

University of  
St Andrews

# Quantitative Analysis of Indic Script Development

Vinodh Rajan

vrs3@st-andrews.ac.uk

Dr Mark-Jan Nederhof

mn31@st-andrews.ac.uk

sicsa\*

## Background

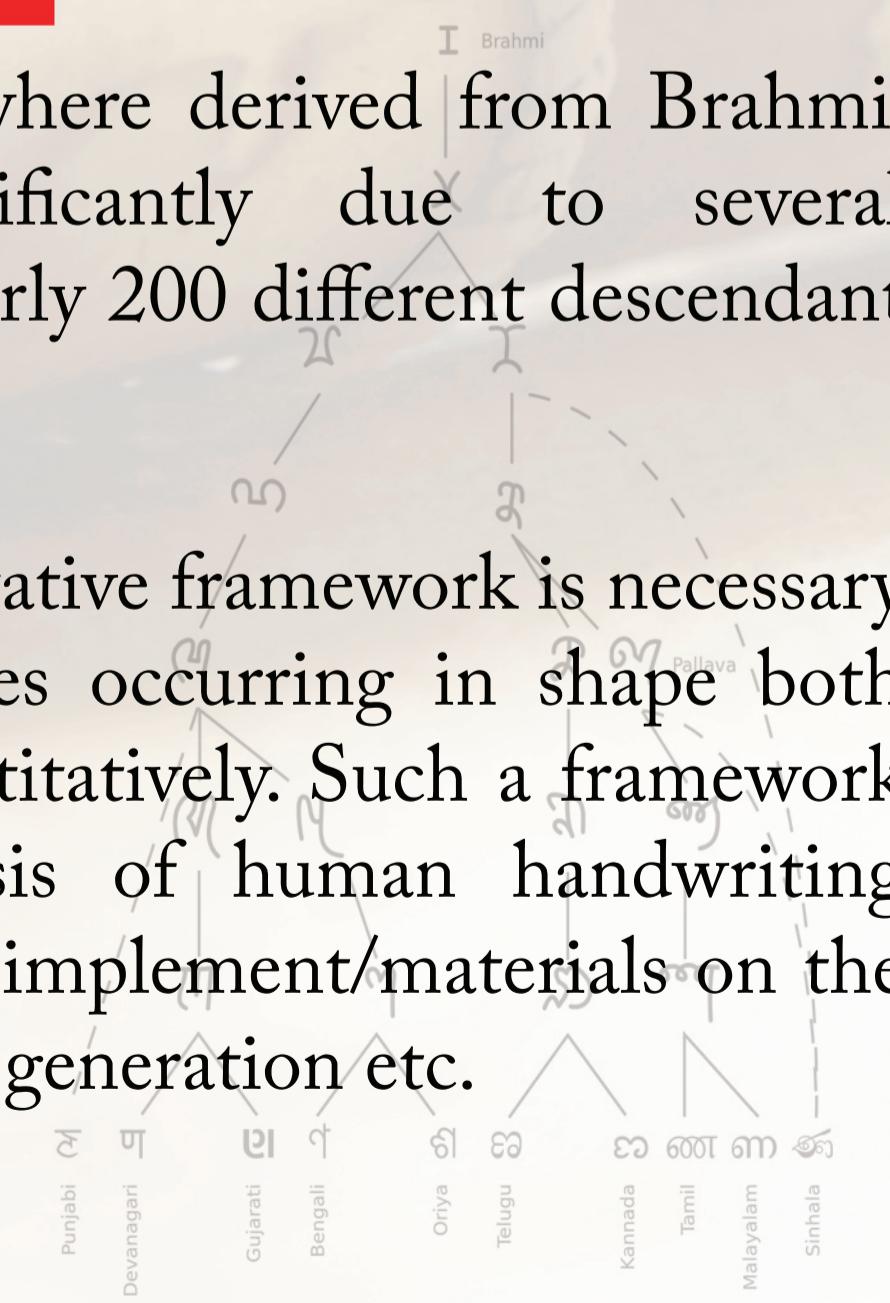
Scripts are usually seen as simple carriers of languages. Research on scripts until recently has been minimal and niche, except for the field of paleography. Scripts are an important part of the cultural heritage of humanity which requires more research to be done in its analysis and study. Fortunately, there is a growing interest in analysis of scripts.

Traditionally, analysis and study in paleography have been usually done manually. Digital paleographic methods are at present making more inroads into the field. However, applying quantitative analysis on scripts is not yet popular and standardized.

## Motivation

Indic scripts, which were derived from Brahmi, have diverged significantly due to several developments into nearly 200 different descendant scripts.

A well-defined quantitative framework is necessary to analyze the changes occurring in shape both qualitatively and quantitatively. Such a framework aids in better analysis of human handwriting behavior, the effect of implement/materials on the script shape, allograph generation etc.



## Quantitative Framework

We propose a analysis framework for scripts that is largely computational and requires minimal user interaction.

### Spline Conversion

The characters of scripts are externally represented as B-splines. B-splines are very efficient in preserving the shape and curvature of glyptic segments and can be manipulated without significant effort.

### Trajectory Reconstruction

Trajectory reconstruction attempts to recover the absent dynamic information which is essential in defining the character. The recovery is performed by conducting a global search using a set of heuristics, such as length minimization and curvature minimization.

