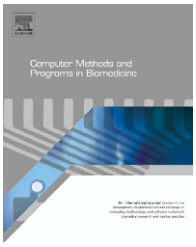
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 108 (2012) 1012–1021

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Parallel perfusion imaging processing using GPGPU

Fan Zhu[a,b,*], *David Rodriguez Gonzalez*[a,b], *Trevor Carpenter*[b], *Malcolm Atkinson*[a], *Joanna Wardlaw*[b]

[a] *Data-Intensive Research Group, School of Informatics, University of Edinburgh, Edinburgh, UK*
[b] *SFC Brain Imaging Research Centre, Division of Clinical Neuroscience, University of Edinburgh, Edinburgh, UK*

ARTICLE INFO

Article history:
Received 10 October 2011
Received in revised form
14 June 2012
Accepted 16 June 2012

Keywords:
Local AIF
Perfusion imaging
Deconvolution
Parallelization
GPGPU

ABSTRACT

Background and purpose: The objective of brain perfusion quantification is to generate parametric maps of relevant hemodynamic quantities such as cerebral blood flow (CBF), cerebral blood volume (CBV) and mean transit time (MTT) that can be used in diagnosis of acute stroke. These calculations involve deconvolution operations that can be very computationally expensive when using local Arterial Input Functions (AIF). As time is vitally important in the case of acute stroke, reducing the analysis time will reduce the number of brain cells damaged and increase the potential for recovery.
Methods: GPUs originated as graphics generation dedicated co-processors, but modern GPUs have evolved to become a more general processor capable of executing scientific computations. It provides a highly parallel computing environment due to its large number of computing cores and constitutes an affordable high performance computing method. In this paper, we will present the implementation of a deconvolution algorithm for brain perfusion quantification on GPGPU (General Purpose Graphics Processor Units) using the CUDA programming model. We present the serial and parallel implementations of such algorithms and the evaluation of the performance gains using GPUs.
Results: Our method has gained a 5.56 and 3.75 speedup for CT and MR images respectively.
Conclusions: It seems that using GPGPU is a desirable approach in perfusion imaging analysis, which does not harm the quality of cerebral hemodynamic maps but delivers results faster than the traditional computation.

1. Introduction

With the development of computed tomography (CT) [1,2] and magnetic resonance (MR) imaging [3,4], perfusion imaging becomes a very powerful clinical tool for evaluation of brain physiology. They can be used to evaluate brain function via assessment of cerebral perfusion parameters.

The main applications of brain perfusion imaging are acute stroke and brain tumors. In the case of acute stroke, the information obtained from brain perfusion imaging can be used to evaluate the appropriateness of administering thrombolytic treatment, which can help to reduce the final volume of dead tissue, but has some risks such as hemorrhages. The results are used to evaluate the possible benefits. In the case of tumors, they are used to distinguish tumor characteristics and follow tumor development, possibly also after treatment to see whether it has been effective.

Evaluating tissue time–concentration curve of a contrast agent intensity after its injection, has become possible on time scales comparable with the mean transit time (MTT). To achieve this, deconvolution is used in perfusion imaging to obtain the Impulse Response Function (IRF) that is then used to create parametric maps of relevant hemodynamic

* *Corresponding author at*: Data-Intensive Research Group, School of Informatics, University of Edinburgh, Edinburgh, UK.
E-mail addresses: F.Zhu@ed.ac.uk (F. Zhu), David.Rodriguez@ed.ac.uk (D.R. Gonzalez), Trevor.Carpenter@ed.ac.uk (T. Carpenter), Malcolm.Atkinson@ed.ac.uk (M.P. Atkinson), Joanna.Wardlaw@ed.ac.uk (J. Wardlaw).
0169-2607
http://dx.doi.org/10.1016/j.cmpb.2012.06.004

quantities such as cerebral blood flow (CBF), cerebral blood volume (CBV) and mean transmit time [5–7]. Cerebral blood flow indicates the volume of blood flowing through a given voxel in a given time. Cerebral blood volume refers to the volume of blood in a given voxel of brain tissue. Mean transit time designates the average time blood takes to flow through a given voxel of brain tissue, it is commonly measured in seconds. Time To Peak (TTP) and Time of Arrival (TA) are two other parameters often be measured [8]. TA refers to the time of arrival of the contrast agent in the voxel after injecting contrast agent. TTP refers to corresponding time of the maximum contrast concentration. In previous studies, singular value decomposition (SVD) and its variants were proved to be applicable to perform deconvolution in perfusion imaging [9]. As the raw data obtained from CT or MR scanners is not noise free and as deconvolution is very sensitive to noise, truncated SVD is used to minimize the noise impact [10–13].

