Analyzing Spectral Information of a Turbulent Jet

Big Data Analysis Final Project Proposal

Eugene Hoffman

This document presents the desired data analysis quest to be pursued for the final project of the big data class ME 5013 by Eugene Hoffman. All critical information will be presented and discussed to give the reviewing committee a full understanding and motivation behind the project.

Project Motivation

This particular project was selected for two distinct reasons which are spelt out below:

a. Handling and analyzing large image data sets: A novel and growing aspect of experimental diagnosis involves the use of non-intrusive diagnostic techniques to obtain data sets from which properties being studied such as velocity, pressure, vorticity, etc. in the case of fluid dynamics can be obtained. This revolution has been spurred on by the advent of high-speed cameras which produce large amounts of image data sets from which desired properties can be obtained. The application of big data handling practices and large matrix manipulations presented in this class can be of great benefit in the analysis of huge image datasets created by these new analysis techniques.
b. Understanding of frequencies in turbulent structures: Several techniques detailed in this class propose various ways of studying and correlating frequency responses between similar signals. The ability to understand the frequency response and characterize the spectral nature of some fluid dynamic interactions will be of great benefit and shed light on some phenomenon currently being studied in the hypersonic aerothermodynamics research areas and hence the desire to study this.

Data Set Information:

a. Size and count

The data set to be used in this research endeavor is made up of a series of images capturing a jet of turbulent gas being discharged from an air duster. Two of these data sets at different pressures would be considered for the analysis. Each data set is made up of at least 100,000 images that are 680 pixels long by 280 pixels wide. Sample images from these two data sets can be found below.

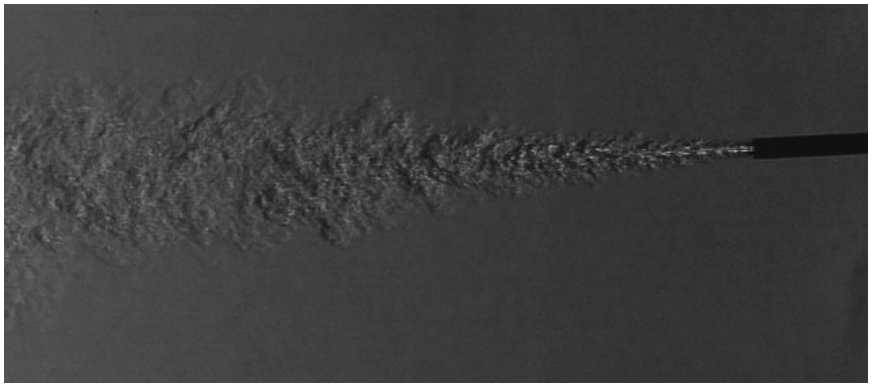

Figure 1. Sample image from first dataset

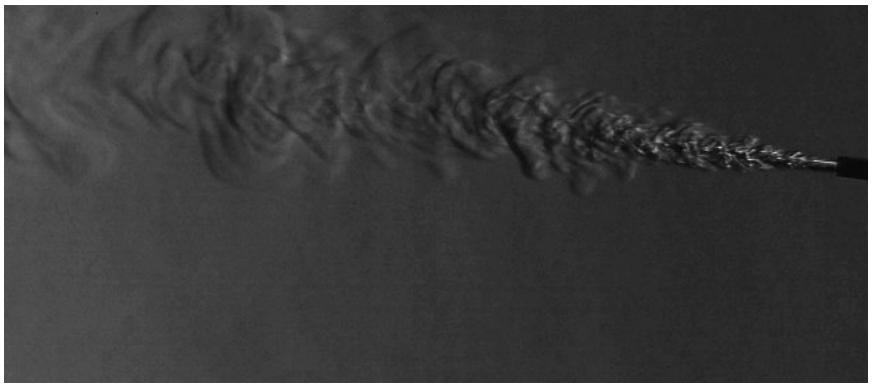

Figure 2. Sample image from second dataset

A breakdown of the image counts and sizes are listed in the table below.

(*Note: - The author decided to change the image file type to PNG's to drastically reduce file size. Test 2a for example was 20GB when originally saved as tiffs but was then reduced to 8.69 GB as PNGs)*

Table 1. Detailed information about image data sets.

Image Set	Pixel size (rows x columns)	Image Count	Data Set Size
Test 1 background	680 x 280	200	14 MB
Test 1 images	680 x 280	100,000	7.62 GB
Test 2 background	680 x 280	2,000	343 MB
Test 2 (a & b) images	680 x 280	150,000	~11 GB

b. Access location

For review purposes, both datasets alongside their respective backgrounds have been uploaded to a Google drive hosting site. A link to these images can be found as follows. (https://drive.google.com/drive/folders/1nu2r1wRfx2k63iDBfLGYdC_K0D3ZFQkS?usp=sharing)

For the project purposes and ease of communication, a secondary location known as Internet Archive, is being explored due to the ease and abundance of python modules available for use with it. This hosting location also provides an abundance of space for uploading datasets. This location is currently under review by the authors and a link to a preliminary data set housed with this hosting site is as follows (https://archive.org/details/@eugene_hoffman)

c. Data set variables

With this current research endeavor, the variables present in the study include the following

1. Pressure,
2. Displacement in the x
3. Displacement in the y,
4. Density shown by pixel intensity.

Grading Metrics

The grading metrics for this particular project is listed below.

a. 15% : Python related work

Use Numpy, Pandas, Matplotlib. Others include Pillow for image analysis, time for tracking processing time and some modules form turbustat.

b. 10% : Data Handling and Manipulation

This will be the majority of my project and will involve shrinking image sizes prior to loading, loading images from web-based applications such as done in turbustat, and getting large image dataset into useable size through various methods including performing an FFT transform of the images to select dominant frequencies to view and analyze.

c. 40% : Data Analytic Methods and analysis

This will basically be centered around Fourier and Spectral analysis of the image data sets (section b and c point allocation might be switched due to the nature of the project. The ajority of the mathematical work will involve Fourier Transforms and cross correlations.

d. 20% : Visualization

Correlations and plotting correlations in an effective manner etc. Investigate various spectral plotting techniques as well as image plotting techniques

e. 15% : Hypothesis-driven results

Major hypothesis of projects. Analyzing spectral relationships in a turbulent jet image set.