Optics and Lasers in Engineering 110 (2018) 7–17

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

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

A flexible heterogeneous real-time digital image correlation system

Tianyi Wang, Qian Kemao*, Hock Soon Seah, Feng Lin

School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Keywords:
Digital image correlation
Real-time processing
Parallel computing
Pipelined system
GPU
Multicore CPU

ABSTRACT

An accurate and flexible real-time digital image correlation (RT-DIC) system utilizing a pipelined CPU and GPU parallel computing framework is proposed. First, the respective advantages of CPU and GPU in performing the fast Fourier transform-based cross-correlation (FFT-CC) algorithm and the inverse-compositional Gauss Newton (IC-GN) algorithm of the employed path-independent DIC (PI-DIC) method are elucidated. Second, based on the different properties and performances of CPU and GPU, a pipelined system framework unifying five Variants of combinations of CPU and GPU is proposed, which can be flexibly applied to various practical applications with different requirements of measurement scales and speeds. Last, both the accuracy and speed of the entire pipelined framework are verified by a PC implementation of the RT-DIC system integrating Variants 2–5. Variants 2 and 5 are also implemented on an iPhone 5S for the feasibility investigation of realizing a portable RT-DIC system on mobile devices using the same framework.

1. Introduction

Digital image correlation (DIC) [1–6] is a very important optical measurement technique for surface deformation of an object/material/structure. DIC estimates deformation between a reference (or undeformed) image and a target (or deformed) image, assuming that the deformation is small. When this assumption is violated, i.e. the deformation becomes large, decorrelation occurs, and the estimated result may be unreliable. However, a high-performance DIC system should have the capability to solve this problem. In the literature, two categories of methods of mitigating large deformation have been attempted, which are the direct methods [7–10] and the incremental methods [11–13].

The direct methods work on one reference image and one target image with a large deformation between them. To converge to satisfactory measurement accuracy, the direct methods always require highly reliable initial guesses. The feature matching method with scale-invariant feature transform (SIFT) features originated from computer vision was first applied to the initialization stage [7–9]. As long as more than three corresponding points are matched between the reference and the target subsets, the initial guesses can be effectively and reliably calculated. Another recent method transformed the rotation in Cartesian coordinates to the translation in polar coordinate, after which the initial guesses were calculated based on the gradient orientation error between the gradient orientation at the seed points and those at the search points [10]. Theoretically, this method can be applied to arbitrary rotation angles. These direct methods, however, are only applicable to large rigid rotational deformation but not effective in analyzing significant tension or compression, in which case SIFT may fail to generate a set of corresponding feature points, while the deformation cannot be effectively characterized in polar coordinates.

The incremental methods resolve the large deformation problem by inserting intermediate images between the reference image and the target image to make the deformation between each two consecutive images small enough to use any existing DIC methods. DIC is thus preferred to be performed dynamically on these images and the final deformation is calculated by accumulating the per-frame results. To avoid decorrelation, the reference image is updated if the deformation is getting too large [11] or the correlation coefficients are less than a certain threshold [12,13]. The incremental methods not only solve the large deformation problem but also provide a means of deformation evolution of the dynamic process, and thus are more generally applicable and useful than the direct methods. The open-source DIC application Ncorr [13] has already employed an incremental method to resolve the large deformation. The similar idea has also been applied to dynamic measurement [14], such as the fracture [15], the strain under dynamic loading [16,17], and the velocity field [18], by incorporating even more intermediate frames. Nonetheless, due to the high computation burden of conventional DIC methods [4], the growth of the number of images during this dynamic process makes DIC even more time-consuming. Therefore, to use the incremental methods in practical DIC applications, the computation efficiency of DIC methods should be improved.

In terms of increasing computation efficiency, the inverse compositional Gauss-Newton (IC-GN) algorithm [19,20], as an alternative to the successful forward additive Newton-Raphson (FA-NR) algorithm,

* Corresponding author.
E-mail address: mkmqian@ntu.edu.sg (Q. Kemao).

https://doi.org/10.1016/j.optlaseng.2018.05.010
Received 28 February 2018; Received in revised form 5 April 2018; Accepted 11 May 2018

was proposed. It eliminated redundant calculations on reference subsets without compromising with accuracy [21,22]. In addition, a reliability-guided initial-guess transferring scheme and an interpolation coefficients loop-up table (LUT) approach [20] were proposed to further reduce the computation cost of IC-GN. Thereafter, IC-GN was applied to develop a video-rate non-contact extensometer recently [23].

