

NHẬN DẠNG DẤU THANH VÀ MŨ TRONG KÝ TỰ TIẾNG VIỆT VIẾT TAY

RECOGNIZING ACCENT IN VIETNAMESE HANDWRITING CHARACTERS

Tác giả: [Huỳnh Hữu Hưng](#), [Nguyễn Trọng Nguyên](#)

Tóm tắt bằng tiếng Việt:

Nhận dạng ký tự viết tay là chủ đề đang được nghiên cứu rộng rãi hiện nay. Nhiều giải pháp được áp dụng cho ký tự tiếng Anh và mang lại hiệu quả cao. Tuy nhiên, bài toán tiếng Việt đang đối mặt với nhiều thách thức, phần lớn vấn đề đến từ dấu mũ và dấu thanh (gọi chung là dấu). Bài báo này đề xuất một giải pháp nhận dạng dấu, kể cả trong trường hợp mũ và thanh trong ký tự dính liền nhau – một vấn đề rất phổ biến và ảnh hưởng đáng kể đến kết quả nhận dạng. Chúng tôi bắt đầu với việc tách riêng dấu và chữ cái bằng kĩ thuật gán nhãn vùng liền thông. Tiếp theo, dấu được phân vào hai lớp: dấu đơn (một dấu mũ hoặc dấu thanh) hoặc dấu kép (mũ và thanh dính liền nhau). Dấu kép (nếu có) tiếp tục được tách ra thành các dấu đơn để chuẩn bị cho bước xử lý cuối cùng. Quá trình nhận dạng được thực hiện bằng phương pháp mô hình Markov ẩn. Giải pháp đề xuất đã được thử nghiệm và thu được kết quả khả quan.

Từ khóa: ký tự viết tay tiếng Việt; dấu mũ; dấu thanh; bộ phát hiện góc; tách nhánh; mô hình Markov ẩn

Tóm tắt bằng tiếng Anh:

Handwriting character recognition is one of the most common research topics. Many approaches have been applied to English characters and achieved high accuracy. However, the complexities in the language of each country are not same. Recognizing Vietnamese handwriting character is facing many problems, most of them come from the accent. This paper focuses on accent recognition, especially when there is a connection between two accents - a common problem which affects the identification result. Our approach starts with separating accents from characters using the connected-component labeling method. The obtained accent then is checked if it is single or multiple (the combination of many accents). In the second case, each accent is extracted for the last processing step. The recognition is performed using hidden Markov models with each single accent. Proposed solution has been tested and obtained with high accuracy.

Key words: Vietnamese handwriting character; accent; corner detector; branch separating; hidden Markov models