

Um Sistema de Visão Artificial Incorporando Múltiplas Representações de Forma

PAULO R. MADRUGA
JOSÉ R. A. TORREÃO

Departamento de Informática, Universidade Federal de Pernambuco
50.732-970 Recife, PE, Brasil

Abstract. This work describes the implementation of a flexible 2-D binary vision system - suitable for industrial applications in moderately restricted environments -, which is capable of choosing the best strategy for the recognition and the localization of objects. Our system incorporates two techniques for the representation of shape: one based on boundaries, and the other based on morphological skeletons, and, from an analysis of the input images, it is able to decide which representation is best for each object depicted. The advantages obtained with the conjugation of those two representation strategies have been corroborated by various experiments, both with synthetic and with real imagery.

1 Introdução

O objetivo deste trabalho é descrever a implementação de um sistema flexível de visão computacional binária, capaz de escolher a melhor estratégia para reconhecer e localizar objetos planares, com vistas a aplicações industriais.

A maioria dos sistemas de visão artificial hoje existentes ainda lida com um número pequeno e classes restritas de objetos. Isto porque mesmo a melhor técnica de reconhecimento e esquema de representação não são flexíveis o bastante para suportar grandes mudanças no domínio dos objetos a reconhecer [10].

Para evitar este tipo de restrição, o nosso sistema incorpora duas técnicas distintas para a representação e o reconhecimento de objetos bidimensionais a partir de suas imagens binárias: uma técnica baseada em contornos [7], e uma outra baseada em esqueletos morfológicos [4]. A partir da análise das imagens de entrada, o nosso sistema é capaz de escolher, em uma fase de aprendizado, o esquema de representação mais adequado para cada uma das peças apresentadas, o que facilita e agiliza a posterior identificação e localização das mesmas, na fase de reconhecimento.

As vantagens da conjugação destas duas técnicas de representação foram verificadas em diversos experimentos, tanto com imagens sintéticas quanto com imagens reais.

2 Descrição Geral do Sistema

Faremos aqui uma descrição geral do nosso sistema, que está voltado para o reconhecimento e localização

de partes industriais. A cena é composta pelas peças colocadas em posição arbitrária no campo visual de uma câmera de vídeo. As peças devem ser rígidas e planares, pois estamos tratando com um sistema que não leva em consideração, para fins de reconhecimento, a forma tridimensional das mesmas, mas apenas a sua silhueta. A aquisição de imagens é feita em vários níveis de cinza, efetuando-se posteriormente uma binarização, de modo a trabalharmos apenas com dois níveis de intensidade.

Na operação do sistema, além da etapa de customização - etapa de definição de parâmetros -, distinguem-se duas fases: a fase de aprendizagem e a fase de reconhecimento. A fase de aprendizagem é utilizada para a criação da representação interna das peças a serem reconhecidas, gerando o 'banco de peças' do sistema. Na fase de reconhecimento, cada peça observada sofre um tratamento semelhante ao da fase de aprendizado, para que se determine a sua representação; a seguir é feita uma busca no 'banco' do sistema, confrontando-se as características da peça em questão com aquelas das peças armazenadas, sob a mesma representação, na fase de aprendizado. Em ambas as fases de operação, um gerenciador seleciona a técnica mais adequada para o tratamento de cada peça, objetivando maximizar a eficiência dos processos.

No nosso sistema, empregamos duas técnicas distintas para a representação e o reconhecimento: uma técnica baseada em contornos, e uma outra baseada em esqueletos morfológico. A decisão sobre qual técnica deve ser empregada para uma dada peça é feita pelo gerenciador do sistema, com base na análise das propriedades globais de forma da mes-

ma, tais como número de Euler, compactividade e excentricidade. O primeiro passo, na determinação das formas de representação mais adequadas para as peças apresentadas ao sistema, é a extração dos seus contornos. Após a extração dos contornos, cria-se uma cadeia de vetores elementares correspondentes aos mesmos e, a partir destes, obtêm-se os atributos globais de cada peça observada.

Muitos são os atributos globais que podem ser obtidos a partir das listas de vetores elementares [1]. Estes atributos podem, por exemplo, fazer referência à forma ou ao posicionamento (isto é, localização e orientação) da peça. Para a escolha da representação de uma dada peça, os atributos globais empregados no nosso sistema são: área, perímetro, excentricidade (medida de alongamento da peça), compactividade (i.e., relação entre o perímetro quadrado e a área), e número de Euler (i.e., número de regiões conectadas menos número de buracos). Na fase de reconhecimento de peças representadas via contornos, atributos adicionais são utilizados (inclusive segmentos salientes), conforme descrito na seção seguinte.

No nosso sistema, estes atributos são analisados para cada peça, e a técnica de representação adequada é escolhida de acordo com os critérios expressos na Tabela I, que segue os preceitos estabelecidos por Yang and Sengupta em [10], e cujos valores numéricos foram obtidos empiricamente a partir de nossos experimentos com o conjunto de imagens sintéticas mostrados na Figura 1.

Após a fase de aprendizagem, portanto, o 'banco de peças' do sistema constará de arquivos contendo atributos globais de forma e segmentos salientes, para as peças representadas via contorno, e arquivos contendo esqueletos morfológicos, para as demais.

Dada esta visão geral da operação do nosso sistema, prosseguimos nas próximas seções com a apresentação das estratégias de reconhecimento via contornos e via esqueletos morfológicos.

	Compact.	Excentr.	No. Euler
Contornos	$< 30 ; > 80$	≤ 2	$\neq 1$
Esqueleto	$> 30 \text{ e } < 80$	> 2	$= 1$

Tabela 1

3 Reconhecimento Via Contornos

Uma das estratégias que o nosso sistema de visão computacional utiliza para o reconhecimento de peças, como já mencionado, está baseada em contornos. Nesta estratégia, após extraídos os contornos da imagem capturada, geram-se as cadeias de vetores elementares correspondentes a cada um deles, e a

partir das listas de vetores elementares, o gerenciador obtém os valores dos atributos globais correspondentes à peça observada, e também os seus segmentos salientes [1].

Os segmentos salientes são aqueles que têm como característica fundamental a capacidade de identificar de maneira inequívoca uma peça, sua posição e orientação, dentro de um conjunto de peças dado. É importante observar que o grau de saliência de cada segmento de um dado objeto depende dos outros objetos em questão. Assim, o segmento saliente poderia ser classificado como um atributo local dependente do conjunto de peças tratado [3].

Para a obtenção dos segmentos salientes, tanto no aprendizado como no reconhecimento, o contorno da peça é dividido em segmentos de um comprimento fixo. Faz-se então uma transformação da representação dos contornos das peças, do espaço (x, y) para o espaço Ângulo-Comprimento (Θ, S) , para facilitar a comparação entre os segmentos [7].

Se um segmento tem uma alta similaridade com algum outro, seja na mesma peça ou em outra qualquer, isto é, se seu coeficiente de checagem (C_m), adequadamente definido, resultar maior ou igual a um limiar previamente estabelecido, o segmento em questão não deverá ser escolhido como segmento saliente da peça em análise. Assim, cada segmento saliente deverá ter baixíssima similaridade com os segmentos de todas as outras peças amostradas, e também baixíssima similaridade com segmentos da própria peça. Caso contrário, isto iria dificultar o reconhecimento desta peça ou a determinação da sua orientação.

Se, na fase de aprendizagem, vários forem os segmentos salientes encontrados, estes serão catalogados de acordo com sua saliência, isto é, o segmento com maior saliência é aquele com menor valor de C_m . O de segundo menor valor será o segundo mais saliente, e assim por diante, até um número máximo de segmentos salientes aceitos pelo sistema.

Desta forma, obteremos, para cada peça, uma lista decrescente por saliência dos seus segmentos salientes. Isto facilitará na fase de reconhecimento, quando a busca da peça é feita a partir do segmento mais saliente encontrado na mesma.

No reconhecimento via contornos, os atributos globais da peça observada, e também os seus segmentos salientes, são comparados com aqueles das peças armazenadas, sob esta representação, no banco do sistema. Além dos citados na seção anterior, fazemos uso dos seguintes atributos: número de buracos da peça, momentos de primeira ordem (que permitem a determinação da posição da peça), momentos de segunda ordem (que permitem estimar a

orientação da peça), além da área e do perímetro do maior e menor buracos [1].

Para a determinação dos atributos globais de uma peça a partir da vetorização do seu contorno, foi utilizada uma técnica baseada na decomposição de uma área complexa em figuras mais simples (retângulos e triângulos), cujos atributos são facilmente determinados. Esta técnica está descrita em [5] e foi satisfatoriamente testada como uma alternativa mais eficiente para cálculo dos atributos globais em visão binária (ver [3]).

Na representação via contornos, o reconhecimento se dará quando se observa uma coincidência dos valores dos atributos globais da peça observada com aqueles de uma dada peça do banco do sistema, dentro da faixa de tolerância previamente estabelecida.

4 Reconhecimento Via Esqueletos

A Morfologia Matemática pode ser utilizada para extrair informações a respeito da forma de um objeto numa imagem a partir da sua estrutura interna, ao transformá-lo através de sua interação com outro objeto, chamado 'elemento estruturador', de forma e tamanho mais simples do que o objeto original. Informações sobre forma, tamanho, distribuição espacial, o quão concentrado é o objeto, a suavidade do seu contorno, e a sua orientação, podem ser obtidas pela transformação deste objeto usando diferentes elementos estruturadores [4]. No nosso sistema de visão artificial, um dos métodos de representação que utilizamos se baseia no uso do esqueleto morfológico para a caracterização da estrutura geométrica interna das peças a serem reconhecidas.

Existem diversas formas de esqueletização. Lantuejoul[2] definiu um tipo particular de esqueleto, usando transformações da morfologia matemática, para aplicação na compressão de dados, com a intenção de que o objeto original pudesse ser reconstruído a partir do seu esqueleto. Este tipo de esqueleto também tem aplicações úteis para a representação e o reconhecimento de objetos, como empregado no nosso sistema. Nesta estratégia, o esqueleto de cada peça é armazenado como a sua representação de forma, e, como critério de reconhecimento de uma dada peça, verifica-se quão bem um esqueleto protótipo armazenado se 'encaixa' na peça, efetuando-se a erosão morfológica da mesma com o simétrico do esqueleto do protótipo, e observando-se o resíduo resultante desta operação [8]. Se o objeto analisado é uma cópia exata do protótipo, o resíduo resultante irá consistir de uns poucos pontos, independentemente da posição da peça (pela invariância sob translações da erosão) [9]. Para a determinação

de orientação, os eixos de alongamento, tanto do esqueleto quanto da peça analisada, são obtidos e rotacionados de modo que coincidam as suas orientações. A medida do ângulo de rotação dá a informação de orientação da peça reconhecida.

A esqueletização proposta por Lantuejoul envolve a criação de subconjuntos de esqueletos que podem ser utilizados numa estratégia de reconhecimento multinível. Numa situação real, em que o contorno da peça observada é em geral corrompido por ruído ou falhas na digitalização ou binarização, é aconselhável desprezar os subconjuntos de esqueleto de ordem mais baixa, que correspondem aos detalhes mais finos do contorno. O critério de reconhecimento em cada etapa é estabelecido como um limiar no tamanho do resíduo deixado pela erosão pelos esqueletos parciais: resíduos menores que o limiar escolhido denotam ocorrências do protótipo na imagem.

5 Operação do Sistema

A seguir daremos algumas informações mais detalhadas sobre o funcionamento do nosso sistema, descrevendo tanto o **software** como o **hardware** utilizados.

No nosso sistema, as imagens são capturadas por uma câmera CCD modelo DXC-107 da Sony, que se acopla a uma placa de aquisição e digitalização de imagens ('framegrabber') DT2853, da Data Translation.

Para o tratamento das imagens adquiridas, o sistema apresenta ao usuário um histograma da imagem capturada, e solicita o valor limiar para a binarização. Após a binarização, a imagem será armazenada em formato compactado.

Feita a aquisição das imagens para o aprendizado de todas as peças de interesse, o usuário terá a incumbência de fornecer alguns parâmetros que serão utilizados tanto na fase de aprendizado quanto na de reconhecimento. Tais procedimentos **off-line** são fundamentais para a operação do nosso sistema.

Na fase de aprendizagem, definimos os seguintes parâmetros:

- *Max - linhas* e *Max - colunas* - Números máximos de linhas e colunas da matriz de píxeis.

Utilizamos imagens de 64 linhas \times 64 colunas nos nossos experimentos, porém, como todas as variáveis relacionadas com as peças têm alocação dinâmica de memória, o limite de tamanho para as imagens tratadas pelo nosso sistema é o limite de capacidade de memória da máquina;

- C_m - Coeficiente de checagem de similaridade entre os segmentos. Utilizado tanto para definir os segmentos salientes na fase de aprendizado, quanto

para checar similaridade entre um segmento de peça a reconhecer e um segmento saliente do 'banco de peças', durante a fase de reconhecimento. Aqui já deverá estar incluída uma faixa de tolerância para levar em conta possíveis imperfeições devidas a rotações e falhas de captura.

Nos testes do nosso sistema, estimamos este coeficiente em 90% de similaridade quando queremos reconhecer uma peça, e 10% quando queremos achar os seus segmentos salientes;

- *Tatrib* - Tolerância com relação a flutuações nas medidas de atributos globais (devidas a possíveis alterações nas condições de aquisição), assim como a rotações e seus efeitos quando discretizamos as imagens. A tolerância assumida para teste do nosso sistema foi 10% dos valores corretos dos atributos, como obtidos na fase de aprendizagem

- *Tesq* - Tolerância associada ao reconhecimento de peças via morfologia matemática. Para nossos testes, foi fixada em 20% do total de subesqueletos gerados na aprendizagem;

- *na* - Número de amostras analisadas para cada peça. Caso seja necessário, pode-se capturar um conjunto de imagens de cada peça, a fim de encontrar a melhor imagem para representá-la. Assim, pode-se escolher as melhores amostras das peças na fase de aquisição;

- *ns* - Número de segmentos em que o contorno de uma dada peça será dividido para determinar aqueles segmentos que serão escolhidos como segmentos salientes da mesma;

- *np* - Número de píxeis que compõem um segmento de contorno. Esta informação depende fundamentalmente do tamanho da imagem da peça. Nos nossos experimentos, estabelecemos 10 píxeis por segmento para cada peça;

- *nmax* - Número máximo de segmentos salientes por peça.

Algumas peças terão um número de segmentos salientes maior do que o estabelecido como parâmetro, e neste caso serão considerados os *nmax* segmentos mais salientes, os quais serão armazenados por ordem de saliência. Nós trabalhamos com *nmax* = 5.

- *ne* - Número de píxeis que indica o espaçamento na translação dos segmentos do modelo quando da detecção de segmentos salientes. Normalmente trabalhamos com *ne* da ordem de 3;

Salientamos que, como cada peça é dividida em *ns* segmentos de *np* píxeis, observa-se que, quanto mais píxeis forem utilizados para formar a imagem, maiores poderão ser *ns*, *np* e *ne*. Em compensação, o tempo de identificação das peças aumentará, neste caso, devido à maior complexidade computacional.

Por outro lado, se for utilizado um número muito pequeno de píxeis na formação das imagens, a complexidade computacional, com *ns* e *np* menores, diminui sensivelmente, mas assim também a confiabilidade quanto ao reconhecimento das peças, pois a detecção de similaridade fica prejudicada.

É importante notar que todos os parâmetros devem ser adaptados aos requerimentos do tempo de processamento, do grau de similaridade entre as peças, e da capacidade de armazenamento de dados do equipamento.

Além dos parâmetros acima, cada peça tem a ela associada uma identificação, que é fornecida pelo usuário na fase de aprendizado. Esta identificação poderá estar associada ao nome da peça gravado em um arquivo.

Assim como na fase de aprendizado, os modelos utilizados na fase de reconhecimento do sistema são formados tomando os parâmetros fornecidos na inicialização. Logo, todo o processamento do sistema utilizará um mesmo conjunto de parâmetros. Existirá necessidade de reprogramação do sistema quando houver qualquer alteração nos parâmetros, por exemplo para a inclusão de novas peças no seu 'banco de peças'.

6 Conclusões

Neste trabalho, nós descrevemos a implementação de um sistema de Visão Computacional binária bidimensional que incorpora múltiplos métodos de representação de forma, com vistas à aplicação em ambiente industriais. O nosso propósito foi o de desenvolver um sistema capaz de reconhecer e localizar uma gama variada de peças, utilizando para isto tanto os seus atributos estruturais, obtidos via esqueletização morfológica, quanto os seus contornos e atributos globais.

O nosso sistema foi inicialmente testado em um banco de 101 imagens sintéticas, obtidas via *scanner*, e posteriormente empregado com sucesso na identificação e localização de peças reais de relojoaria a partir de imagens capturadas por uma câmera de vídeo.

7 Referências

- [1] KITCHIN, P.W. & PUGH, A., *Processing of Binary Images*, In: PUGH, A. ed. Robot Vision. Springer Verlag, 1983. p.21-42.
- [2] LANTUEJOL, C., *Issues in Digital Image Processing*, R.M. Haralick and J.C. Simon eds., Sijthoff and Noordoff, The Netherlands, 1980.
- [3] MADRUGA, P.R., *Um Sistema de Visão Binária*

- Incorporando Múltiplas Representações*, Tese de Mestrado, Departamento de Informática, UFPE, Recife, 1992.
- [4] MARAGOS, P. A. & SCHAFER, R. W., *Morphological Skeleton Representation and Coding of Binary Images*, IEEE Trans. ASSP, Vol. ASSP-34, oct. 1986
- [5] PISSARENKO, G. S. & LAKOVLEV A. P. & MATVELEV V. V., *Prontuário de resistência de materiais*, Capítulo 2, Editora Mir, Moscou.
- [6] PUGH, A., Ed., *Robot Vision*, IFS, Ltd., and Springer-Verlag, 1983.
- [7] RILLO, A.H.R.C., *Sistema de Visão Binária com Reconhecimento de Peças parcialmente Oclusas*, Tese de Mestrado, Faculdade de Engenharia Industrial, São Bernardo do Campo, 1989.
- [8] SENGUPTA, S. & YANG, H.S., *Morphological Shape Representation and Recognition of Binary Images*, Proc. SPIE Conf. Intelligent Robots and Computer Vision, Nov. 1987.
- [9] SERRA, J., *Image Analysis and Mathematical Morphology*, Academic Press, 1982.
- [10] YANG, H. S. & SENGUPTA, S., *Intelligent Shape Recognition for Complex Industrial Tasks*, In: International Conference on Systems, Man, and Cybernetics, Alexandria, Virginia, Oct. 1987

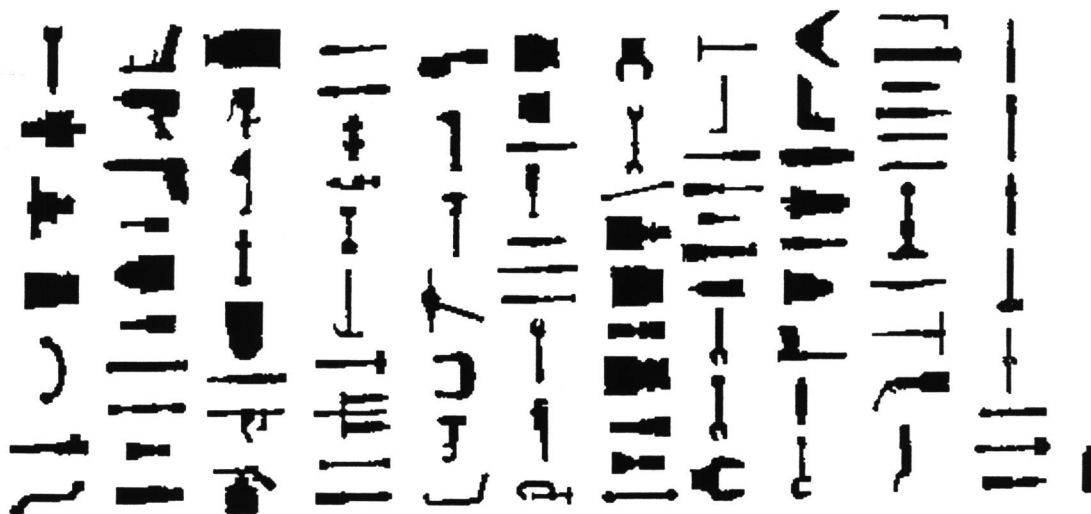


Figura 1

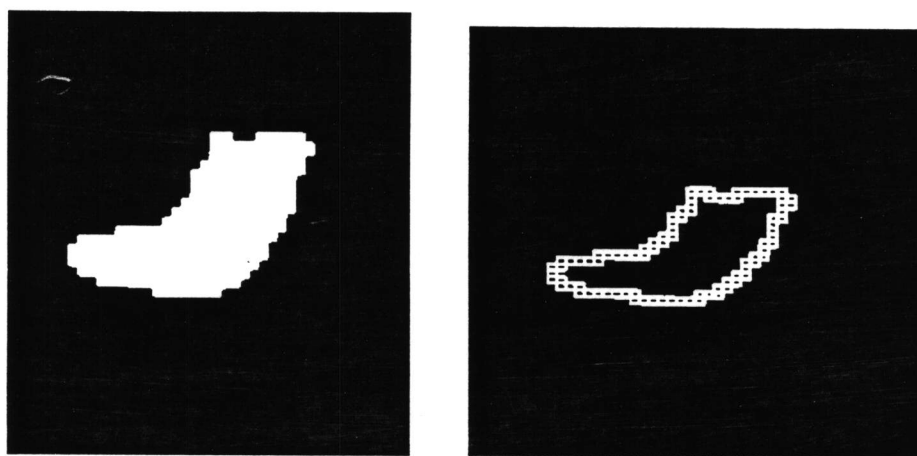


Figura 2: Peça Representada via Contornos



Figura 3: Peça Representada via Esqueletos, com Alguns dos seus Esqueletos Parciais

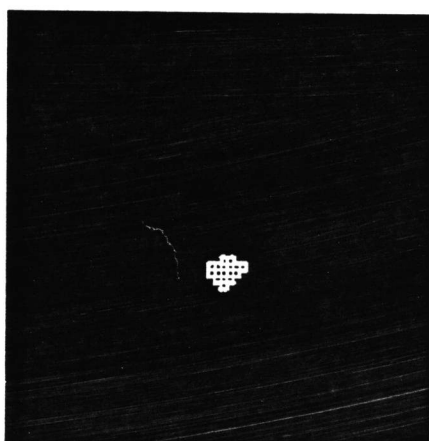


Figura 4: Resíduo da Erosão da Peça da Figura 3 por seu Esqueleto