right-hand helix, the temperature properties of point are saved as a tuple in data structure

in which, $a_{i-l}, b_{i-l}, c_{i-l}, d_{i-l}$ is constant value, V is the volume of voxels. T_i, T_j, T_k, T_l is the pouring temperature of points.

When the metal liquid is poured into the cavity, pouring temperature (T) refers to the temperature of the metal liquid. This temperature is composed of two parts: liquidus temperature (T_y) and super-heat (T_g). Liquidus temperature is determined by the chemical composition of the metal liquid itself. The chemical compositions of the metal liquid used in this paper are as follows: 3.6% C, 0.1% Cu, 0.036% Mg, 0.15% Mn, 0.007% Mo, 0.029% Ni, 0.016% P, 0.02% S, 2.65% Sn [32]. The liquidus temperature is calculated as follows:

$$T_y = 1538 - \sum_{\substack{i=chemical \\ composition}} a_i w_i \tag{3}$$

in which, $a_i = a_c, a_{cu}, a_{Mg}, a_{Mn}, a_{Mo}, a_{Ni}, a_p, a_S, a_{Sn}$, it is the temperature coefficient of the chemical composition.$w_i = w_c, w_{cu}, w_{Mg}, w_{Mn}, w_{Mo}, w_{Ni}, w_p, w_S, w_{Sn}$, it is the ratio of chemical composition contained in the metal liquid. 1538 is melting point of pure iron. Based on the pouring temperature of each points, the temperature of casting is computed by the formula (1), the computed result is stored as a property of TSDF which is integrated into the deformation field. In this paper, we get the temperature of casting after matching correspondence.

At the top of the deformation field, the rotation angle (R) and displacement (t) are defined to simulate the global motion of the scene. Each grid point is initialized to zero, except for the position of t_i, and t_i is initialized as the displacement of regular grid. Unlike the dynamic fusion method, the regular volumetric mesh g is more close to the shape of the real scene, it is more suitable for dynamic relocation of the spatial position of the non-rigid object. The property values of the grid points are obtained by tri-linear interpolation. The change of the position of the points is achieved by the space deformation function:

$$S(x) = R \cdot (\sum_{i=1}^{length} \alpha_i(x) \cdot t_i) + t \tag{4}$$

in which, *length* is the number of points, $\alpha_i(x)$ is the tri-linear interpolation weight. P is the current deformed surface, $P = s(\hat{P})$.

We get $\hat{P}$ from the volumetric block, the deformation is applied to the grid point. First, a thread for each

grid cell is used for extracting the final set of triangles. The resulting vertex is deformed immediately according to the current deformation field, and the P is obtained. The deformed grid is used for matching correspondence. The algorithm steps are as follows:

input. P //deformed camera attitude map，represented by R and t。
i //latest added point
k //pouring rate
output. S //deformation field with grid points

Step1. $m = length(S)$
if $m = 0$ then
//initialize the deformation field according to the initial camera attitude
$S_m^g = t_0$
$m = m + 1$
$= 0$
$P = P_i$
Step2. for i to $length(P)$ //add a new node to the deformation field
if $\|t_i - t_{last}\|_2 > k$ then
$S_m^g = t_i$
$m = m + 1$
$P_{last} = P_i$
Step3. end

5 Matching Correspondence

The deformation field is updated by matching correspondence between the current shape P and the new frame. The new frame consists of depth data and color data. The dense depth-correspondence is obtained through data-parallel projection (see Sect. 5.1); The sparse color-correspondence is obtained through RGB data (see Sect. 5.2). Since the sparse color descriptors do not change with space deformation, it is used as a global anchor for dynamic and robust tracking of deformation targets. The matching process with temperature feature is shown in Fig. 3.

5.1 Dense Depth Correspondence

Depth correspondence is established by fast projection association steps. Different from the traditional method, we extract a mesh-based isosurface (S) and rasterize it, then, the sample points (S_c) of the current isosurface is obtained. Each S_c is projected into the current depth map (D_t), the sample points are read at the projection position to generate a correspondence between S_c and S_c^a. In order to measure the similarity, we calculate their world spatial distance ($\|S_c - S_c^a\|_2$) and the normality of the normals ($S_c \circ S_c^a$).

In order to ensure the accuracy of reconstruction, we set three thresholds: distance (ε_d), normal offset (ε_n) and camera angle (ε_v). The correspondence is filtered by setting the confidence of the correspondence to zero ($w_c = 0$). The confidence of the correspondence is computed by formula (5):

$$w_c(x) = \left(1 - \frac{\left\| \left(\sum_{c \in S} w_c((1 - \varepsilon_d/x)S_c^a(x - S_c^g) + (1 - \varepsilon_n/x)S_c^g + (1 - \varepsilon_v/x)S_c^t)\right) - S_c^g \right\|_2}{d}\right)^2 \quad (5)$$

