

read me file

## 1 *Devanagari* Characters

The following table illustrates the classes of frequently used *Devanagari* characters<sup>1</sup> in roman English<sup>2</sup>. Note that name of the folders in dataset corresponds to *Devanagari* classes of characters.

Consonants					
ka क	kha ख	ga ग	gha घ	cha च	chha छ
ja ज	jha झ	tta ट	ttha ठ	dda ड	ddha ढ
ta त	tha थ	da द	dha ध	na न	pa प
pha फ	ba ब	bha भ	ma म	ya य	ra र
la ल	va व	sa स	ha ह	kshya क्ष	tra त्र
gya ज्ञ					

  

Vowels					
a अ	i इ	ae ए	u उ	uu ऊ	

Table 1: *Devanagari* class of characters in Roman English

## 2 On-line Data

### 2.1 Dataset

We have used a simple Graphite tablet (WCACOM Co. Ltd.), model ET0405A-U, US patent, which is working under 5V DC and 40 mA captures the pen-tip position in the form of 2D coordinates. The data set<sup>3</sup> was composed of 1800 characters from 36 classes of characters, where 25 natives were used. Each writer had given a chance to write two times per class of character. Therefore, for every  $i^{th}$  writer/user, we have stored writing as  $useri\_1$  and  $useri\_2$ .

As no directions, constraints, and instructions were given to the users, the dataset was completely composed of natural handwritings as if they were writing on a piece of paper.

<sup>1</sup>For more information: <http://en.wikipedia.org/wiki/Devanagari>

<sup>2</sup>Roman English is just to label classes of characters which are not strictly identical to officially published one.

<sup>3</sup>The dataset is strictly intended for scientific purpose only.

## 2.2 Data Representation

The digitizer captures a series of strokes during pen movement. A string of coordinates (pen-tip positions) from pen down to pen up movement represents a stroke. Variable numbers of strokes are used to complete a character from time to time writing and from one user to another.

For simplicity, we have inserted  $[-1, -1]$  for the termination of the stroke that makes easier to count and separate strokes in a complete character. The following is an example.

```
200.0, 166.0
201.0, 164.0
203.0, 163.0
.
.
.
323.0, 244.0
320.0, 239.0
-1.0, -1.0
265.0, 178.0
267.0, 178.0
.
.
.
353.0, 180.0
351.0, 178.0
-1.0, -1.0
```

Figure 1: two-stroke character – an example

## 2.3 Visual Representation

In the following Fig. 2, let us consider a class of character क to show different writing styles. Writing style refers to variation of stroke number and order, including structural information. In Fig. 2, each red dot represents stroke initiation. Furthermore, it is helpful for counting the number of stroke.



Figure 2: six different styles of writing क having different stroke number and order

### For more information

**Santosh K. C.**

INRIA Nancy-Grand Est Research Centre,

LORIA - Campus Scientifique

BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex, France

*Email:* Santosh.KC@inria.fr