

UMA PROPOSTA DE RECONHECIMENTO DE CARACTERES UTILIZANDO  
CORRELAÇÃO COM MÁSCARAS BINÁRIAS (BINARY TEMPLATE MATCHING)

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** - São apresentados resultados de experiências práticas em Reconhecimento Óptico de Caracteres (OCR), utilizando a técnica de correlação com máscaras binárias. Rotinas de reconhecimento para todas as letras maiúsculas do alfabeto e suas respectivas taxas de confiabilidade são avaliadas.

## 1. INTRODUÇÃO

O Reconhecimento Óptico de Caracteres (*Optical Character Recognition - OCR*) emergiu como uma das áreas de maior interesse científico e comercial da década de 80 e tende a canalizar esforços e interesses crescentes nos próximos anos. Suas aplicações são inúmeras, desde a inspeção de códigos de peças e produtos em sistemas de manufatura até a leitura, armazenamento e transmissão de documentos, no processo conhecido como "documentação eletrônica".

Inúmeras técnicas de OCR foram desenvolvidas ao longo dos últimos 40 anos por diversos pesquisadores, no sentido de reconhecer automaticamente caracteres impressos e/ou manuscritos. Estas técnicas variam consideravelmente de acordo com as propriedades físicas (*features*) escolhidas do caractere, a maneira pela qual estas propriedades são extraídas e o esquema de classificação utilizado.

As metodologias de OCR costumam tratar o problema do reconhecimento em duas etapas: segmentação (decomposição da imagem original em imagens contendo caracteres separados) e clas-

sificação. Estas duas fases são consideradas como o ponto de partida da maioria das pesquisas publicadas nesta área. Neste trabalho propõe-se um esquema no qual os caracteres são reconhecidos independentemente de informação a respeito de suas posições.

Nesta abordagem (fig. 1), qualquer caractere pode ocorrer em qualquer posição da imagem. Uma etapa de pré-processamento executa operações de realce e limiarização, produzindo uma imagem P&B com a qual o sistema trabalha. Uma vez que não há segmentação, as operações de correlação são efetuadas sobre todos os pontos da imagem, ou seja, a existência de um caractere é detetada independentemente de sua posição na imagem (excetuando-se, obviamente, os casos em que o caractere esteja rotacionado de um ângulo considerável). Devido ao elevado número de operações de correlação necessárias, o sistema requer alta velocidade de processamento, obtida através de *hardware* específico. Com base nestas operações de correlação de todos os pontos da imagem com as máscaras representativas dos principais atributos de um caractere, são geradas hipóteses sobre o caractere que se deseja reconhecer. Estas hipóteses podem ser testadas via *software*, obtendo-se, então, o resultado final da etapa de reconhecimento. Se todos os atributos especificados estão presentes e guardam entre si uma posição geométrica correta, o caractere é dito reconhecido.

O trabalho apresentado objetiva a implementação de um esquema para geração de hipóteses sobre o caractere. Uma vez que nenhuma informação sobre o tipo (*font*) utilizado está disponível neste estágio, o sistema deve obrigatoriamente ser tolerante a variações de tipos, sem prejuízo de sua performance.

A solução proposta utiliza a técnica da correlação com máscaras binárias (*Binary Template Matching*), pesquisando os atributos presentes e/ou ausentes de um dado caractere e combinando-os logicamente. O sistema utiliza os módulos de *hardware* do sistema de aquisição e processamento de imagens PAPS

(*Picture Acquisition and Processing System*) controlados por rotinas de *software* que permitem a criação de máscaras bem como a execução das operações de correlação e combinação lógica dos diversos atributos.

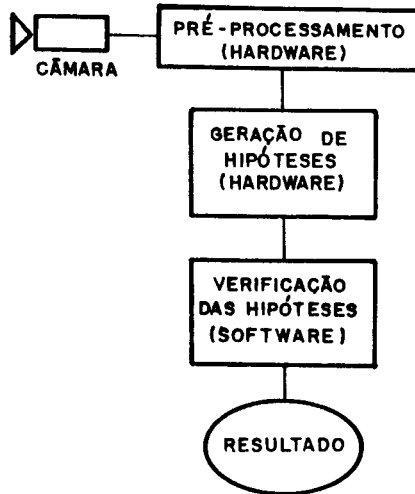


Fig. 1 : Esquema de reconhecimento proposto.

O ambiente no qual o trabalho foi executado consiste de câmara de TV, monitor P&B, PAPS com alguns de seus módulos de *hardware* e o computador principal (SUN 3), no qual se pode executar o *software* de controle e as rotinas de reconhecimento, escritas em linguagem POP-11 (Barrett, Ramsay e Sloman, 1985), utilizando os recursos do ambiente conhecido como "Sistema POPLOG" (*POPLOG System*). Um diagrama básico ilustrando os principais componentes do sistema é mostrado na fig. 2.

O esquema de reconhecimento proposto objetiva obter descrições para todas as letras maiúsculas do alfabeto latino. Estas descrições devem ser válidas, com taxa de confiabilidade semelhante, para qualquer tipo de impressão utilizado em documentos (p. ex. os tipos "*helvetica*", "*courier*" e "*times*" produzidos pela impressora a laser e suas variações - normal e negrito (*boldface*) - ). A tolerância a espaçamento deve ser tal

que permita que os caracteres toquem um no outro. A fusão de dois ou mais caracteres em um não é permitida, uma vez que isto causaria o desaparecimento de algumas propriedades fundamentais dos caracteres.

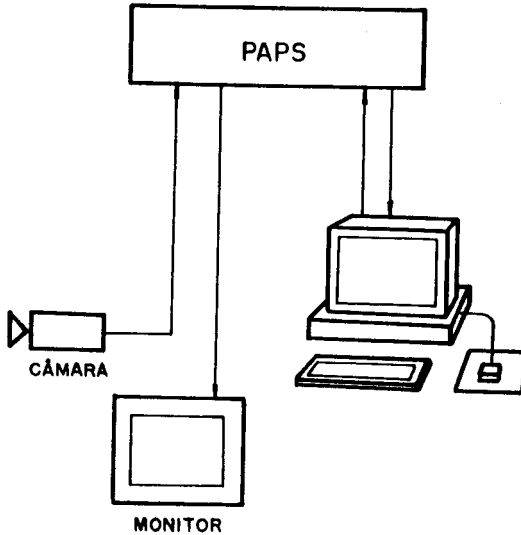


Fig. 2 : O hardware utilizado

## 2. A CORRELAÇÃO COM MÁSCARAS BINÁRIAS APLICADA AO RECONHECIMENTO DE CARACTERES

A correlação com máscaras (*template matching*) é uma técnica através da qual os padrões presentes em uma cena são determinados através da comparação com padrões previamente armazenados. A técnica de correlação com máscaras costuma ser aplicada diretamente a imagens digitalizadas, efetuando uma comparação pixel a pixel com o padrão de referência. As máscaras pertencentes a esta categoria são usualmente denominadas "máscaras a nível de pixel" (*pixel-level templates*).

As máscaras criadas e utilizadas em nosso trabalho assumem que um padrão, em nosso caso um caractere, pode ser dividido em segmentos que o constituem. Estes segmentos podem ser reconhecidos separadamente, correlacionando cada um deles com sua respectiva máscara.

A fim de executar as operações de correlação com máscara binárias, criou-se uma biblioteca de máscaras representativas dos principais atributos considerados significativos na descrição de um caractere. No sistema implementado, estas máscaras são arrays bidimensionais de dimensão menor ou igual a  $12 \times 12$  pixels, onde cada elemento do array pode assumir os valores "preto", "branco" ou "neutro" (*don't care*). A propriedade que se deseja reconhecer na imagem é codificada em "preto" e "branco". Os pixels do array que não especificam uma parte da propriedade são codificados como "neutro" (*don't care*).

Associado a esta máscara está o conceito de limiar (*threshold*). Este limiar pode ser positivo, identificando o número mínimo de pixels na imagem que devem ter o mesmo valor dos pixels "ativos" (preto/branco) da máscara, ou negativo, estabelecendo o número máximo de pixels na imagem que podem diferir dos pixels "ativos" da máscara.

Dentre as máscaras criadas e incorporadas à biblioteca, encontram-se máscaras de tamanho fixo, destinadas à identificação de arcos, cantos e orifícios, e máscaras de tamanho variável, proporcional à altura média dos caracteres presentes na cena, para detecção de bordas horizontais e verticais, bem como identificação de barras verticais, horizontais e diagonais, brancas ou pretas.

### 3. A METODOLOGIA ADOTADA

O trabalho iniciou com um cuidadoso estudo geométrico dos caracteres pertencentes aos tipos "helvetica", "times" e "courier", objetivando identificar a presença ou ausência de certas propriedades e as relações geométricas entre elas, fornecendo uma primeira indicação de quais propriedades seriam úteis na descrição de um determinado caractere. Estas observações qualitativas e quantitativas foram feitas utilizando impressões ampliadas das letras maiúsculas para os 3 tipos de impressão considerados. Os resultados das medições das principais dimensões dos caracteres foram normalizados para uma i-

magem de 100 pixels de altura. Com base na computação da variância destas dimensões é possível ter uma idéia, de antemão, de quais propriedades são mais adequadas a uma descrição capaz de abranger os 3 tipos considerados.

Após o estudo das relações geométricas para todas as letras maiúsculas do alfabeto, decidiu-se dividi-lo em 4 sub-conjuntos, cada um deles contendo letras que possuem propriedades comuns. Os objetivos desta decisão foram o de concentrar esforços na obtenção de boas descrições para caracteres similares e reutilizar parte de uma descrição confiável quando da descrição de outra letra com propriedades semelhantes. Além disto, ela possibilita uma melhor visão de quão bem as descrições são capazes de distinguir caracteres similares, uma vez que algumas das possibilidades mais críticas de reconhecimento errôneo já estão incluídas no mesmo sub-conjunto.

Seguindo este raciocínio, escolheu-se os seguintes sub-conjuntos:

1. 'B', 'E', 'F', 'H', 'P' e 'R': onde uma barra vertical esquerda conectada a uma barra horizontal de altura média é a propriedade comum;

2. 'A', 'K', 'M', 'N', 'V', 'W', 'X', 'Y' e 'Z': por possuírem barras diagonais;

3. 'C', 'D', 'G', 'J', 'O', 'Q', 'S' e 'U': por possuírem arcos;

4. 'I', 'L' e 'T': caracterizados por uma barra vertical e poucas propriedades adicionais.

Para cada um deles foram efetuadas inúmeras experiências visando otimizar as descrições dos caracteres baseadas em seus atributos. O resultado destas experiências é apresentado na forma de descrições cujos passos contêm as máscaras com as quais o caractere deve ser comparado e o deslocamento ( $x_s$ ,  $y_s$ )

do resultado da correlação até um ponto de referência. Cada um destes passos é combinado logicamente com seu antecessor através de uma operação lógica "and". Desta forma, partindo-se de uma imagem contendo quaisquer (combinações de) caracteres, após a última linha da descrição ter sido processada, apenas as ocorrências do caractere desejado terão sido selecionadas.

#### 4. O SISTEMA SOB TESTE

A fim de analisar quantitativamente o desempenho do sistema, decidiu-se testá-lo usando imagens padronizadas contendo todas as possíveis combinações de letras maiúsculas para os 3 tipos, usando as opções "normal" e "bold" para espessura e as opções "spread 0" (espaçamento normal) e "spread -1" (caracteres comprimidos) para espaçamento entre as letras. Foram realizados testes completos nas imagens do tipo "normal spread 0", enquanto que nas imagens "normal spread -1" os valores percentuais foram obtidos por amostragem. Embora esta análise quantitativa não seja completa (exaustiva), ela certamente ilustra os pontos positivos e negativos do sistema proposto bem como subsidia argumentos para conclusões intermediárias que podem ser usadas para direcionar o raciocínio em próximos passos da pesquisa.

Os resultados das experiências estão resumidos na tabela 1. 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).

