University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Dimitri Degkwitz

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, November 13, 2020

Bachelor's Thesis
Database Technology

Topic: A Convolved Single Point Fourier Transform (SPIFT) for Radio-Astronomy Data using Apache Flink

The Fourier transform is an important technique in signal and image processing. In radio astronomy, the Fourier transform is applied to frequency data of radio signals to compute the sky image. The common practice is to apply the Fourier transform after the observation stage has been completed and all the data has been acquired. The *Single-Point Incremental Fourier Transform* (SPIFT) [1] computes the sky image incrementally during the observation stage and continuously refines the image as new observations arrive.

The radio antennas sense the electric field of radio signals and produce a voltage, termed visibility. A visibility vis_t is a complex value measured at coordinate u, v and time t. We denote the continuously arriving observations with $S(t) = (\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_t})$, i.e., a time indexed stream with all visibility records up to time t. An element $\mathbf{s_t}$ is a triple

$$\mathbf{s_t} = \langle u_t, v_t, vis_t \rangle$$

where $u_t, v_t \in \mathcal{R}$ and $vis_t \in \mathcal{C}$. In order to compute the Fourier transform, the visibility vis_t measured at coordinates u_t, v_t in the UV-plane, must be placed at discrete grid points. This discretization is termed gridding [2]. After gridding the SPIFT algorithm [1] is applied on the gridded visibility value to incrementally refine the sky image $\mathbf{I_t}$:

$$\langle u_t, v_t, vis_t \rangle \xrightarrow{\text{grid}(u_t),\text{grid}(v_t)} \langle gu_t, gv_t, vis_t \rangle \xrightarrow{\text{SPIFT}(gu_t, gv_t, vis_t)} \mathbf{I_t} \quad (1)$$

where **grid** refers to the gridding function that is used to place coordinates at discrete grid points. gu_t, gv_t are the gridded coordinates with $0 \leq gu_t, gv_t < N$. A visibility grid $\mathbf{V_t}$ contains all visibilities up to time t. $\mathbf{V_t}$ at time t can be computed incrementally from the visibility grid at time $t-1$ as follows:

$$\mathbf{V_t}[u,v] = \begin{cases} \mathbf{V_{t-1}}[u,v] & \text{if } (u,v) \neq (gu_t, gv_t) \\ \mathbf{V_{t-1}}[u_t,v_t] + vis_t & \text{otherwise} \end{cases} \tag{2}$$

Due to gridding, some parts of sky brightness may be aliased resulting in undesired artifacts in the sky image. The aliasing can be minimized by applying a convolution to visibility values. A convolution is an operation that multiplies a cell of a grid and its neighboring cells with a small matrix known as kernel. The convolved visibility grid $\mathbf{V_t}$ at time t can be computed as:

$$\mathbf{V_t} = \sum_{\substack{a=\frac{-m}{2},\ldots,\frac{m}{2} \\ x=0,\ldots,m}} \sum_{\substack{b=\frac{-m}{2},\ldots,\frac{m}{2} \\ y=0,\ldots,m}} \mathbf{V_{t-1}}[gu_t + a, gv_t + b] + (vis_t \cdot \mathbf{C}[x,y]) \tag{3}$$

where $\mathbf{C}$ is the convolution kernel matrix and m is the size of the kernel matrix.

The goal of this Bachelor's thesis is to efficiently implement the gridded convolution operation in (3) in combination with the SPIFT algorithm to incrementally refine the image.

- **Task 1 - Implementation of SPIFT algorithm in JAVA**
  Study the SPIFT algorithm [1] and implement it in JAVA. Use grayscale images to verify the correctness of your implementation.
- **Task 2 - Integration of SPIFT algorithm in Apache Flink**
  Familiarize yourself with the architecture of the Apache Flink stream processing platform and the basic concepts of stream processing [4]. Integrate the SPIFT implementation from Task 1 into Apache Flink to handle Radio Astronomy Data Streams. The goal is to compute the sky image after gridding the visibility values using the SPIFT algorithm given in (1).
- **Task 3 - Implementation of convolution kernel and computation**
  Implement the prolate spheroidal function and integrate it into the streaming pipeline implemented in Task 2. The first step is a baseline approach for computing the convolved visibility values by applying SPIFT to all convolved values of $\mathbf{V_t}$ as given in (3). Propose and implement an efficient approach that has better performance than the baseline.
- **Task 4 - Experimental Evaluation**
  Evaluate the performance of your approach empirically.
- **Task 5 - Write a Bachelor's thesis**
  Write the Bachelor thesis that describes your solution, analyzes it analytically, and evaluates it empirically. Give a 20 minutes presentation at a DBTG meeting.

References

1. M. Saad, A. Bernstein, M. H. Böhlen, D. Dell'Aglio, "Single Point Incremental Fourier Transform on 2D Data Streams," 2021 IEEE 37th International Conference on Data Engineering (ICDE), Chania, Greece.

2. A. Thompson, J. Moran, and G. Swenson, Jr, Interferometry and Synthesis in Radio Astronomy, 01 1991, vol. -1
3. Taylor, G.B. and Carilli, C.L. and Perley, R.A. and National Radio Astronomy Observatory (U.S.), Synthesis Imaging in Radio Astronomy II: A Collection of Lectures from the Sixth NRAO/NMIMT Synthesis Imaging Summer School Held at Socorro, New Mexico, USA, 17-23 June, 1998
4. P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, and K. Tzoumas. Apache flink: Stream and batch processing in a single engine. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 36(4), 2015.

Supervisor: Muhammad Saad (saad@ifi.uzh.ch)
Start Date: November 16, 2020
End Date: May 16, 2021
Presentation Date: TBD

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor