

SISTEMA AUTOMÁTICO DE RECONHECIMENTO ÓPTICO DE CARACTERES A PARTIR DE UM CONJUNTO DE EXEMPLOS

Ogê Marques Filho

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DO PARANÁ
CURSO DE PÓS-GRADUAÇÃO EM INFORMÁTICA INDUSTRIAL (CPGII)
Av. 7 de setembro 3165. CEP 80230 - CURITIBA - PR - BRASIL

RESUMO - É apresentado um sistema para reconhecimento de caracteres no qual cada caractere é descrito como uma combinação de diversas máscaras de correlação representando seus principais atributos. O sistema permite que tais descrições possam ser obtidas automaticamente, a partir de uma biblioteca de máscaras e de um conjunto de exemplos de cada caractere.

1. INTRODUÇÃO

1.1. Estudos prévios sobre o problema

Em trabalho anterior (Marques, 1988) foram publicados resultados de um primeiro estudo sobre a viabilidade de implementação de um esquema de reconhecimento de caracteres baseado na técnica de correlações com máscaras (*template matching*), utilizando um *hardware* específico (PAPS - *Picture Acquisition and Processing System*) controlado por rotinas de *software*. As principais características do esquema proposto eram:

(i) cada caractere deveria ser reconhecido independentemente de informação a respeito de sua posição na imagem, isto é, não havia segmentação precedendo a etapa de reconhecimento.

(ii) as rotinas de reconhecimento deveriam ser tolerantes a variações de tipos de impressão (*fonts*), fornecendo resultados confiáveis para qualquer tipo de impressão utilizado normalmente na documentação escrita.

(iii) cada caractere era reconhecido utilizando operações de correlação com máscaras, buscando identificar propriedades

físicas (*features*) presentes e/ou ausentes em um dado caractere, combinando-as através de operações lógicas.

Rotinas de reconhecimento tolerantes a variações de tipos de impressão para todas as letras maiúsculas do alfabeto foram obtidas e testadas. Os resultados obtidos naquela ocasião incentivaram a sugestão de aprimoramentos adicionais a fim de evoluir para um sistema automático e com maior confiabilidade.

1.2. O problema atual

Em estudos anteriores percebeu-se que alguns caracteres necessitam descrições diferentes quando se utilizam tipos de impressão diferentes. Apesar de todas as tentativas no sentido de tornar as descrições suficientemente flexíveis para operar com alta confiabilidade independentemente do tipo utilizado, observou-se que a especificação "manual" das descrições dos caracteres pode conduzir a resultados falhos, por exemplo descrições inconsistentes ou redundantes. Além disso, torna-se uma atividade pouco criativa, bastante cansativa e sujeita a erros.

No trabalho atual, deseja-se implementar um método a partir do qual as descrições dos caracteres para um dado tipo de impressão são geradas automaticamente, tomando-se um conjunto de exemplos como referência. A derivação automática das descrições a partir de um conjunto de exemplos garante uma análise uniforme (imparcial) de todos os caracteres do conjunto de símbolos. Pequenas modificações na biblioteca de máscaras podem ser implementadas facilmente se a geração de uma nova descrição baseada na biblioteca modificada puder ser feita automaticamente.

Formalizando o problema, deseja-se um sistema que, dado um conjunto de exemplos impressos em um tipo específico, seja capaz de:

(i) extrair automaticamente as propriedades características de cada caractere a partir dos exemplos;

(ii) compilar uma descrição - escrita sob a forma das máscaras que devem ser aplicadas ao caractere - capaz de reconhecer as propriedades selecionadas, isto é, que reconheça apenas as ocorrências do caractere desejado em uma imagem.

1.3. A solução proposta

O sistema projetado opera assumindo que existe uma biblioteca contendo máscaras capazes de reconhecer as principais propriedades de um caractere, bem como um conjunto de exemplos criteriosamente escolhidos.