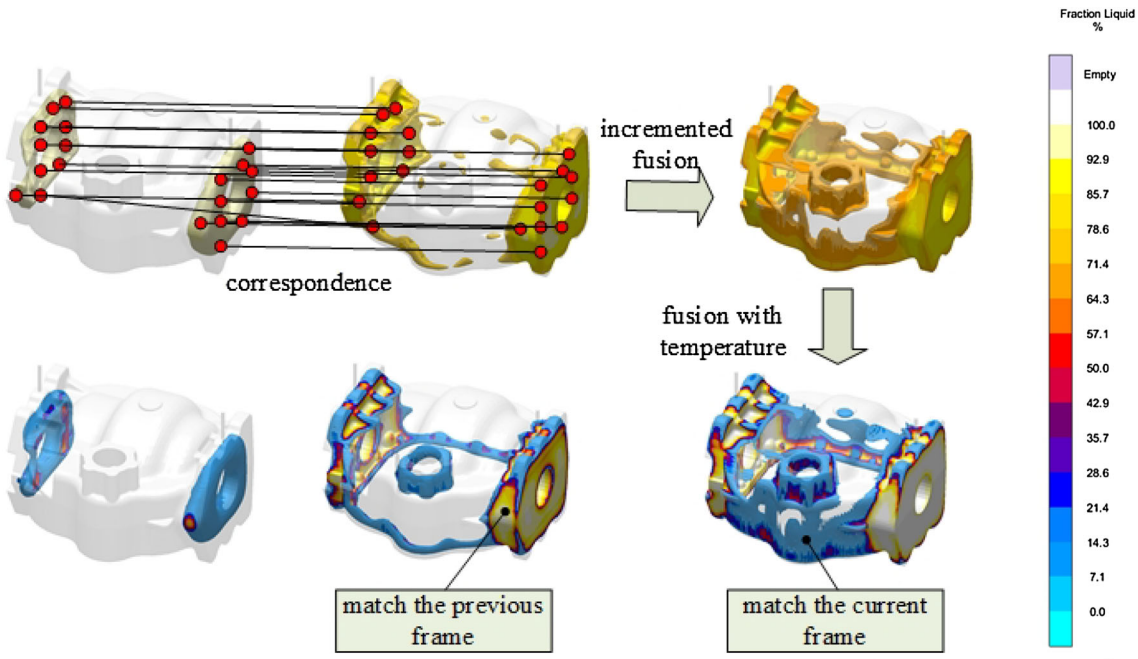


Fig. 3 Correspondence matching process with temperature feature. The deformation field is updated by matching correspondence between the current frame and the next frame. After incremented fusion, the frame with temperature properties is incrementally updated to match the new frame

in which, d is the Euclidean distance to the nearest point x.

5.2 Sparse Color Correspondence

In this paper, dense and sparse correspondence are combined to improve tracking stability and reduce drift. To this end, we use GPU to compute SIFT to match all previous input frames. When a new frame is captured, all feature points are projected to the previous frame through deformation field. Assuming that a rigid transformation is used for the matching between the previous frame and the current frame, the reconstruction pipeline consists of four aspects: key detection, feature extraction, correspondence matching and correspondence filtering. The correspondence matching algorithm is as follows:

input. L //the control points of frame
D_{max} //the maximum change of the depth of point
$[I_{n-1}, M_{n-1}]$ //the depth correspondence of the previous frame
$[I_n, M_n]$ //the depth correspondence of the current frame
R^I //rotation angle of camera
t^I //displacement of camera
输出. C //correspondence list
K_c //global corresponding points
Step1. $i = \alpha \times \gamma \times block + \gamma \times thread$
//initialized points
if $thread = 0$ then
$\kappa = 0$
syncthreads() //synchronize all points within a thread

Step2. for $\iota = 0$ to γ //all pixels within a thread

$p = L_{i+\iota}$

$z = M_n(p)$

if z then

//if the depth data is valid then transform coordinate

$(x', y', z')^T = z(R^I(p,1)^T) + t^I$

$p' = (\frac{x'}{z'}, \frac{y'}{z'})^T$ //project to previous frame

if isInImage(p')

$d = M_{n-1}(p')$,

$\iota_i = (p, p', I_n(p) - I_{n-1}(p'))$, K_c ++

//added correspondence

Step3. syncthreads()//synchronize thread

$b = \gamma \times thread$

$C_i = \iota_a$

Step4. end

In this paper, data parallel strategy is used for searching key position in RGB-D image, the searching result is set as the maximum value of zooming field. It contains 4 octaves, each with 3 levels. Where only the extremes with valid depth are used, then the key points are integrated into three-dimensional space. All key points of the same scale are stored in an array and managed by an atomic counter. The maximum number of key points is 150 per image. Our algorithm associates each key point with up to two gradient directions.

A 128-dimensional SIFT descriptor is calculated for each valid key. Each key consists of its three-dimensional position, scale, orientation, and SIFT descriptors. Key points and descriptors are extracted in resolution of 640 × 480, and the required GPU processing time is 8 ms.

In this paper, all extracted features are stored for matching correspondence in subsequent frames. The feature distance from the current to all previous frames is calculated, and vice versa. The best matching correspondence in both directions is determined by the minimum value of the shared memory. There are 128 correspondence between two frames. The correspondences are stored in order by sorting the feature distance. The algorithm preserves 64 best correspondence between frames.

6 Space Deformation

In order to reconstruct the non-rigid surface in real time, the space deformation is updated at the frame rate, and the robust reconstruction of the casting process is realized. The algorithm uses the dense projection ICP to estimate the global pose parameters of the correspondence. In this paper, objective function is defined on the deformation field to optimize the non-rigid transformation. The deformation process is shown in Fig. 4.

In order to guarantee the robustness of deformation, the optimal parameter is defined as a nonlinear variational optimization problem. The following objective functions are defined:

$$E_{total}(X) = w_r E_{reg}(X) + w_s E_{sparse}(X) + w_d E_{dense}(X) + w_{co} E_{con}(X) \quad (6)$$

in which, E_{reg} represents that the object is regularized by local rigid deformation in the previous frame. E_{sparse} represents the sparse color correspondence, E_{dense} represents the dense depth correspondence. Confidence weights refer to w_r, w_s, w_d and w_{co}, it controls the relative impact of different object, it remain constant during the reconstruction process.

Our method is build upon non-rigid fusion strategy, the depth data of each RGB-D frame is incrementally integrated into the TSDF. Our non-rigid fusion method defines a deformation field through the whole reconstruction. We only integrate to the voxel M (a ring of the current isosurface). It contains at least three optimization step ($K_{\min} = 3$) to ensure that the data is fused into the specific region of deformation, avoiding the overlap of the geometric details. The isosurface is expanded to handle previously invisible geometric features. This extension needs to add new points to the grid, because the new point has not yet been included in the optimization process, the location of these grid points and rotation properties does not match the current deformation. We initialize the position and rotation of each new grid point by extrapolating the current deformation field.