TABELA 1 Taxas de confiabilidade

CARACTERE	Taxa de Caracteres Não-detetados (%)		Taxa de Reconhecimento Falso (%)	
	"spread 0"	"spread -1"	"spread 0"	"spread -1"
A	0.00	0.00	0.04	0.31
B	0.00	2.27	0.00	0.00
C	0.00	7.69	0.11	0.39
D	0.00	2.13	0.13	0.86
E	0.00	3.45	0.15	0.08
F	0.00	4.00	0.07	0.08
G	0.31	4.35	0.26	1.09
H	1.26	0.00	0.00	0.00
I	15.00	58.33	4.38	0.62
J	14.91	58.14	3.12	2.03
K	0.00	0.00	0.00	0.00
L	0.00	34.00	0.02	0.47
M	14.81	0.00	0.00	2.18
N	0.00	2.04	0.26	2.10
O	0.64	6.82	0.02	1.09
P	1.27	4.17	0.00	0.00
Q	0.63	3.03	0.00	0.47
R	0.61	3.45	0.04	1.33
S	1.27	35.00	0.00	0.00
T	2.53	20.00	0.00	0.08
U	0.00	6.25	0.20	0.55
V	0.00	24.49	0.11	0.00
W	18.47	14.29	0.28	0.23
X	0.00	0.00	0.00	0.70
Y	0.00	2.56	0.43	3.20
Z	0.00	0.00	0.00	0.00

Analisando os resultados obtidos, podemos concluir que:

- Considerando aceitáveis as descrições que apresentam TCN menor que 10% e TRF menor que 0,5%, verifica-se que 10 (dez) caracteres ('A', 'B', 'C', 'E', 'F', 'H', 'K', 'P', 'Q' e 'Z') se enquadram nesta classe. A performance das descrições para as letras 'K' e 'Z' pode ser distinguida devido ao excelente resultado obtido (0,00% para ambas as taxas).

- Ambas as taxas normalmente aumentam ao passar de imagens "spread 0" para "spread -1", o que já era esperado. Desvios nesta tendência podem ser atribuídos ao fato de que os



testes com imagens "spread -1" foram amostrados, enquanto os realizados com imagens "spread 0" foram exaustivos.

- Os resultados pouco expressivos apresentados pelas letras 'I', 'J', 'L' e 'T' já eram esperados. Há uma dificuldade natural em descrever estas letras usando o método proposto devido à falta de atributos passíveis de utilização em suas descrições.

- Para caracteres como 'I', 'J', 'L', 'T', 'V' e 'W' o problema do reconhecimento é agravado pela necessidade de se manter as descrições compatíveis com os 3 tipos de impressão e toma dimensões ainda maiores quando os caracteres estão pouco espaçados entre si.

- Alguns reconhecimentos indesejáveis ocorreram não apenas com caracteres diferentes do que se desejava reconhecer, mas sim com combinações de caracteres semelhantes ao caractere desejado (p. ex. 'VV' e 'W').

## 5. CONCLUSÕES

Com base nos resultados obtidos e comparando o estado final do trabalho com seus objetivos, podemos concluir que:

- A meta geral do trabalho proposto foi alcançada. Implementou-se um esquema de reconhecimento para todas as letras maiúsculas do alfabeto, cujas taxas de confiabilidade podem ser consideradas boas para mais de 80% das descrições quando aplicadas a letras normalmente espaçadas impressas com tipos convencionais na documentação escrita.

- Caracteres pouco espaçados representam um problema técnico bastante conhecido no contexto do OCR. Algoritmos de segmentação ainda não são capazes de resolvê-lo satisfatoriamente (com alto grau de confiabilidade). Neste sentido, os bons resultados apresentados por 10 (dez) letras podem encorajar um aprofundamento nas pesquisas sobre o método proposto.

- Os resultados deste trabalho não são, sob hipótese alguma, definitivos. Eles devem ser interpretados como um estudo quantitativo preliminar da confiabilidade do método. Seu principal significado, portanto, é o de servir como referência para futuras investigações da técnica proposta.

#### REFERÊNCIAS BIBLIOGRÁFICAS

Barrett, R. & Ramsay, A. & Sloman, A., (1985). POP-11: A Practical Language for Artificial Intelligence, Chinchester, Inglaterra, Ellis Horwood Ltd.

Marques Filho, O., (1988). "Character Recognition Using a PAPS". Philips International Institute, Eindhoven, Holanda.

Marques Filho, O., (1989). "Automatic Derivation of Template Descriptions for Character Recognition From a Set of Examples". Philips International Institute, Eindhoven, Holanda.