Dispondo-se das máscaras e exemplos, correlacionam-se todas as máscaras com todos os exemplos, gerando hipóteses sobre o caractere que se deseja reconhecer. Numa etapa subsequente, estas hipóteses são testadas a fim de verificar sua utilidade na descrição do caractere. Neste estágio o sistema deve verificar se houve algum resultado positivo de correlação, quão significativo e discriminativo é este resultado e qual sua frequência de ocorrência. Se o sistema obtiver conclusões positivas sobre estas verificações, acredita-se que a hipótese em questão é válida e, portanto, deve ser anexada à descrição do caractere.

O sistema foi projetado de forma modular, de modo que todos os passos intermediários são facilmente acessíveis, permitindo que aprimoramentos em qualquer um deles sejam implementados e testados separadamente, caso isto seja conveniente.

O equipamento utilizado consistia de câmara, monitor, o sistema PAPS com seus módulos de hardware e o computador principal (SUN-3), no qual se pode executar as rotinas de controle do hardware, utilizando os recursos do Sistema POPLOG ("POPLC; System") (Marques, 1988). Os programas que compõem os diversos módulos do sistema foram escritos em linguagem POP-11 (Barrett, Ramsay e Sloman, 1985).

2. DESCRIÇÃO DO SISTEMA

2.1. Introdução

O sistema desenvolvido baseia-se no princípio do aprendizado a partir de exemplos. Partindo de uma biblioteca de máscaras e de um conjunto de exemplos convenientemente escolhidos, ele produz uma descrição - escrita como uma lista de máscaras acompanhadas de suas respectivas operações de deslocamento - com a qual o programa encarregado de processar estas descrições ("recognizer") irá trabalhar.

Durante a fase de aprendizagem ou treinamento, todas as máscaras e diversas combinações de máscaras são correlacionadas aos exemplos, gerando hipóteses sobre o caractere que se deseja reconhecer. Estas hipóteses são então avaliadas para verificar quão significativos e discriminativos são os resultados e qual sua frequência de ocorrência na fase de treinamento. Se um determinado resultado de correlação com máscara aparece na maioria dos - preferencialmente em todos os - exemplos e auxilia a discriminar o caractere que se deseja reconhecer, ele é considerado um bom candidato para o bloco "compilador de descrições". Este bloco avaliará os melhores resultados de correlação com máscara sobre um conjunto de imagens padronizadas e escolherá quais passos utilizar, bem como decidirá sobre o uso de operações de "crescimento" (dilatação). A saída do "compilador de descrições" é uma lista utilizada pelo "recognizer" para identificar o caractere desejado numa imagem real.

Durante a operação, o programa "recognizer" processará essas listas de descrição, linha por linha, executando as seguintes operações:

- (i) correlação das máscaras indicadas com a imagem.
- (ii) deslocamento dos resultados da correlação a um ponto de referência.
- (iii) combinação de cada resultado intermediário com o seu antecessor através de uma operação "and".

2.2. A fase de treinamento ou aprendizagem

A fig. 1 ilustra os principais blocos componentes do sistema e a forma através da qual eles estão correlacionados. Estes blocos serão discutidos em maior detalhe nos parágrafos subseqüentes.