### Stroke Segmentation

Stroke segmentation retrieves the structure of the character based on its trajectory. This is performed by segmentation of the character at various important landmark points of the recovered trajectory such as the extrema of curvature, breaking them down into basic strokes.

### Metrics Computation

For quantitative analysis, metrics need to be computed from the characters. These serve to quantify several aspects of the characters such as production, geometry, stroke & cognition. These features/metrics could additionally serve as descriptors for the scripts. As quantitative features these can be widely used in analysis and/or visualization.

## Prototype Implementation

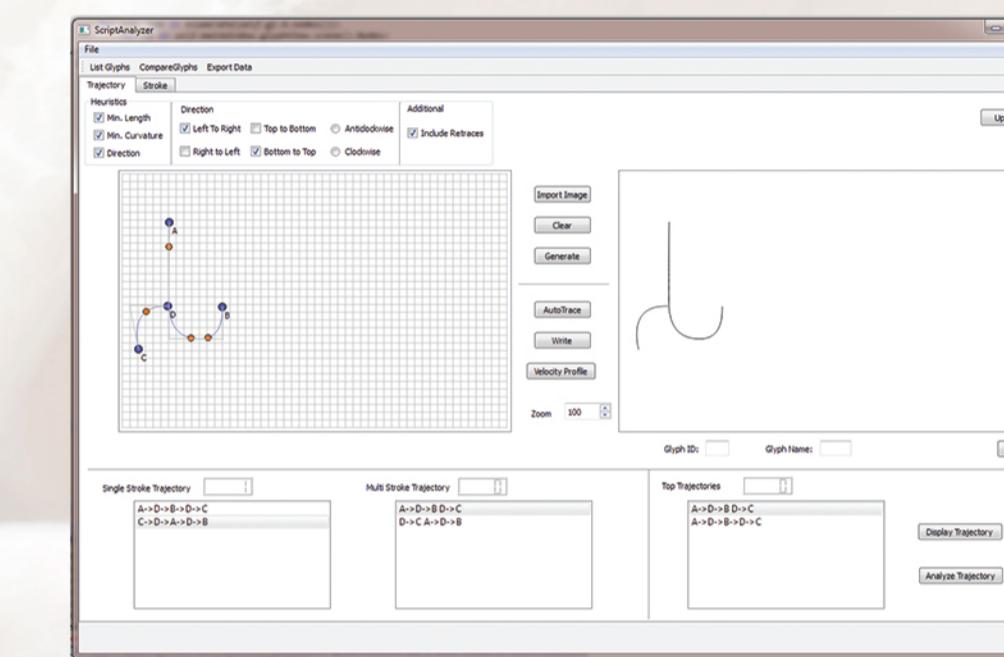


Figure 1: Trajectory Reconstruction

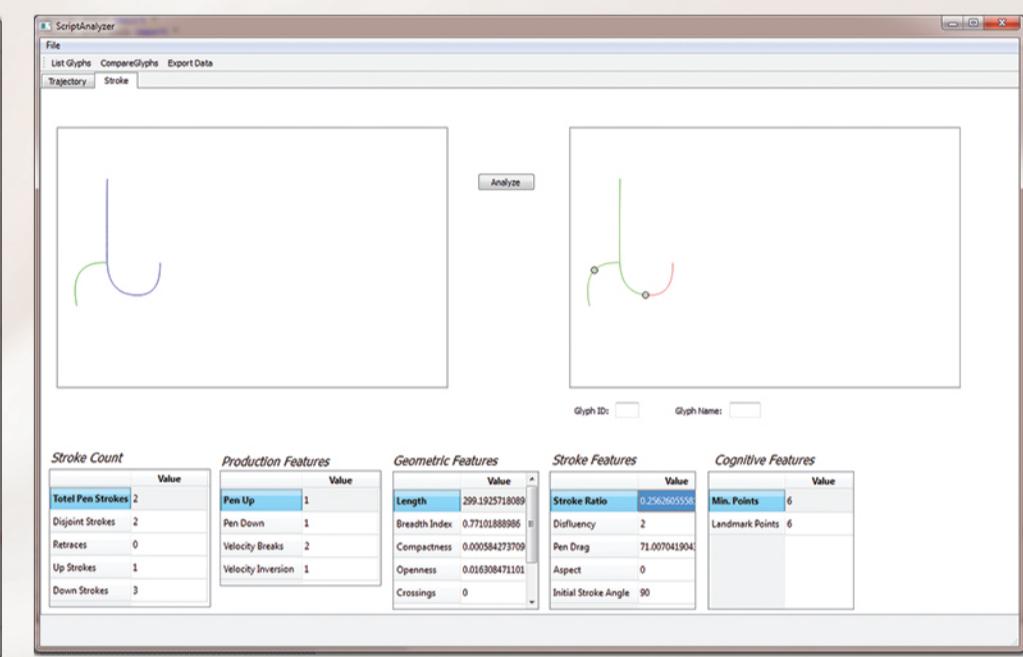


Figure 2: Stroke Segmentation & Metrics Computation



Figure 3: Script Repository

## Future Work

Proper analysis of the various quantitative features would help to better study and understand the characters within a script and also compare several scripts. Visualization is particularly helpful in studying paleographic scripts and analyzing the changes that took place over time.

Various statistical analyses of the features and visualization techniques would be built into the implementation.

Analysis of how the metrics perform with variants of the same script is also to be performed.