Oral presentation: 254

Sinhala handwritten address recognition for postal sorting automation

M. F. A. Ifhaam* and S. G. V. S. Jayalal

Department of Industrial Management, Faculty of Science, University of Kelaniya, Sri Lanka
*ahmedifhaam1@gmail.com

Sri Lankan post office mail sorting process is done manually, even today. Even though employees are well experienced, it takes considerable time and pushes employees to work overtime in places like Central Mail Exchange (CME). With major issues like unclear handwriting, having trouble to recognize some uncommon or ambiguous names and carrying these duties twice a day create a negative impact on the efficiency of the postal delivery system. In the prevailing system, forward mails and delivery mails are the two categories of separating mails at the sorting centers. Delivery mails are the posts which can be delivered to its destination directly. Forward mails are the ones which needs to be sent to an appropriate post office that can deliver the particular post to its destination. Majority of Sri Lankans use Sinhala language for their day to day activities. Then again, less researches had been done on Sinhala handwriting recognition. However, the research recognized only the city names by using the postal sorting domain in 2004. This finding is useful for main post offices, but the number of main post offices in Sri Lanka are limited to 501 whereas, number of sub post offices in Sri Lanka are mentioned as 2953. Sorting process at sub post offices needs street address or one line below in the address line, which cannot be done only using city names. These requirements made the above research less successful. Though similar systems have been implemented in other languages like English, there haven't been a highlighting findings except the above mentioned research due to the complexity of languages like Sinhala. The proposed system is focused on recognition of Sinhala handwriting using Optical Character Recognition (OCR) and image processing technologies. Implementing these techniques to recognize postal address will increase the efficiency of postal mail sorting. Handwritten postal envelopes will be used as the training, testing and validating materials. Therefore, this research is not limited only to a single restricted writing style, but also for unstructured writing styles. In this system, one of the major impediment is touching characters. Segmentation of handwritten touching characters become a crucial step in such systems. Conventional segmentation methods are incapable of handling the complexity of Sinhala handwriting. The proposed method separates touching characters into isolated character models in two steps as described in literature such as basic projection profile method and water reservoir concept. Recognition quality will be extended using heuristics since population of recognizing words are finite. Genetic Algorithms (GA) will be used to generate more optimized results faster with higher accuracy. The Primary Objective of the research would to identify the automatic way of forwarding the letter to the next post office from the current post office. Given addresses are written in the default format. This format can be extended to more formats as improvements in future. Since current methods are completely manual, evaluation should be done with the help of experienced employees at sorting centers.

Keywords: Image processing, machine learning, postal address sorting, Sinhala OCR