



Master Thesis Defense

Speaker: Puntis Jifroodian-Haghighi

Supervisor: Dr. Suen

Examining Committee: Drs. Lam, Shiri and Dr. Witte (Chair)

Title: A Discrete HMM for Recognition of Handwritten Frasi Words

Date: Tuesday January 12, 2010

Time: 10:30 am.

Place: EV3.101

ABSTRACT

Handwriting recognition systems (HRS) have been researched for more than 50 years. Designing a system to recognize specific words in a handwritten clean document is not difficult nowadays. However, it is still challenging to develop an HRS with a high recognition rate. Previously, most of the research in the handwriting recognition domain was conducted on Chinese and Latin languages, while recently more people have shown an interest in the Indo-Iranian script recognition systems.

In this thesis, we present an automatic handwriting recognition system for Farsi words. The system was trained, validated and tested on the CENPARMI Farsi Dataset, which was gathered during this research. CENPARMI's Farsi Dataset is unique in terms of its huge number of images (432,357 combined grayscale and binary), inclusion of all possible handwriting types (Dates, Words, Isolated Characters, Isolated Digits, Numeral Strings, Special Symbols, Documents), the variety of cursive styles, the number of writers (400) and the exclusive participation of Native Farsi speakers in the gathering of data.

The words were first preprocessed. Concavity and Distribution features were extracted and the codebook was calculated by the vector quantization method. A Discrete Hidden Markov Model was chosen as the classifier because of the cursive nature of the Farsi script. Finally, encouraging recognition rates of 98.76% and 96.02% have been obtained for the Training and Testing sets, respectively.