



UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA



**RECONHECIMENTO DE CARACTERES
MANUSCRITOS JAPONESES**

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Marconi Emanuel Madruga Filho (memf@cin.ufpe.br)
Orientador: George Darmiton da Cunha Cavalcanti (gdcc@cin.ufpe.br)

Recife, 18 de março de 2008

Índice

1. Contexto.....	3
2. Projeto.....	4
3. Cronograma	5
4. Referências	6
5. Assinaturas	7

1. Contexto

A análise e reconhecimento de documentos é uma área muito desafiadora dentro de visão computacional que possui aplicações que vão desde o reconhecimento de caracteres, símbolos, palavras, imagens, dentre outros, até o que pode ser considerado mais alto, a compreensão semântica do texto. Apesar dos grandes avanços que surgiram na área nas últimas décadas, ainda há muitos problemas que permanecem em aberto.

No âmbito de reconhecimento textual, pode-se dizer que, para texto digitado (*typewritten text*), o problema já foi resolvido. Muitos algoritmos eficientes já atingem, comprovadamente, ótimas taxas de acerto. No entanto, se olharmos para o texto manuscrito (*handwritten text*), a situação muda bastante. Apenas em condições muito restritas de domínio consegue-se obter taxas comparáveis às de texto digitado.

O reconhecimento de texto manuscrito é uma grande área e pode ser vista a partir de diversos ângulos. Pode-se tratar tanto de palavras como um todo quanto de caracteres individualmente. No que tange a caracteres, há uma diferença clara entre caracteres ocidentais e orientais. Essa diferença consiste principalmente na grande disparidade em relação ao tamanho do conjunto de classes de caracteres, como também à relevância do número de traços em um caracter. No Japão, por exemplo, falantes nativos têm domínio de mais de 3000 caracteres, e 2000 são exigidos para ler um jornal ou revista.

As técnicas de reconhecimento para esses dois tipos de caracteres são, obviamente, influenciadas por essas distinções. A quantidade enorme de caracteres orientais torna difícil o uso de redes neurais, classificadores comumente utilizados para caracteres orientais. Os caracteres orientais, então, recaem quase sempre nos classificadores de vizinho mais próximo (*nearest neighbour classifiers*) [1].

Tais técnicas de reconhecimento ainda podem ser divididas de acordo com a abordagem utilizada, podendo ser *on-line* ou *off-line*. Técnicas *off-line* utilizam apenas as imagens dos caracteres, enquanto técnicas *on-line* consideram informações da dinâmica de escrita. Atualmente, independente de *on-line* ou *off-line*, a melhoria das taxas de acertos das técnicas é fundamental para o desenvolvimento da área.

2. Projeto

Este trabalho de graduação se propõe a: estudar o estado da arte das técnicas de reconhecimento de caracteres japoneses, obter e verificar uma base de dados de caracteres e implementar uma aplicação de reconhecimento utilizando a base de dados obtida.

O estudo do estado da arte tem como objetivos ampliar a compreensão do domínio e, principalmente, a escolha da técnica que será utilizada na aplicação a ser desenvolvida. Há técnicas *off-line* como [2] e *on-line* como em [3] e a escolha entre essas duas abordagens será determinante para o próximo passo.

As bases de dados para técnicas *on-line* e *off-line* são também diferentes, visto que as informações necessárias em cada caso são bem distintas. Ao escolher o tipo de técnica a ser desenvolvido, será buscada uma base para atender à sua necessidade.

Por fim, de posse da base e da técnica, será desenvolvido o software de reconhecimento. A aplicação, diante da entrada de um carácter manuscrito por parte do usuário, sugerirá que caracteres possivelmente ele representa, de acordo com a técnica escolhida.

O principal foco deste trabalho é estudar e entender o universo desafiador do reconhecimento de caracteres orientais, ainda mais complexo que o ocidental, a que estamos acostumados.

4. Referências

[1] JAEGER, S., LIU, C.-L., NAKAGAWA, M. *The state of the art in Japanese online handwriting recognition compared to techniques in western handwriting recognition*. International Journal on Document Analysis and Recognition (2003) pg. 75-88

[2] ISHIDERA, E., NISHIWAKI, D., SATO, A. *A confidence value estimation method for handwritten Kanji character recognition and its application to candidate reduction*. International Journal on Document Analysis and Recognition (2004) pg. 263-270

[3] KITADAI, A., NAKAGAWA, M. *Prototype learning for structured pattern representation applied to on-line recognition of handwritten Japanese characters*. International Journal on Document Analysis and Recognition (2007) pg. 101-112

5. Assinaturas

Recife, 18 de março de 2008

George Darmiton da Cunha Cavalcanti
Orientador

Marconi Emanuel Madruga Filho
Aluno