

# An Eigenspace Method for Spatiotemporal Multiple Hotspot Detection in Image

CHAUHAN SAURABH

*M.tech (CSE)*

*Institute of Engineering Technology*

*Ahmedabad, Gujarat*

saurabh.chauhan@iet.ahduni.edu.in

**Abstract**—Space and Time are two critical components of many real word systems. For this reason, analysis of anomalies in spatiotemporal data has been great of interest. Hotspot detection aims at identifying subgroups in the observations that are unexpected, with respect to the some baseline information. For instance, in disease surveillance, the purpose is to detect sub-regions in spatiotemporal space, where the count of reported diseases (e.g. Cancer) is higher than expected, with respect to the population. The state-of-the-art method for this kind of problem is the Space-Time Scan Statistics (STScan), which exhaustively search the whole space through a sliding window looking for significant spatiotemporal clusters. A novel methodology called Eigen Spotis proposed where instead of an exhaustive search over the space, tracks the changes in a space-time correlation structure. Not only does the new approach presents much more computational efficiency, but also makes no assumption about the data distribution, hotspot shape or the data quality. The principal idea is that with the joint combination of abnormal elements in the principal spatial and the temporal singular vectors, the location of hotspots in the spatiotemporal space can be approximated.

**Keywords**—Divide & Conquer, Hotspot Detection, Spatiotemporal Data, Eigenspace, SVD, Outbreak Detection

## I. INTRODUCTION

Hotspot detection is used in many disciplines, as in crime analysis, for analyzing where crimes occur with a certain frequency, in fire analysis for studying the phenomenon of forest fires, and in disease analysis for studying the localization and the focuses of diseases. Nowadays, there is great interest in spatiotemporal data analysis due to the huge amount of data available. Among various analysis tasks that can be carried out on spatiotemporal data, hotspot analysis is recognized as an important tool in security informatics and bio-surveillance. For instance in crime hotspot application, an outcome such as City Center between hours 8 to 11pm would be a spatiotemporal hotspot. Outcome like City Center or City Park would be strict spatial hotspots and 8 to 11pm and 10 to 12pm are samples of temporal hotspots. Hotspot analysis goal consists of detecting spatiotemporal regions among data. For instance, in face recognition the specific set of the largest eigenvectors can be used to approximate the images of the human face. In structural engineering, both the eigenvalue and eigenvectors are used to estimate the vibration of structures. In control engineering, the eigenvalues of the linear system are used to assess the stability and response of the system. The report later stated that SVD was the key data analysis tools that

they used. A realistic scenario of the application of the hotspot detection is in disease surveillance. Suppose that we have the population of different postal codes during a range of years as the baseline information and the count of the reported diseases in a range of postal codes, throughout different years as the cases dataset. The goal is to detect those spatiotemporal regions that contain unexpected counts. For instance, the output like zones S1, S2 and S3 during the years T1 to T5 might be considered a spatiotemporal hotspot. The detection of such hotspots enables the officials to better understand their target of interest for essential medical care and preventive measures.

## II. PROPOSED APPROACH

### The Problem

Given a spatiotemporal count matrix for the cases needed for the detection of those spatiotemporal regions (hotspots) that seem unexpected, given the baseline spatiotemporal matrix. Each cell in each matrix represents a count corresponding to a specific region and time. In particular, for disease outbreak detection, each cell in the baseline matrix represents the population corresponding to a region in a specific time period. Each cell in the matrix cases also represents the count of reported disease in a specific region, within a given time period, as well. The purpose is to determine those subgroups of the spatiotemporal space whose reported cases are unexpected. We are interested in developing a system, which has the following characteristics: 1) does not require any input parameter; 2) weighs all the possible hotspots, based on a standard metric like statistical significance (p-value). The benefit is this that the output can be compared to relevant systems or methods. The alpha threshold is also easy to estimate (usually  $\alpha=0.15$ ).

### The Method

Initially, We have two Lena images with a few changes in both images. After that we have converted images into two  $n \times m$  matrices. Assume that we have B (baseline) and C (cases) such that n be the number of components in the spatial dimension and m be the number of components in the temporal dimension. After that Matrix is divided into submatrices of  $3 \times 3$  dimension. The SVD of the sub matrix is a factorization of the form  $M = U \Sigma V^T$ . The 3 columns of U and the 3 columns of V are called the left-singular vectors and right-singular vectors of the matrices, respectively. The left singular vectors correspond to spatial dimension, while the right-singular ones correspond to the temporal dimension. In order to clarify and elaborate, a new terminology spatial singular vector is used along with

temporal singular vector that respectively refers to the principal left singular vector and the principal right singular one. Note that we take only the singular vector corresponding to the largest eigen value for the comparison, due to the fact that the first principal singular vector represents the largest possible variance. Hence, it explains or extracts the largest part of the inertia of the data.