7 Experimental Results

7.1 Qualitative Analysis

The core of our method is to integrate the temperature feature into TSDF and use the sparse RGB feature as a global anchor for robust tracking. Figure 5 shows the

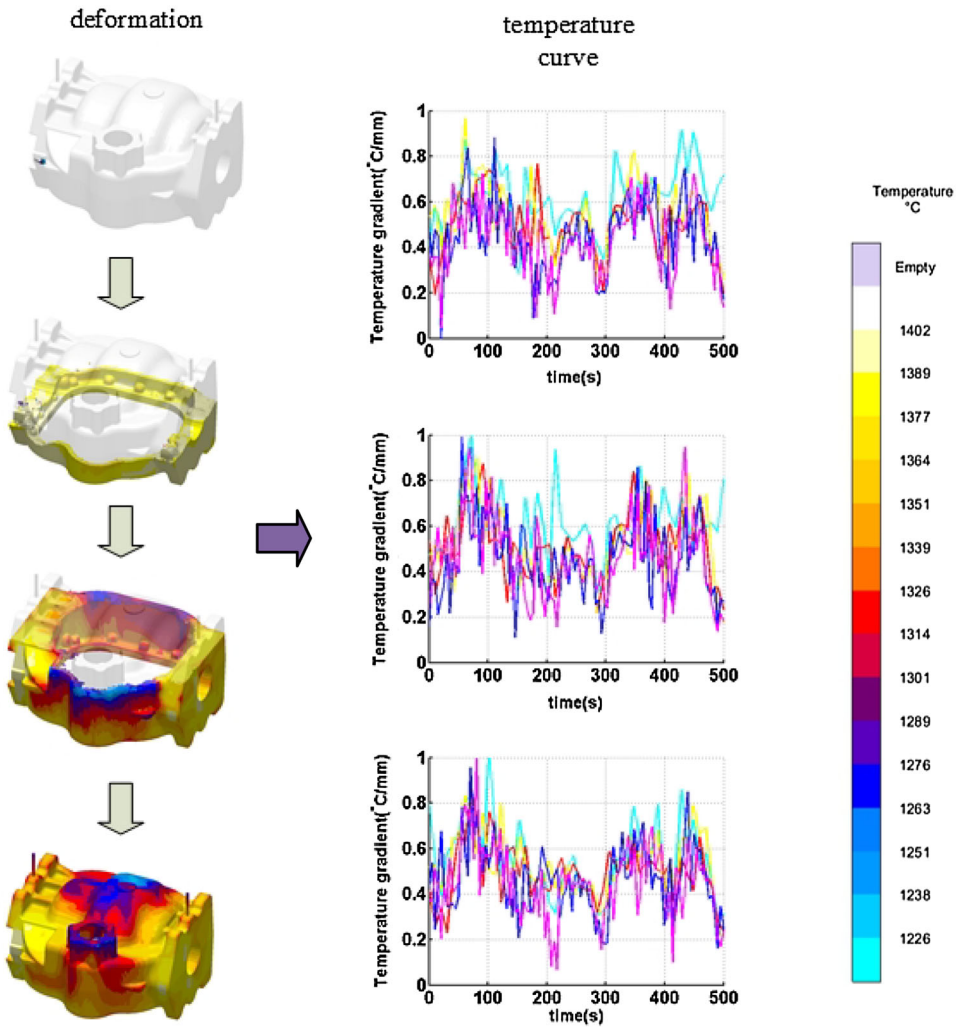


Fig. 4 Casting deformation and temperature gradient. Illumination curves of casting are presented under 3-level deformation. The albedo of the first level is higher than that of the adjacent area, the illumination curve of the first layer changes drastically. Compared with other level, the lighting mutation occurs continuously in first layer of 1-level. While the shading optimization is done level by level, the illumination curves tend to be smooth, the mutation is reduced

reconstruction result of our method. In the absence of geometric feature, dense depth correspondence will cause drift, especially for tangential motion. We successfully track and reconstruct scene by including color feature. For low-resolution deformed grids, the deformation field is not soft enough to handle fine-scale deformations. Using only depth data results in inaccurate tracking and local drift. When the target is violent and the geometry is not obvious, it will cause failure tracking. In this paper, we combine the RGB-D map with the color feature to reduce the drift, the reconstruction result matches the deformation grid with high resolution.

We compare our method with method proposed by Li et al. In Li's method, the pre-scanned template is reconstructed from the static sequence. In our method, we do not need the pre-scanned template, the performance of our method can be quantitatively evaluated by calculating the geometric distance of the template which is fused into the first frame at last. In Fig. 6, the fusion result of our method is shown in yellow one, the blue one is the fusion result of Li's method. Compared with the template-based method, the reconstruction error of our method is less than that of Li's method, our method obtains a higher quality without the need for an initial template.

Our method reconstruct casting in better way compared with other method. The local deviation of our method is less than 1 ms, the RGB feature is extracted as the anchor of correspondence, this anchor reduces the drift, it guarantees the robustness of tracking. In addition, our method reconstruct casting in higher resolution compared with DF method, our method keeps the geometric details of reconstruction in large scale.

In Fig. 7, the tracking stability of our method is shown, and each surface is fused according to the position of the point in the canonical grid. In the case of non-rigid tracking, the surface features remain constant and the surface will change its structure over time in the event of tracking failure or drift. Our method robustly tracks the surface points throughout the sequence, and all points remain stable in their undeformed position without drift. Table 1 presents the noise intensity, the external point ratio, the reconstructed projection error (RPD), and the 3D

