

# **A novel SIFT-based codebook generation for handwritten Tamil character recognition**

**Subashini, A. And Kodikara, N.D.**

## **Abstract**

A method for the off-line recognition of Tamil handwriting characters based on local feature extraction is investigated. In the proposed method each pre-processed character is represented by a set of local SIFT feature vectors. From a large set of SIFT descriptors, the key idea is to create a codebook for each character using K-means clustering algorithm. K-means is an optimisation algorithm but this algorithm takes very long time to converge. We construct an initial codebook by using the Linde Buzo and Gray (LBG) algorithm so that the convergence time for K-means is reduced considerably. Target character is recognised into one of twenty categories by k-nearest neighbour classification. An average recognition rate of 87% on the character level has been achieved in experiments using six thousand training and two thousand testing images of twenty selected characters. Further study may include more characters and more samples being recognised with better classifier.