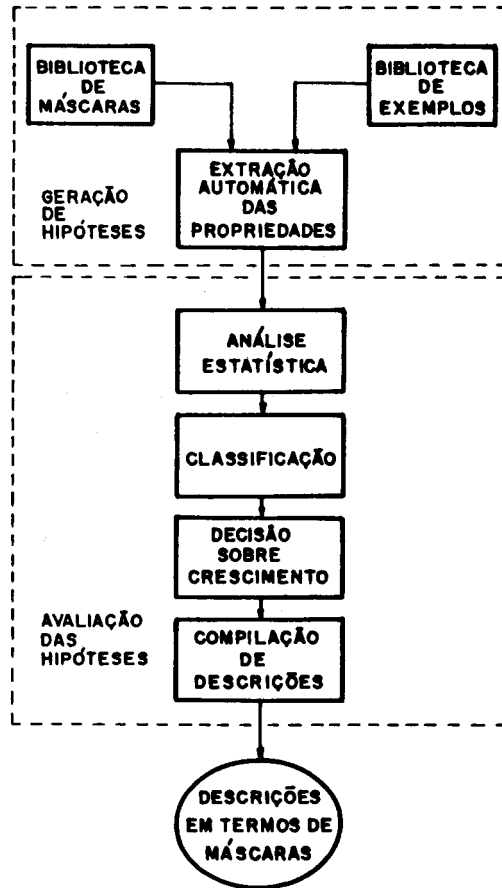


Fig. 1 : Fase de treinamento - diagrama em blocos.

(i) A biblioteca de máscaras

A biblioteca de máscaras utilizada contém máscaras para detecção de barras horizontais, verticais e diagonais, além de cantos, orifícios, bordas horizontais e verticais, arcos e outros atributos passíveis de inclusão na descrição de um

caractere. Além disso, existem máscaras capazes de executar operações de "crescimento" (dilatação), direcionais ou não. Estas operações de dilatação são úteis para minimizar os erros causados por caracteres mal impressos, aumentando a flexibilidade e a confiabilidade do processo.

(ii) A biblioteca de exemplos

A biblioteca de exemplos contém imagens de caracteres impressos com o tipo "helvetica" em 2 tamanhos diferentes, com 5 diferentes qualidades de impressão para cada tamanho. Cada imagem contém 6 exemplos do caractere em questão, totalizando, portanto, 60 (sessenta) exemplos por caractere.

(iii) A extração automática das propriedades

Neste estágio são efetuadas operações de correlação de todas as máscaras com todos os exemplos. Como saída, é gerada uma tabela que indica, para cada exemplo, quais máscaras fornecem resultados positivos de reconhecimento, quais as dimensões (em pixels) destes resultados e quais as coordenadas (xs,ys) destes resultados em relação a um par de eixos cuja origem é o centro de gravidade do caractere. As informações sobre coordenadas foram normalizadas para um caractere de 100 pixels de altura. Dentro deste estágio há um programa encarregado de eliminar os resultados de operações de correlação cuja dimensão e/ou formato pouco ou nada acrescentem à descrição do caractere em questão.

(iv) A análise estatística

Este bloco compõe-se basicamente de 2 *procedures*. O primeiro tem por função agrupar resultados de operações de correlação obtidos com máscaras diferentes mas que se referem à mesma propriedade física do caractere; o segundo calcula a média e o desvio padrão dos deslocamentos (xs,ys), bem como a frequência de ocorrência de cada propriedade. Com base nestes dados, pode-se avaliar a utilidade de uma certa propriedade, levando-se em consideração tanto sua frequência de ocorrência como a precisão espacial dos resultados das operações de

correlação. Estes dois aspectos serão ponderados na compilação da descrição final do caractere.

(v) Classificação e decisão sobre o uso de operações de crescimento (dilatação)

A esta altura do processamento, já se possui informações a respeito do comportamento estatístico das propriedades dos caracteres. O problema passa então a ser como utilizar estes resultados, convertendo-os em uma descrição para o caractere. No sistema implementado, este problema foi solucionado da seguinte maneira:

1. Os resultados estatísticos são ordenados segundo sua frequência de ocorrência: apenas resultados com frequência maior ou igual a 98% são considerados.

2. Para cada valor de frequência de ocorrência, os resultados são novamente ordenados, de acordo com os menores desvios padrões.

3. Com base no desvio padrão de cada passo, decide-se sobre o uso de operações de "crescimento" ou não.

À saída deste estágio dispõe-se de uma lista de candidatos para o bloco "compilador de descrições".