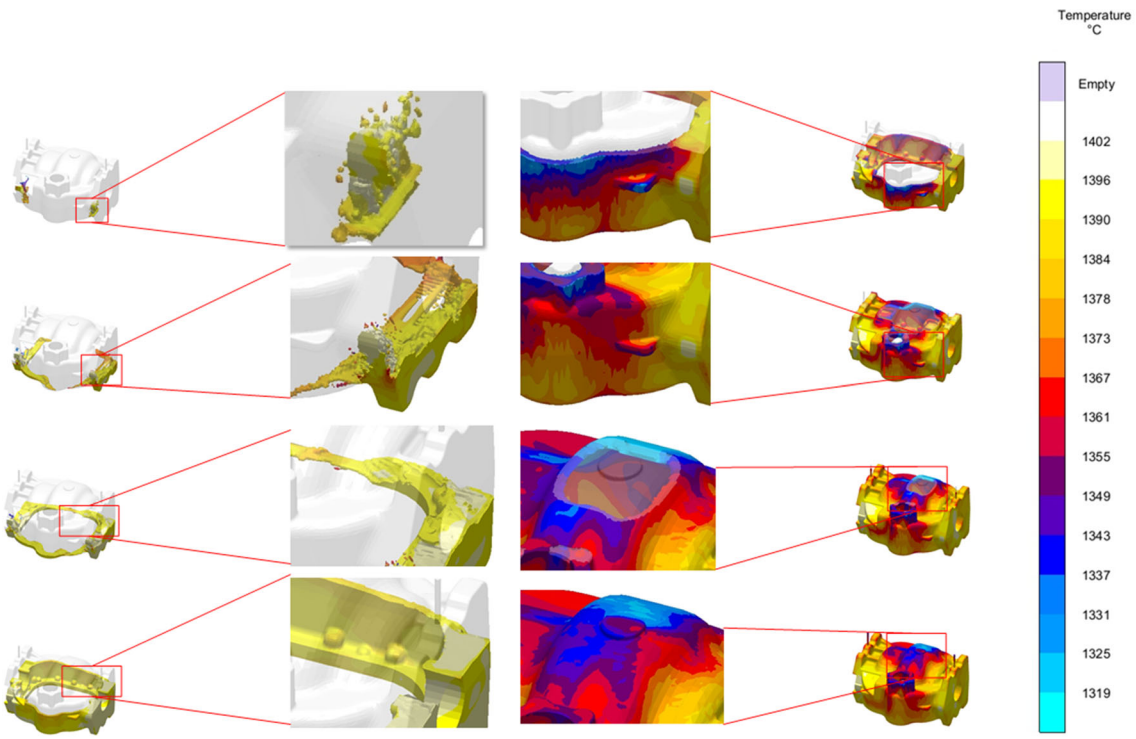
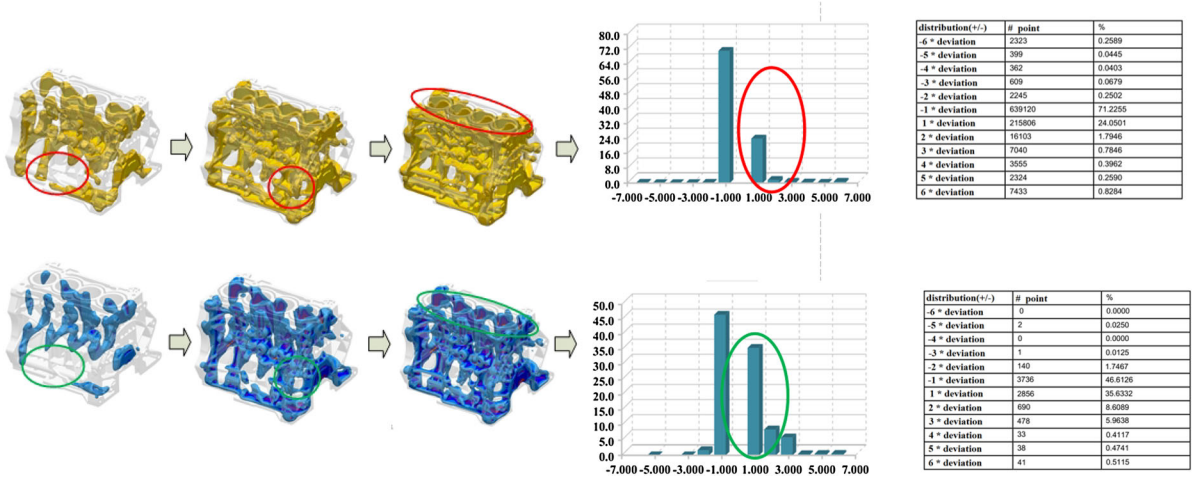


Fig. 5 Non-rigid reconstruction with temperature feature (casting C1). In the absence of geometric feature, dense depth correspondence will cause drift, especially for tangential motion. We successfully track and reconstruct scene by including color feature



distribution(+/-)	# point	%
-6 * deviation	2323	0.2589
-5 * deviation	399	0.0445
-4 * deviation	362	0.0403
-3 * deviation	609	0.0679
-2 * deviation	2245	0.2502
-1 * deviation	639120	71.2255
1 * deviation	215806	24.0501
2 * deviation	16103	1.7946
3 * deviation	7040	0.7846
4 * deviation	3555	0.3962
5 * deviation	2324	0.2590
6 * deviation	7433	0.8284

distribution(+/-)	# point	%
-6 * deviation	0	0.0000
-5 * deviation	2	0.0250
-4 * deviation	0	0.0000
-3 * deviation	1	0.0125
-2 * deviation	140	1.7467
-1 * deviation	3736	46.6126
1 * deviation	2856	35.6332
2 * deviation	690	8.6089
3 * deviation	478	5.9638
4 * deviation	33	0.4117
5 * deviation	38	0.4741
6 * deviation	41	0.5115

Fig. 6 Comparison between our method and Li's method (casting C2). The fusion result of our method is shown in yellow model, the blue one is result from Li's method. Compared with the template-based method, the average error of reconstruction of our method is less than that of Li's method, our algorithm obtains a higher quality without the need for an initial template

point estimation error (3PEE) in the dynamic fusion process. 3PEE refers to the Euclidean distance between 3D point and the estimated value. It can be seen from the table that the average error of reconstruction of C2 casting is between 0.025×10^{-7} and 4.374×10^{-11}. It can be seen from the experimental results that the reconstruction accuracy decreases smoothly with the increase of noise, and no mutation occurs.