In clinical practice, a global AIF for the entire brain can be determined from voxels near a major artery feeding the brain. However, the global AIF technique is based on the assumption that the contrast agent reaches every voxel of the brain at the same time. In the case of acute stroke, the contrast arrival time can be different and the assumption is not satisfied. As a result, using a global AIF for the entire brain is not very accurate [14,15].

The other solution is to use local AIFs [16–19]. Instead of using a global AIF generated from voxels near the major artery for whole brain, different local AIFs are used for a single scan. Each local AIF is generated by measuring a small set of blood vessels in a specified area near the voxel of interest. Lorenz et al. [16] had shown that localized AIFs are feasible and provide more useful perfusion results.

However, using local AIFs leads to fairly slow performance, in the worst case, the perfusion-imaging analysis takes more than half an hour compared with the running time of global AIFs based methods which is a couple of minutes. According to Saver's experiment in 2006 [20], during 30 min, 57.6 million neurons die. In the same minutes, your brain loses 41.4 billion synapses and 360 km of axonal fibers. Since a stroke is a medical emergency and every second counts, the sooner results are delivered in diagnosis, the less damage will be caused to a patient's brain. Obviously, half an hour is not a reasonable option for clinical diagnosis. Therefore, a parallel implementation of perfusion-imaging analysis which brings performance speedup without quality lose is very promising to help using local AIFs in perfusion imaging.

In this paper, we present a GPGPU-based brain perfusion imaging analysis implementation using the CUDA programming model. We also compared the performance of the serial and parallel perfusion imaging analysis methods.

2. Background and methods

2.1. Perfusion imaging algorithm

Ostergarrd et al. [11,12] and Wurestan et al. [13] have shown that an accurate CBF can be determinated using deconvolution of a tissue time–concentration curve and an AIF. From a CT or MR scanner, we get a series of brain images at different sampling times. For each voxel, we collect data at specific time intervals to build a tissue time–concentration curve of contrast agent intensity, which is also called volume of fluid (VOF) curve. This curve will be referred to as C_t.

A local AIF matrix is created from the local AIF vector as follows:

$$C_a = \Delta t \begin{pmatrix} C_a(t_1) & 0 & \cdots & 0 \\ C_a(t_2) & C_a(t_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_a(t_N) & C_a(t_{N-1}) & \cdots & C_a(t_1) \end{pmatrix} \quad (1)$$

where $(t_1, t_2, \ldots, t_N)$ is the sampling time, $(Ca(t_1), Ca(t_2), \ldots, Ca(t_N))$ is an arterial input function given as an input and Δt is time scale.

In perfusion imaging, the output we want to obtain is Impulse Response Function (IRF), which is referred to as h.

The volume of fluid, C_t, the C_a, and IRF h satisfies the following equation:

$$C_t = C_a \otimes h + \epsilon \quad (2)$$

where $\otimes$ denotes convolution and ϵ is the noise.

Finally, the CBF, CBV and MTT for each voxel are calculated as follows:

$$CBF = Max(h) \quad (3)$$

$$CBV = \int_0^\infty h(t)\,dt \quad (4)$$

$$MTT = \frac{CBF}{CBV} \quad (5)$$

Singular value decomposition (SVD) is one of the most popular techniques to solve deconvolution problems in perfusion imaging. Suppose C_a from Eq. (1) is an m-by-m matrix, there exists a factorization such that:

$$C_a = U \cdot W \cdot V^T \quad (6)$$

where U is an $m \times m$ unitary matrix, W is $m \times n$ diagonal matrix and V^* is the transpose of an $n \times n$ unitary matrix V. A common convention is to order the diagonal matrix W in a decreasing order and these diagonal entries of W are known as the singular values of original matrix C_a.

The C_a^{-1} can then be written as:

$$C_a^{-1} = V \cdot W^{-1} \cdot (U^T) \quad (7)$$

To solve the deconvolution problem in Eq. (2), the solution can be simply delivered after applying SVD:

$$h = V \cdot W^{-1} \cdot (U^T \cdot C_t) \quad (8)$$

Furthermore, as rows in C_a in Eq. (2) are close to linear combinations, the deconvolution is an ill-posed problem, hence, it is very sensitive to noise. Truncated SVD is introduced to minimize the effect of noise. In truncated SVD, a threshold is added