## The EigenSpot Algorithm

First Divide Matrix into Sub Matrix of Dimension 3X3

n: number of items in the spatial dimension

m: number of items in the temporal dimension

B: Baseline n \_ m spatiotemporal matrix

C: Cases n \_ m spatiotemporal Matrix

$\alpha$ : Statistical significance level (e.g. 0.15)

Input: B, C,  $\alpha$

Output: Hotspot

$[\vec{a}_{sb}, \vec{a}_{tb}] = 1\text{-rank SVD (B)}$

$[\vec{a}_{sc}, \vec{a}_{tc}] = 1\text{-rank SVD (C)}$

for i=1:n do

dsi = sci - sbi

end for

for j=1:m do

dtj = tcj - tbj

end for

## Illustrative Example

All Experiments are performed on a Laptop with Linux OS with Intel Core i5 Processor having 8GB RAM. We have used R Language for algorithm implementation.

Initially, we have two 3X6 matrices. After that given two matrices are divided into four 3X3 matrices. Then we apply SVD on all four 3X3 matrices and find low rank SVD of matrices. After that Z Norm is applied to above resultant matrices and finally we compute P-norm of above matrices.

Consider following Result:

```
> a
[1, ] [1, 2, 3, 4, 5, 6]
[2, ] [7, 8, 9, 10, 11, 12]
[3, ] [13, 14, 15, 16, 17, 18]
> b
[1, ] [1, 200, 3, 4, 5, 6]
[2, ] [7, 8, 9, 10, 11, 1200]
[3, ] [13, 14, 15, 16, 17, 18]
```

In Above Snap-Shot we have, One matrix(baseline) that contains value from 1 to 18 row-wise and in another matrix(case)only two elements are changed compare to baseline matrix.

After that above matrices are divided in 3X3 which are shown in below snapshot:

```
>a1
[1, ] [1, 2, 3]
[2, ] [7, 8, 9]
```

```
[3, ] [13, 14, 15]
>b1
[1, ] [1, 2, 3]
[2, ] [4, 5, 6]
[3, ] [10, 11, 12]
>a2
[1, ] [1, 2, 3]
[2, ] [7, 8, 9]
[3, ] [13, 14, 15]
>b2
[1, ] [1, 2, 3]
[2, ] [4, 5, 6]
[3, ] [10, 11, 1200]
[3, ] [16, 17, 18]
```

Result of 1-Rank SVD is shown below:

```
>u1
[1, ] [1]
[2, ] [-0.8704128]
[3, ] [0.4521520]
>v1
[1, ] [1, 2, 3]
[2, ] [-0.4239339]
[3, ] [0.6060769]
>u2
[1, ] [1]
[2, ] [0.2361655]
[3, ] [-0.4720620]
>v2
[1, ] [1, 2, 3]
[2, ] [0.5239453]
[3, ] [0.5668436]
[3, ] [-0.3799268]
```

Z score is applied to matrices and output is shown below:

```
>zu1
[1, ] [1]
[2, ] [-1.1332669]
[3, ] [0.3748536]
>zv1
[1, ] [1, 2, 3]
[2, ] [0.4930033]
[3, ] [-1.150775]
>zu2
[1, ] [1]
[2, ] [0.07600906]
[3, ] [-1.03583566]
>zv2
[1, ] [1, 2, 3]
[2, ] [0.5367685]
[3, ] [0.6170025]
[3, ] [-1.153771]
```

Finally apply P-norm to output of Z score and it is shown below:

>pu1

```
      [, 1]  
[1, ] 0.1285511  
[2, ] 0.3538847  
[3, ] 0.2241018
```

>pv1

```
      [, 1]      [, 2]      [, 3]  
[1, ] 0.3110051 0.1249123 0.2553423
```

>pu2

```
      [, 1]  
[1, ] 0.4697059  
[2, ] 0.1501394  
[3, ] 0.1685712
```

>pv2

```
      [, 1]      [, 2]      [, 3]  
[1, ] 0.2957138 0.2686166 0.124297
```

### III. LIMITATION OF PROPOSED ALGORITHM

Proposed algorithm is fail when Multiple hotspot are there within same matrix because it can detect only one hotspot within matrix.

### ACKNOWLEDGMENT

This work is performed in the context of the Academic project AOBD Winter- 2015 under the guidance of the Prof. Ratnik Gandhi at IET-Ahmedabad University.

### References

- [1] Brand, M.(2006). Fast low-rank modifications of the thin singular value decomposition.Linear algebra and its applications 415(1),20-30J.
- [2] Hadi Fanaee -T and Joao gama ,Eigenspace Methoid for Spatiotemporal Hotspot Dectection.