7.2 Quantitative Analysis

In this set of experiment, our algorithm is quantitatively tested by synthetic depth and RGB data. A grid is generated from 10 virtual observation points with a given camera track. By radiometric projection, a quantitative depth map is obtained from the real model, Gaussian noise is added to simulate the true depth sensor characteristics. In order to generate color data, we set the surface reflectivity to a uniform value and use a spherical harmonic function to illuminate the model. Our method is compared with the micro-scanning method (WMSM) proposed by Wang et al. [33]. Table 2 shows comparison result of two methods. The results of WMSM include W_a, W_b and W_c. The results of our method are named S_a, S_b and S_c. The geometrical model of casting is cut by virtual plane, and the cross section is projected into horizontal coordinate system. The casting before optimization is then projected onto the horizontal plane using the same method. The two projection curves are fitted to calculate the error distance between the two surface geometries, several key points are extracted from the projection and the error results are displayed at these critical points. The maximum threshold is set to 2, the maximum nominal value is set to 0.2, the minimum threshold is set to -2, and the minimum nominal value is set to -0.2. Table 2 includes the maximum positive deviation (Mpd), the maximum negative deviation (Mnd) and the standard deviation (sd). The excess of the maximum threshold (emc) is calculated as a percentage, and the excess minimum threshold (emt) is also expressed as a percentage. When the tangent plane is projected to the vertical direction, the deviation results are shown as W_a and S_a. The mean standard deviation of W_a is 2.3921 mm and S_a is only 0.6102 mm. Compared with WMSM, the matching accuracy of our method is improved by 18%.

Since WMSM needs to enter the initial mesh, we extract the initial mesh from the implicit surface. Our method takes about 11 s to reconstruct C1 and C2 casting. Because WMSM needs to minimize the objective function, the computational complexity is higher. WMSM reduces the artifacts by assuming a set of albedo clustering, but the texture error is still large. In this paper, the sparse RGB feature is used for optimizing the geometric and albedo changes in the hierarchical structure, the texture distortion is effectively reduced for the geometric details. The

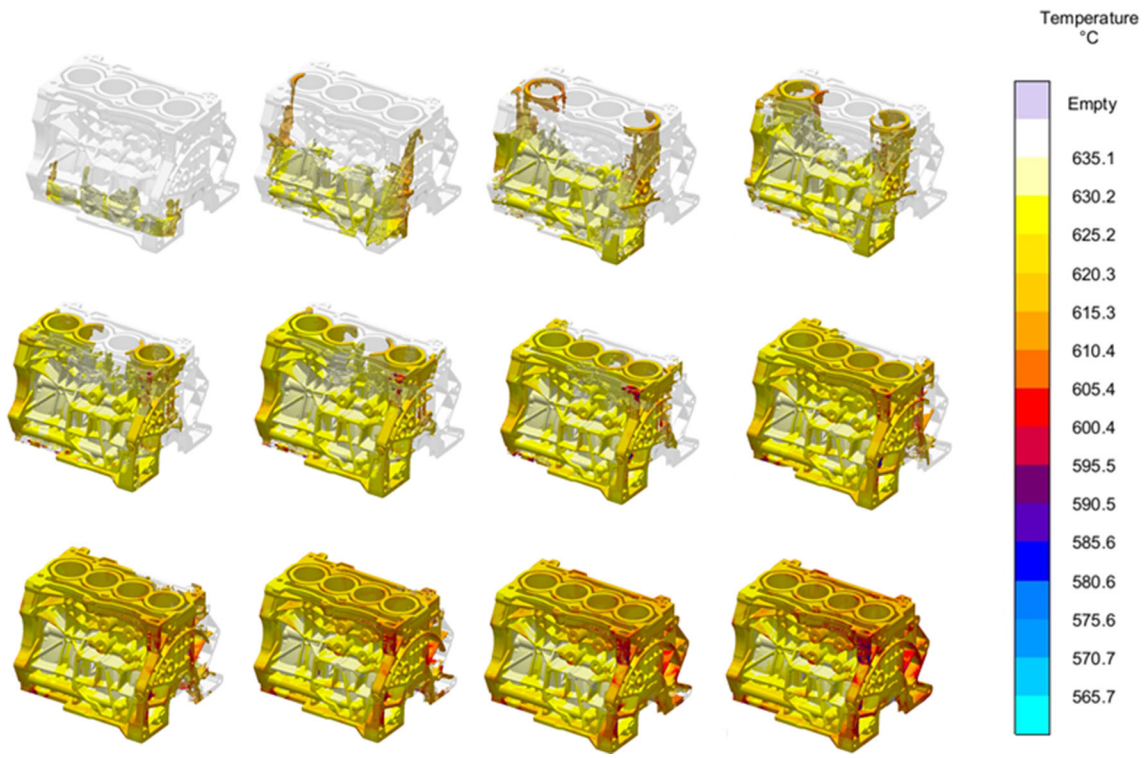


Fig. 7 Dynamic reconstruction of C2 casting. Each frame is fused according to the position of the point in the canonical grid. Our method robustly tracks the surface points throughout the sequence, and all points remain stable in their undeformed position without drift

Table 1 Reconstruction error of C2

σ	%	RPD	3PEE	Time (ms)
2.0	30	0.377×10^{-7}	0.120×10^{-9}	0.12
1.9	20	5.784×10^{-9}	4.374×10^{-11}	0.16
1.8	10	2.287×10^{-8}	2.477×10^{-10}	0.25
1.7	5	5.284×10^{-8}	4.452×10^{-10}	0.31
1.6	2.5	0.936×10^{-7}	0.518×10^{-9}	0.35
1.5	0	0.204×10^{-7}	0.236×10^{-9}	0.36

reconstruction accuracy is higher than that of WMSM, as shown in Fig. 8. Our reconstruction is optimized after the fusion step, and WMSM is optimized before fusion. WMSM's optimization strategy smoothes the local details of the geometry. The total run time of the two methods is basically the same, however, our method can achieve better reconstruction quality.

For different scans of the same casting, the reconstruction performance is evaluated by testing whether the two points match in correspondence. Figure 9 shows accuracy and performance of several methods [15–17] on matching confidence threshold.