Besides the mere algorithmic optimizations of DIC, CPU multithread computing strategies have also been attempted to accelerate DIC to an even higher frame rate [17,23–25], which is crucial for a growing need of on-line monitoring and measurement of dynamically evolving displacements and strains [15,26,27]. Pan and Tian proposed a parallel implementation of their reliability-guided DIC algorithm by starting from multiple seed points of interest (POIs) and achieved a 7 times speedup compared with its sequential implementation [24]. Similarly, Shao et al. designed a real-time 3D DIC system [25,28], which reached a computation speed of more than 50,000 POIs/s when the subset size was set at 15×15. Wu et al. proposed a real-time DIC method [17] employing an efficient integer-pixel estimation method with the combination of an improved particle swarm optimization (PSO) algorithm and a block-based gradient descent search (BBGDS) algorithm. In order to achieve a high frame rate, they then applied OpenMP [29] to parallelize both the integer-pixel search algorithm and the IC-GN algorithm for sub-pixel registration. This system could run at 60 Hz when four POIs were interrogated simultaneously in one frame of images. One of the well-known commercialized DIC systems VIC-2D was also announced to achieve a processing speed up to 1.35×10^5 POIs/s using a PC with a single quad-core CPU [30].

A general-purpose graphical processing unit (GP-GPU), as a more powerful parallel computing device than a multicore CPU, has also been attempted to further accelerate DIC in the past few years. To fully utilize the computation power of GPU, the path-independent DIC schemes have been developed. Jiang et al. proposed a novel path-independent DIC (PI-DIC) algorithm to process each POI independently [31], and accelerated it on GPU (paDIC), reaching a 1.66×10^5 POIs/s computation speed [32]. Afterwards, paDIC was extended and applied to digital volume correlation (paDVC) with a computation speed of 1750 POIs/s [33]. A super-fast, path-independent DIC algorithm called FOLKI-D was proposed by Besnerais et al. for dense 2D and 3D displacement field estimation [34]. GPU was applied to pixelwise operations including solving the 2×2 linear systems and doing the sub-pixel interpolation on GPU's texture memory [35]. Their method achieved a computation speed closed to 2.5×10^6 points/s for a 75% overlap among subsets in the 2D DIC. As the zero-order shape function was employed for matching, the performance of using higher-order shape functions needs further investigation. Very recently, Huang et al. proposed to use heterogenous CPU and GPU parallel computing to accelerate the PI-DIC algorithm which achieved higher speed performance than the GPU-based paDIC [36]. As will be shown in Section 4.2, this method can be simply integrated into the proposed DIC system framework.

Although the computation efficiency of DIC algorithms has been greatly improved, current DIC methods cannot be directly applied to dynamic analysis without investigating the appropriate scheduling scheme of the entire measurement procedure including data acquisition (or data capture), data analysis, and result saving and visualization. To our best knowledge, such investigation has not been conducted in the literature, even though there are commercial DIC systems available in the market [13,30,37]. For convenience, from now on, we refer to correlation analysis as DIC algorithm and the process from data acquisition to result saving and visualization as DIC (software) system. In this study, an accurate and flexible real-time DIC (RT-DIC) system is proposed with the following considerations: (i) since both multicore CPU and GPU have been used to accelerate DIC algorithms, choosing appropriate computing platforms for different measurement requirements is demanded. Thus, a performance comparison between CPU and GPU in implementing FFT-CC and IC-GN in the employed PI-DIC method is first performed; (ii) Guided by the comparison result, it is found that a pipelined heterogeneous system framework that unifies the computing strengths of CPU and GPU is necessary to accelerate the entire DIC system. Based on different measurement scales and different properties and performances of CPU and GPU, five Variants of the pipelined framework are explained and discussed. These Variants can be unified into a single RT-DIC system, which offers it the flexibility to fulfill different requirements in practical DIC applications; (iii) Both the accuracy and speed of the entire framework are verified by a PC implementation of RT-DIC integrating Variants 2-5. Also, as the computation power of mobile devices keeps increasing, the feasibility of porting RT-DIC to mobile devices is investigated. We successfully implement Variants 2 and 5 on CPU and GPU of an iPhone 5S, which opens the door of realizing portable high-performance DIC systems using the proposed pipelined framework.

