

DIRECTION HISTOGRAM: NOVEL DISCRIMINATIVE GLOBAL FEATURE
FOR THAI OFFLINE HANDWRITTEN OCR

EKAWAT CHAOWICHARAT 5237083 SCMA/D

Ph.D. (MATHEMATICS)

THESIS ADVISORY COMMITTEE: KANLAYA NARUEDOMKUL, Ph.D.

NICK CERCONE, Ph.D., LALITA NARUPIYAKUL, Ph.D.

ABSTRACT

The image feature used for classification is a crucial part of an optical character recognition system. To achieve a high accuracy for offline handwritten character recognition, the feature should capture and emphasize the differences between different characters classes and ignore the differences among the various drawings of the same character class. In this research, we present a novel image feature called direction histogram (DH) and a feature extraction algorithm called bag of histogram (BoH). Unlike the traditional image feature, DH is a global feature that describes pixel density in every direction around each center, which tolerates stroke thickness and variation, and omits the stroke connectivity (if any). BoH is the algorithm to count the proportion of distinct DHs in an unseen image and generate a one-dimensional feature vector. This vector is compressed by PCA and is classified by a neural network. Fifty-two datasets, each containing 30 drawings of 80 Thai characters, are used for training our OCR system, and the original, thick, and distorted handwritten datasets are used for testing. The recognition system with our proposed DH and BoH feature extraction yielded higher recognition accuracy when compared to the popular convolutional neural network.

KEY WORDS: OPTICAL CHARACTER RECOGNITION / DIRECTION HISTOGRAM /
BAG OF HISTOGRAM / HANDWRITTEN / OFFLINE / NEURAL
NETWORKS

107 pages