Table 2 Comparison of our method and WMSM for two-dimensional deviation (unit: mm)

	Point	Mpd	Mnd	sd	emc (%)	emt (%)
W_a	6280	4.9732	−3.8561	1.0404	10.5573	0.0478
W_b	8015	3.0446	−1.9560	0.3948	0.4991	0.1540
W_c	3327	6.4600	−6.1895	0.9569	5.4707	0.4809
S_a	6302	3.9822	−2.0966	0.6500	9.2250	0.0354
S_b	8083	2.5055	−1.6550	0.3060	0.3547	0.0866
S_c	3274	5.3866	−5.4020	0.8746	4.6506	0.2636

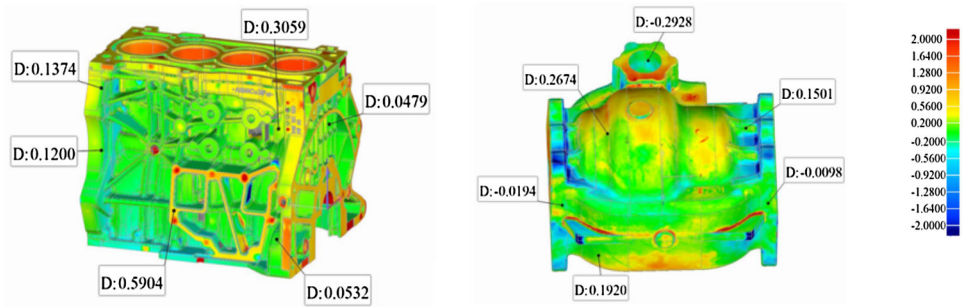


Fig. 8 Comparison of our method and WMSM for three-dimensional deviation. The sparse RGB feature is used for optimizing the geometric and albedo changes in the hierarchical structure, the texture distortion is effectively reduced for the geometric details

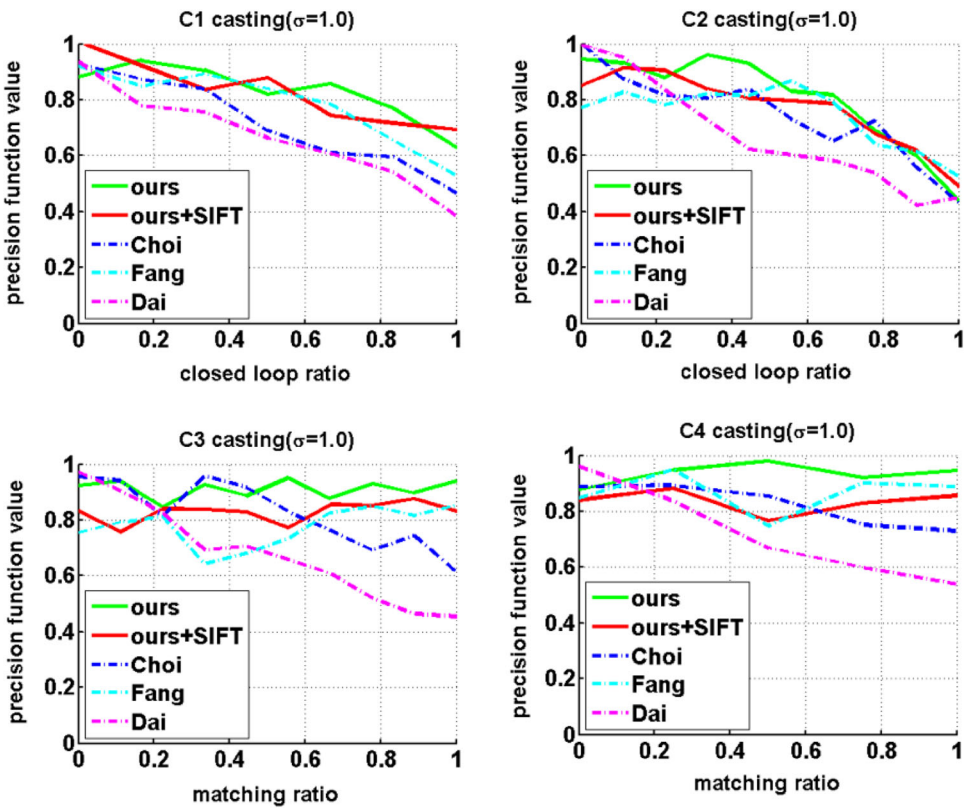


Fig. 9 Comparison of several methods for matching accuracy. This graph shows the performance of several state-of-art geometric methods on matching confidence threshold

Our method performs well by retaining 92% accuracy at 95% closed loop. The matching threshold for each method is calculated by matching the optimal ratio of volumes, while the ratio is changed when correction is involved continuously during TSDF optimization step, and reconstruction accuracy keeps relative to that ratio. The experimental results show that manually produced correspondences are poor in performance when noise and data loss are involved in initial RGB-D input. Our method can maintain a more robust reconstruction in this condition. Our method owns a higher precision compared with other manual geometric matching methods.

In this paper, Asus Xtion Pro sensor is used to capture scenes of automobile castings under natural lighting conditions. Testing scenes are named as C1,

C2, C3 and C4, as shown in Fig. 10. Depth sensor operates in two modes: the first is to capture 640 × 480 resolution depth data and RGB images at 30fps frame rate, the second is to capture 640 × 480 depth data and 1024 × 1024 RGB images at 12fps frame rate. The sensor uses a structured lighting template to obtain independent depth values only at points with visible projection IR. The effective depth resolution is lower than flow resolution. For these test scenes, weight parameters are set as follow: $w_g = 0.2$, $w_r = 20$, $w_s = 10 \sim 100$, $w_a = 0.1$. The weight value increases from 10 to 100 during reconstruction. For a target scene with a uniform albedo, $w_a \to \infty$ is used to indicate that albedo is a constant parameter.

First, depth data with noise is fused into initial geometries represented by TSDF. Noises are regularized by fusion, but it will cause surfaces to be too smooth, resulting in the lack of high-frequency details of geometric reconstruction and drift. Figure 10 shows deformation of castings before and after, red points refer to matching keys before optimization, green casting refers to fusion result after optimization from three different stereo. The geometric details of automobile casting can be robustly tracked in real time. Table 3 shows fusion time, optimization time, and GPU memory exhausted by our method. Following parameters are used for estimating running-time: (1) There are three levels, up to 9 outer Gauss–Newton iterations and 10 PCG iterations; (2) We measure fusion input data (F) and optimize TSDF (O) at 3 levels; (3) It includes GPU memories occupied by variables (#V), hierarchy number of Gauss–Newton iteration (#I), and total time (time). Our method corrects implicit functions with 31 M dynamic variables in seconds. For C1 casting, highest resolution is achieved by 130 M variable on 3-level hierarchy, which takes about 3 min to calculate. Reconstruction result of C1 shows that optimized voxels own resolution of 1 mm and other scenes have a resolution of 1.5 mm. The optimized TSDF achieves high-resolution reconstruction of automobile castings.

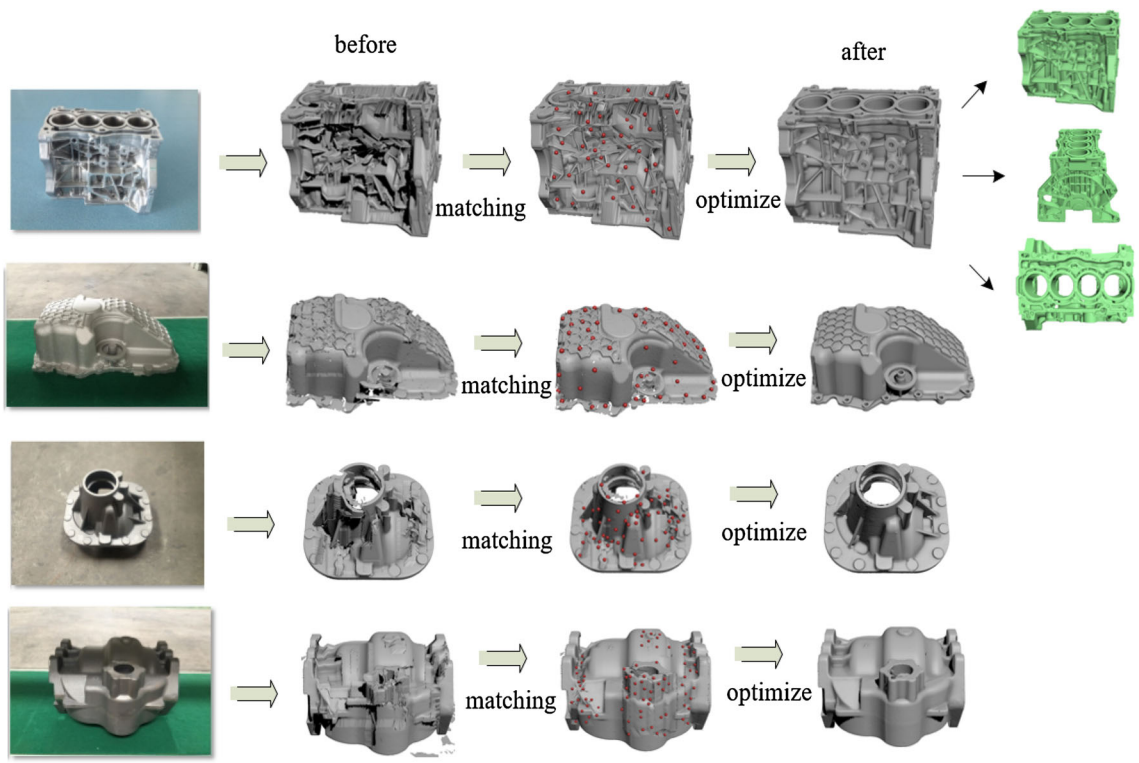


Fig. 10 Before and after result of our method for automobile castings. *Red* points refer to matching keys before optimization. *Green* casting refers to fusion result after optimization from three different stereo

Table 3 Reconstruction time of automobile casting (#V: MB, Time: s)

Level	L1			L2			L3			Total	
Casting	F	O	#V	F	O	#V	F	O	#V	#I	Time
C1	4.0	1.1	31	3.4	0.5	14	1.7	0.3	4	12	11
C2	2.6	1.0	20	2.0	0.5	10	0.9	0.3	3	11	7.3
C3	1.9	0.8	18	1.6	0.4	9	0.9	0.2	2	9	5.8
C4	1.5	0.7	16	1.2	0.3	7	0.7	0.1	1	9	4.5

8 Conclusion

In this paper, an on-line volumetric reconstruction method based on optimization of non-rigid deformation field is proposed to achieve real-time tracking of geometric shapes and motion characteristics of non-rigid deformation scenes. An uniform volumetric representation is presented to achieve parametric representation of geometric and kinematic characteristics. In our method, GPU fast parallel optimization strategy is combined with sparse color and dense depth constraints to achieve motion tracking. The experimental results show that the whole casting process is reconstructed in fast frame rate, the temperature field is generated dynamically to predict the defects of casting robustly.

Acknowledgements This work was supported by National High-tech R&D Program (Grant numbers 2014AA7031010B); Science and Technology Project of The thirteenth Five-Year Plan (Grant number 2016345).