(vi) A compilação das descrições

O compilador de descrições, a partir da lista de candidatos fornecida pela etapa de classificação, fornece uma lista de descrição ótima para o caractere em questão. Há diversas maneiras de obter esta lista ótima e os resultados obtidos com cada método podem variar significativamente. No sistema proposto, foram implementados e testados dois métodos simples de compilação. Em ambos os métodos a decisão é tomada iterativamente da seguinte maneira:

A partir de uma imagem de referência contendo o alfabeto maiúsculo impresso com o tipo "helvetica" e da "lista de candidatos" à descrição de um determinado caractere, o compilador verifica a importância de cada passo da lista de

candidatos na descrição como um todo. Basicamente, este programa acrescenta um passo à lista de descrição do caractere (originalmente vazia) se o passo em questão auxilia na discriminação do caractere desejado e não degrada seriamente os resultados positivos de reconhecimento já obtidos.

A principal diferença entre os dois métodos testados reside no programa responsável pela inclusão ou não de um determinado passo na lista de descrição ótima: enquanto o primeiro método utiliza a mesma imagem de referência com o alfabeto maiúsculo impresso com o tipo "helvetica" durante todo o tempo de compilação, o segundo comuta para outra imagem assim que a "lista de descrição mínima" (i.e., a lista cujos passos são necessários e suficientes para distinguir o caractere desejado de qualquer outro caractere presente na imagem de referência) for obtida. A partir deste ponto, ele utiliza uma imagem com 6 exemplos de tamanho médio, com má qualidade de impressão, do caractere desejado e elimina qualquer passo cuja inclusão degrade qualquer um dos resultados de reconhecimento obtidos até então. Apesar desta mudança parecer bastante simples, os resultados obtidos com o segundo método foram sensivelmente melhores, como exposto no item 3.

3. O SISTEMA SOB TESTE

3.1. Testes realizados

A fim de analisar quantitativamente a performance do sistema decidiu-se testá-lo usando imagens contendo todas as combinações possíveis de letras maiúsculas impressas com o tipo "helvetica", usando as opções "normal" para espessura e "spread 0" e "spread -1" para espaçamento entre os caracteres. "Spread 0" significa espaçamento normal entre os caracteres, enquanto na opção "spread -1" os caracteres estão mais próximos um do outro.

3.2. Resultados obtidos

Os resultados das experiências estão resumidos na tabela

1. Nela são apresentadas as taxas de confiabilidade para imagens "normal spread 0", comparando os dois métodos utilizados. Foram definidas e medidas duas taxas percentuais de confiabilidade:

- Taxa de Caracteres Não-detetados (TCND): indica a diferença entre o número total de vezes em que um caractere aparece e o número de resultados positivos de reconhecimento, expressa em %. Portanto, seu valor ideal é 0 (zero).

- Taxa de Reconhecimento Falso (TRF): expressa o número de reconhecimentos indesejados obtidos ao executar uma descrição, sobre o número total de caracteres presentes no conjunto de imagens considerado, em %. Ela pode ser interpretada como a probabilidade de se reconhecer um caractere como sendo algum outro caractere. Seu valor ideal também é 0 (zero).

Analisando os resultados obtidos, pode-se dizer que:

- consideram-se boas as descrições que apresentaram TCN menor que 10% e TRF menor que 0.5%, quando um texto normalmente espaçado é utilizado como imagem de entrada. Tomando por base o segundo método, este foi o caso para 11 (conze) caracteres, a saber 'B', 'E', 'G', 'K', 'O', 'P', 'Q', 'S', 'T', 'Y' e 'Z'. A performance das descrições dos caracteres 'B', 'E', 'P', 'Q', 'S', 'T', 'Y' e 'Z' pode ser distinguida devido aos excelentes resultados obtidos (0.00% em ambas as taxas).

- os maus resultados apresentados pelas letras 'I', 'J' e 'L' eram esperados. Existe uma dificuldade natural em descrever estes caracteres utilizando o método proposto, porque não há muitos atributos para usar em suas descrições. Além disso, estes atributos são comuns a muitos outros caracteres ou combinações de caracteres, o que aumenta a probabilidade de reconhecimento indesejado.

