Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language 35 (2016) 1–13
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Parallel SRP-PHAT for GPUs☆

Taewoo Lee [a], Sukmoon Chang [b], Dongsuk Yook [a,*]

[a] *Speech Information Processing Laboratory, Korea University, Seoul 136-701, Republic of Korea*
[b] *Pennsylvania State University, Middletown, PA 17057, USA*

Received 27 December 2014; received in revised form 13 April 2015; accepted 7 May 2015
Available online 15 May 2015

Abstract

The steered response power phase transform (SRP-PHAT) is one of the widely used algorithms for sound source localization. Since it must examine a large number of candidate sound source locations, conventional SRP-PHAT approaches may not be used in real time. To overcome this problem, an effort was made previously to parallelize the SRP-PHAT on graphics processing units (GPUs). However, the full capacities of the GPU were not exploited since on-chip memory usage was not addressed. In this paper, we propose GPU-based parallel algorithms of the SRP-PHAT both in the frequency domain and time domain. The proposed methods optimize the memory access patterns of the SRP-PHAT and efficiently use the on-chip memory. As a result, the proposed methods demonstrate a speedup of 1276 times in the frequency domain and 80 times in the time domain compared to CPU-based algorithms, and 1.5 times in the frequency domain and 6 times in the time domain compared to conventional GPU-based methods.
© 2015 Elsevier Ltd. All rights reserved.

Keywords: Sound source localization; SRP-PHAT; GPUs

1. Introduction

The steered response power with a phase transform filter (SRP-PHAT) is a robust sound source localization (SSL) algorithm that can handle noisy sound signals in reverberant environments (DiBiase et al., 2001). However, since it performs a grid search examining a large number of candidate sound source locations, the SRP-PHAT is computationally expensive and may not be used in real time. Many approaches have been proposed for a fast SRP-PHAT (Zotkin and Duraiswami, 2004; Peterson and Kyriakakis, 2005; Do and Silverman, 2007; Dmochowski et al., 2007; Cho et al., 2009; Yook et al., 2015). For instance, a hierarchical search method (Zotkin and Duraiswami, 2004; Do and Silverman, 2007) gradually prunes the candidate sound source locations in a coarse-to-fine search. A drawback of this method is that it may prematurely prune the sound source with the highest power before it reaches the final decision. Alternatively, a hybrid method (Peterson and Kyriakakis, 2005) first generates a small set of candidate sound source locations using a time difference of arrival (TDOA) based search. It then performs an SRP-PHAT based grid search on this small set of candidate locations. If TDOA estimation is unsuccessful in the first step, then the SRP-PHAT based SSL will fail in the final decision. In Dmochowski et al. (2007), an inverse mapping function relates a relative TDOA to a set of candidate

☆ This paper has been recommended for acceptance by Dr. S. Narayanan.

* Corresponding author. Tel.: +82 02 3290 3202; fax: +82 02 3290 3641.
E-mail addresses: twlee@speech.korea.ac.kr (T. Lee), sukmoon@psu.edu (S. Chang), yook@speech.korea.ac.kr (D. Yook).

http://dx.doi.org/10.1016/j.csl.2015.05.002
0885-2308/© 2015 Elsevier Ltd. All rights reserved.

Table 1
Notations.

Symbol	Meaning
$\|\cdot\|$	Number of elements in (cardinality of) a set or the magnitude of a complex number
C	Number of cores
κ	Subset of frequency bins
M	Number of microphones
n	Microphones m' and m''
N	Number of microphone pairs; $N = M(M - 1)/2$
p_q	Steered response power for location q
Q	Number of candidate sound source locations
$R_n[t]$	Cross correlation of the nth microphone pair at time lag t
ρ	Subset of candidate sound source locations
$\tau[q, m]$	Time of arrival of a sound signal from location q to the mth microphone
$\tau[q, n]$	Time difference of arrival between the nth microphone pair; $\tau[q, n] \equiv \tau[q, m'] - \tau[q, m'']$
T	Number of sound signal samples
$x_m[t]$	Sound signal sample of the mth microphone at time t
$X_m(\omega)$	Fourier transform of the mth microphone signal at frequency ω
$\mathrm{Z}_n[k]$	Cross spectrum of the nth microphone pair at frequency bin k

locations. Only the output powers of the locations that are inversely mapped by the TDOA are considered to find the maximum power location. This method may also fail to find the maximum power location, because it searches only those few locations that are inversely mapped by the relative TDOA. A real-time SRP-PHAT method was proposed for humanoid robots (Cho et al., 2009). It uses a pre-computed look-up table that contains a complete set of regions having unique TDOA values. Although it is guaranteed to find a global maximum power location, this method is suitable for small-scale microphone array systems because there may be too many candidate coordinates in the look-up table for a large-scale microphone array system (Cho et al., 2009). Recently, improvement of this method was proposed by adopting the idea of the hierarchical search method (Yook et al., 2015).

All of the above methods focus on removing repetitive computation from the algorithm or reducing the number of candidate sound source locations that must be searched. However, with the support of single-instruction multiple-data (SIMD) on modern central processing units (CPUs), a parallelized design of the SRP-PHAT on a CPU has also been suggested (Lee and Kalker, 2010). In addition, as general purpose computing on graphics processing units (GPUs) becomes common, parallelized designs of the SRP-PHAT on GPUs have been proposed (Silveira et al., 2010; Lee and Yook, 2011; Minotto et al., 2012). It is known that the bottleneck for most GPU-based parallel processing tasks is the amount of memory access, rather than the computation itself (Williams et al., 2009; Jang et al., 2011). To maximize the efficiency of the SRP-PHAT, therefore, it is imperative to process memory access efficiently so that it can take the best advantage of GPUs. The previous GPU-based parallel SRP-PHAT (Minotto et al., 2012) demonstrated much improvement in their performance compared to the CPU-based implementations. However, it did not exploit the maximum performance from GPUs since on-chip memory usage was weakly considered.

In this paper, we propose a GPU-based parallel algorithm of the SRP-PHAT optimized for the efficient use of memory on GPU devices. The proposed method reduces space complexity by maximizing the use of on-chip memory on the GPU. In addition, it reduces the bank conflicts and utilizes the cache line in its maximum capacity by vectorizing memory access. As a result, the proposed method greatly reduces the overall execution time of the SRP-PHAT-based sound source localization.

The remainder of this paper is organized as follows. In Section 2, we review the sequential SRP-PHAT in the frequency domain (FD) and time domain (TD). In Section 3, we analyze the parallelism of the SRP-PHAT. In Section 4, we present the details of GPU-based parallel algorithms of the SRP-PHAT. Section 5 discusses the experimental results of the proposed methods. Finally, we draw conclusions in Section 6. Some notations used in this paper are summarized in Table 1.

Download English Version:

https://daneshyari.com/en/article/558243

Download Persian Version:

https://daneshyari.com/article/558243

Daneshyari.com