The rest of the paper is organized as follows. Section 2 briefly reviews the principle of the PI-DIC method used in the proposed system. Section 3 studies the speed performance between multicore CPU-based and GPU-based implementations of PI-DIC. Section 4 explains the design of the five Variants of the proposed RT-DIC framework, whose implementation and validation are given in Section 5 by two applications on a PC and an iPhone 5S. Section 6 concludes the paper.

2. Principle of the PI-DIC method

For the sake of completeness of this paper and the convenience of the description of the following sections, the PI-DIC method is briefly described. More details can be found in [31]. Given a reference subset R centered at a POI in the reference image, DIC searches the subset T in the target image with the highest correlation coefficient, from which the displacement vector of this POI is determined. The PI-DIC method is able to process different POIs independently [31], and the principle of processing one POI-centric subset is schematically shown in Fig. 1. This method includes two parts, the fast Fourier transform-based cross-correlation (FFT-CC) for initial guess of the displacement vector and the IC-GN for its sub-pixel refinement.

For FFT-CC, the zero-order shape function is assumed for a subset. The u- and v- displacements with integer-pixel accuracy can be easily and quickly estimated by Fourier transforms as,

$$C_{ZNCC} = FFT^{-1}\left\{FFT^{*}\left(\left[\bar{R}\right]\right) \cdot FFT\left(\left[\bar{T}\right]\right)\right\}, \tag{1}$$

where $\bar{R}$ and $\bar{T}$ are two matrices containing the zero-normalized values in the reference and the target subsets, respectively; *FFT* and *FFT*$^{-1}$ represent forward and inverse Fourier transforms, respectively; and * indicates complex conjugate. The results are augmented into an integer-pixel deformation parameter vector $\mathbf{p}_0 = [u, 0, 0, v, 0, 0]^T$ as the initial guess for IC-GN. The assumption of zero-order shape function in FFT-CC requires small in-plane strain and small rigid body rotation [19] and may yield unreliable estimations otherwise. This problem can be compensated by capturing intermediate frames between the reference and target states.

In IC-GN, the first-order shape function is adopted in this study although higher order shape functions can also be used. The initial guess $\mathbf{p}_0$ calculated from FFT-CC is passed to IC-GN to obtain the sub-pixel deformation parameter vector $\mathbf{p} = [u, u_x, u_y, v, v_x, v_y]^T$, where u_x, u_y, v_x, and v_y are the gradients of u and v, respectively. IC-GN minimizes the zero-normalized sum of squared difference (ZNSSD) criterion iteratively. In each iteration, the incremental deformation parameter $\Delta\mathbf{p} = [\Delta u, \Delta u_x, \Delta u_y, \Delta v, \Delta v_x, \Delta v_y]^T$ is calculated as,

$$\Delta\mathbf{p} = \mathbf{H}^{-1}\sum_{\xi}\left\{\left[\nabla R(\mathbf{P}+\xi)\frac{\partial\mathbf{W}}{\partial\mathbf{p}}\right]^T\left[\frac{\bar{R}_n}{\bar{T}_n}\bar{T}\left[\mathbf{P}+\mathbf{W}(\xi;\mathbf{p})\right]-\bar{R}(\mathbf{P}+\xi)\right]\right\} \tag{2}$$

where $\mathbf{P}(x_0, y_0)$ is the position of the POI; $\xi = (\Delta x, \Delta y)$ represents the local coordinates of pixels within the reference subset R; $\nabla R(\mathbf{P}+\xi)$ is the gradient within the reference subset; $\mathbf{W}(\xi;\mathbf{p})$ is the warp function,