- houve uma clara redução na maioria das TCNs com a utilização do segundo método, algumas vezes às custas de um aumento na respectiva TRF. Em termos globais, pode-se considerar os resultados obtidos com o segundo método como sendo os melhores.

TABELA 1 Taxas de confiabilidade

CARACTERE	Taxa de Caracteres Não-detetados (TCN) (%)		Taxa de Reconhecimento Falso (TRF) (%)	
	1o. método	2o. método	1o. método	2o. método
A	43.40	35.85	0.00	0.00
B	1.75	0.00	0.00	0.00
C	86.79	0.00	0.07	1.87
D	0.00	0.00	1.11	2.56
E	0.00	0.00	0.00	0.00
F	17.86	12.50	0.97	1.04
G	26.79	5.36	0.00	0.14
H	0.00	0.00	3.32	3.39
I	0.00	0.00	100.00	100.00
J	0.00	0.00	100.00	100.00
K	1.79	1.79	0.00	0.00
L	100.00	100.00	13.57	13.57
M	12.96	12.96	0.55	0.48
N	94.55	25.45	0.00	0.00
O	50.00	1.79	0.00	0.07
P	0.00	0.00	0.21	0.00
Q	15.25	0.00	0.00	0.00
R	0.00	0.00	0.00	3.95
S	0.00	0.00	0.00	0.00
T	0.00	0.00	0.00	0.00
U	0.00	0.00	0.28	1.39
V	16.07	16.07	0.14	0.48
W	98.18	96.36	0.00	0.00
X	0.00	0.00	39.75	39.75
Y	5.45	0.00	0.00	0.00
Z	0.00	0.00	0.00	0.00

4. CONCLUSÕES

Com base no trabalho relatado neste artigo, implementou-se um sistema completo capaz de gerar automaticamente listas de descrições de caracteres a partir de um conjunto de exemplos.

Um dos pontos fracos do sistema foi a necessidade de adoção de limiares inflexíveis em decisões em várias fases intermediárias. Estas decisões muitas vezes limitaram a performance dos estágios subsequentes e fizeram com que uma parcela valiosa de informação fosse irremediavelmente perdida.

Comparando as listas de descrições derivadas automaticamente com aquelas escritas manualmente durante o estudo anterior (Marques, 1988), conclui-se que as principais diferenças foram:

- o problema da tolerância a variações de tipos impressos. Enquanto nas descrições anteriores alguns passos tinham que ser incluídos apenas para satisfazer este critério, nas listas geradas automaticamente esta restrição não mais existe.

- as listas geradas manualmente são normalmente mais completas que as automáticas. Na maioria dos casos, isto se deve ao fato de que as listas manuais contêm passos para a detecção de propriedades ausentes (orifícios, aberturas, etc.), enquanto as automáticas só incluem atributos presentes (barras, cantos, etc.) no caractere.

Foram obtidos 11 (onze) bons resultados para as taxas de confiabilidade quando um texto utilizando tipo e espaçamento convencionais na documentação escrita é tomado como referência. Embora este número sozinho represente menos da metade do alfabeto, sabe-se quais as razões pelas quais os demais caracteres não apresentaram desempenho comparável.

Com base nos aprimoramentos obtidos quando da passagem do primeiro para o segundo método de compilação, acredita-se que estudos adicionais visando um melhor estágio compilador de descrições possivelmente conduzirão a melhores taxas de confiabilidade. Sabe-se, entretanto, que outras fases do processamento necessitam aprimoramentos que podem vir a ser tão ou mais importantes no contexto global.

REFERÊNCIAS BIBLIOGRÁFICAS

- Barrett, R. & Ramsay, A. & Sloman, A., (1985). POP-11: A Practical Language for Artificial Intelligence, Chinchester, Inglaterra, Ellis Horwood Ltd.
- Marques O., (1988). "Character Recognition Using a PAPS". Philips International Institute, Eindhoven, Holanda.