References

1. Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., & et al. (2011) Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of IEEE international symposium on mixed and augmented reality*, IEEE. New York, USA, pp. 127–136.
2. Ming, Z., Fukai, Z., Jiaxiang, Z., & Xin-Guo, L. (2013). Octree-based fusion for realtime 3D reconstruction. *Graphical Models, 75*(3), 126–136.
3. Chen, J.-W., Bautembach, D., & Izadi, S. (2013). Scalable real-time volumetric surface reconstruction. *ACM Transactions on Graphics, 32*(4), 1–16.
4. Huibin, D. U., Yiwen, Z. H. A. O., Jianda, H. A. N., Xingang, Z. H. A. O., Zheng, W. A. N. G., & Guoli, S. O. N. G. (2016). Data fusion of human skeleton joint tracking using two kinect sensors and extended set membership filter. *Acta Automatica Sinica, 42*(12), 1886–1898.
5. Steinbruecker, F., Sturm, J., & Cremers, D. (2014). Volumetric 3D mapping in real-time on a CPU. In *Proceedings of IEEE international conference on robotics and automation*, IEEE. Hong Kong, China, pp. 2021–2028.
6. Zollhöfer, M., Nießner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M., et al. (2014). Real-time non-rigid reconstruction using an RGB-D camera. *ACM Transactions on Graphics, 33*(4), 1–12.
7. Dou, M., Taylor, J., Fuchs, H., Fitzgibbon, A., & Izadi, S. (2015). 3D scanning deformable objects with a single rgbd sensor. In *Proceedings of computer vision and pattern recognition*, IEEE. New York, USA, pp. 493–501.
8. Newcombe, R. A., Fox, D., & Seitz, S. M. (2015). Dynamicfusion: reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of computer vision and pattern recognition*, IEEE. New York, USA, pp. 343–352.
9. Sorkine, O., & Alexa, M. (2007). As-rigid-as-possible surface modeling. In *Proceedings of eurographics symposium on geometry processing*. Barcelona, Spain: Eurographics Association, pp. 109–116.
10. Keller, M., Lefloch, D., Lambers, M., Izadi, S., Weyrich, T., & Kolb A. (2013). Real-time 3D reconstruction in dynamic scenes using point-based rusion. In *Proceedings of 3D vision-3dv*, IEEE. Seattle, WA, USA, pp. 1–8.
11. Zhou, Q.-Y., & Koltun, V. (2013). Dense scene reconstruction with points of interest. *ACM Transactions on Graphics, 32*(4), 112.
12. Curless, B., & Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of computer graphics and interactive techniques*, ACM. New York, USA, pp. 303–312.
13. Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. In *Proceedings of 3DIM*, pp. 145–152.
14. Steinbrücker, F., Kerl, C., & Cremers, D. (2013). Large-scale multi-resolution surface reconstruction from RGB-D sequences. In *Proceedings of IEEE international conference on computer vision*, IEEE. Sydney, Australia, pp. 3264–3271.
15. Choi, S., Zhou, Q. Y., & Koltun, V. (2015). Robust reconstruction of indoor scenes. In *Proceedings of IEEE conference on computer vision and pattern recognition*, pp. 5556–5565.
16. Fang, Y., Xie, J., Dai, G., & et al. (2015). 3D deep shape descriptor. In *Proceedings of IEEE conference on computer vision and pattern recognition*, pp. 2319–2328.
17. Dai, A., Nießner, M., Zollhöfer, M., & et al. (2016). Bundlefusion: Real-time globally consistent 3D reconstruction using on-the-fly surface re-integration. arXiv preprint.
18. Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., et al. (2015). High-quality streamable free-viewpoint video. *ACM Transactions on Graphics, 34*(4), 1–13.

19. Guo, K.-W., Xu, F., Wang, Y.-G., Liu, Y.-B., & Dai, Q.-H. (2015). Robust non-rigid motion tracking and surface reconstruction using L0 regularization. In *Proceeding of IEEE international conference on computer vision*, IEEE. Santiago, Chile, pp. 3083–3091.
20. Garg, R., Roussos, A., & Agapito, L. (2013). Dense variational reconstruction of non-rigid surfaces from monocular video. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE. New York, USA, pp. 1272–1279.
21. Li, H., Vouga, E., Gudym, A., Barron, T. J., Luo, L.-J., & Gusev, G. (2013). 3D self-portraits. *ACM Transactions on Graphics, 32*(6), 187.
22. Wang, R.-Z., Wei, L.-Y., Vouga, E., Huang, Q.-X., Ceylan, D., Medioni, G., & et al. (2016). Capturing dynamic textured surfaces of moving targets. In *Proceedings of the European conference on computer vision*. doi:10.1007/978-3-319-46478-7_17.
23. Agudo, A., Moreno-Noguer, F., Calvo, B., & Montiel, J. M. M. (2016). Sequential non-rigid structure from motion using physical priors. *PAMI, 38*(5), 979–994.
24. Agudo, A., Moreno-Noguer, F., Calvo, B., & Montiel, J. M. M. (2016). Real-time 3D reconstruction of non-rigid shapes with a single moving camera. *CVIU, 153*(12), 37–54.
25. Agudo, A., Montiel, J. M. M., Agapito, L., & Calvo, B. (2017). Modal space: A physics-based model for sequential estimation of time-varying shape from monocular video. *JMIV, 57*(1), 75–98.
26. Agudo, A., Montiel, J. M. M., Calvo, B., & Moreno-Noguer, F. (2016). Mode-shape interpretation: Re-thinking modal space for recovering deformable shapes, WACV, Lake Placid (New York, USA).
27. Agudo, A., & Moreno-Noguer, F. (2017). Combining local-physical and global-statistical models for sequential deformable shape from motion. *IJCV, 122*(2), 371–387.
28. Zhu, S., Zhang, L., & Smith, B. M. (2010). Model evolution: An incremental approach to non-rigid structure from motion. In *CVPR*.
29. Zhou, Q. Y., & Koltun, V. (2014). Color map optimization for 3d reconstruction with consumer depth cameras. *ACM Transactions on Graphics, 33*(4), 1–10.
30. Zhou, B., Peng, S., & Liu, X. (2017). Two-layer motion semantic recognition by fusing the restricted Boltzmann machine based generative model and discriminative model. *Journal of Computer-Aided Design & Computer Graphics, 29*(4), 689–698.
31. Chen, L., Fan, X., Wang, M., & Xu, G. (2017). Parametric reconstruction of four-sided finite element mesh model. *CVPR*, *29*(4), 680–688.
32. Zheng, J., Wang, Q., Zhao, P., et al. (2009). Optimization of high-pressure die-casting process parameters using artificial neural network. *The International Journal of Advanced Manufacturing Technology, 44*(7), 667–674.
33. Wang, X., Ouyang, J., Zhang, G., & Yang, H. E. (2017). Super-resolution reconstruction of infrared images based on micro-scanner. *Journal of Jilin University, 47*(1), 235–241.