1. Publicação nº	2. Versão	3. Data	5. Distribuição
INPE-4225-PRE/1102		*Julho, 1987*	☐ Interna ☒ Externa ☐ Restrita

4. Origem	Programa
DPI	*ANIMA*

6. Palavras chaves - selecionadas pelo(s) autor(es)

INTELIGÊNCIA ARTIFICIAL
VISÃO POR COMPUTADOR
COMPREENSÃO DE IMAGENS
VISÃO TRIDIMENSIONAL
SATISFAÇÃO DE RESTRIÇÕES

7. C.D.U.: *681.3.019*

8. Título *INPE-4225-PRE/1102*

ORGANIZAÇÃO DE SISTEMAS DE VISÃO POR COMPUTADOR

10. Páginas: *20*

11. Última página: *19*

12. Revisada por

Nelson D. A. Mascarenhas

9. Autoria

Flávio Roberto Dias Velasco

Assinatura responsável

13. Autorizada por

Dr. Marco Antonio Raupp
Diretor Geral

14. Resumo/Notas

Neste trabalho procuramos mostrar como sistemas de visão por computador são organizados e ilustrar alguns conceitos usados nesta organização. No modelo exposto, a tarefa de visão é dividida em níveis de processamento onde cada nível tem características e preocupações próprias. A grosso modo, três níveis podem ser distinguidos: a visão baixo nível (extração de características), a visão de nível intermediário (segmentação) e visão de alto nível (reconhecimento de cena). Dois problemas de visão são examinados neste trabalho (visão estereoscópica e rotulação de junções) que ilustram os níveis baixo e intermediário da visão. Eles deixam claro o papel dos modelos, exemplificam algumas técnicas de processamento paralelo e cooperativo e satisfação e propagação de restrições comumente usadas.

15. Observações *Trabalho apresentado no III Encontro de Matemática Aplicada e Computacional, de 18 a 20 de maio de 1987, INPE, São José dos Campos, SP.*

ORGANIZAÇÃO DE SISTEMAS DE VISÃO POR COMPUTADOR

1. INTRODUÇÃO

Em inúmeras aplicações, imagens são a forma preferida de aquisição de dados. Essas aplicações incluem, na área médica, as diversas formas de radiografia; na área espacial, o monitoramanto por satélites artificiais da Terra e outros corpos celestes e, na área industrial, a visão robótica usada para a inspeção e montagem.

Computadores digitais podem ser usados de diversas formas como ferramenta auxiliar na análise de imagens. A utilização mais simples consiste no armazenamento, visualização e transmissão de imagens. Ao serem colocadas na forma digital, imagens ficam relativamente imunes à degradação devida ao desgaste, descoloração de pigmentos e outras tantas misérias físico-químicas que costumam afligir os meios analógicos de armazenamento.

A forma digital permite, também, que imagens sejam manipuladas numericamente, ou seja, é possível fazer "contas" com as imagens. Toda uma área de computação (ou da engenharia elétrica, como querem alguns) - o processamento de imagens - dedica-se a teoria e prática dessas manipulações. E possível, por exemplo, aumentar o contraste de imagens, restaurá-las, alterar a geometria, etc, usando somente operações aritméticas adequadas.

Neste trabalho procuramos descrever "visão por computador", que definimos como o uso do computador na extração da informação contida na imagem. Esta informação refere-se aos objetos imageados, ou seja, à cena. Chamaremos de "descrição da cena" a esta informação. Assim, embora a entrada do processo de visão por computador seja uma (ou mais imagens), a saída do processo não é, também, uma imagem (como no caso de processamento de imagens), mas sim uma descrição simbólica e não-pictórica da cena, onde a informação de interesse é tornada explícita.

Dissemos que a entrada do processo de visão é uma imagem (ou conjunto de imagens). Para o computador, uma imagem é,

simplesmente, uma matriz de números (em geral inteiros, **limitados** e não negativos). Para chegar neste formato, a **imag**em original sofre duas transformações: uma chamada **amostra**gem e outra chamada "quantização". Na amostragem são consideradas só algumas posições onde o sinal é medido; na quantização, o valor do sinal, é discretizado num intervalo pré-definido (por exemplo, de 0 a 255). Estes valores discretos são chamados "níveis de cinza". Imagens "coloridas" são, na realidade, a combinação de três imagens diferentes, correspondendo a regiões diferentes do espectro eletromagnético.

Imgens constituem volumes formidáveis de dados. Uma imagem típica de televisão corresponde a 512 por 512 pontos, cada qual com 128 níveis de cinza. Imagens transmitidas pelos satélites da série LANSAT (sensor TM) têm cerca de 36 milhões de pontos (6000 por 6000), cada um com 256 níveis de cinza. Isto ainda é pouco se compararmos ao olho humano, com seus 120 milhões de receptores.

Em aplicações industriais, as imagens tendem a ter menos pontos e menos níveis de cinza (em geral só 2). Por exemplo, no sistema "Vision Module", desenvolvido no Stanford Research Institute, as imagens são binárias e com 128 por 128 pontos (Agin, 1980).

Dissemos, também, que a saída do processo de visão é uma descrição da cena. A forma e o conteúdo desta descrição dependem, obviamente, do propósito do sistema. Num sistema de visão robótica, do tipo "cata e coloca", a descrição pode ser a identidade e atributos de posição e orientação da peça. Num sistema de diagnóstico automático de doenças do pulmão, uma saída possível é a natureza e extensão de problemas respiratórios eventualmente existentes. Finalmente, um sistema de compreensão de desenhos mecânicos, a descrição pode consistir na enumeração das peças existentes no desenho, suas identidades, posições e relacionamento com as outras peças. Um sistema deste tipo teria sua compreensão confirmada se fosse capaz de, por exemplo, gerar outras vistas do mesmo desenho.

Pode-se conjecturar que o processo da visão humana é essencialmente o mesmo que o descrito para a visão por computador. Visão é exercida pelos seres humanos sem nenhum esforço aparente. Entretanto, esta facilidade esconde as imensas dificuldades inerentes ao processo. Só recentemente, quando se procurou reproduzir em computador parte da funcionalidade da visão, é que se deu conta dessas dificuldades. Para os que estudam o problema da visão em computador, é importante ter conhecimento de como funciona a visão humana, pois esta serve como demonstração que a tarefa é possível (apesar de tudo) e como um exemplo de como a tarefa pode ser executada.

Existem, contudo, enormes diferenças, principalmente a nível de implementação física, entre o aparato humano e o computador. Os computadores são seriais e rápidos enquanto que o cérebro humano é lento mas massivamente paralelo. A nível de especificação, ou seja, quais as entradas, quais as saídas e qual informação auxiliar é usada, as diferenças são muito menores; do mesmo modo, ao nível das transformações realizadas e da especificação dos resultados intermediários obtidos, modelos computacionais são úteis na compreensão dos mecanismos da visão humana.

A organização do trabalho é a seguinte: a Secção 2 apresenta os paradigmas em voga para a estruturação de sistemas de visão por computador. O processamento é dividido em diversos passos, também chamados níveis de processamento onde cada nível faz uso de um banco de conhecimento específico para o nível. As secções seguintes tratam, repectivamente, de duas tarefas típicas de dois níveis diferentes de processamento. A Secção 3 trata de visão estereoscópica (tridimensional) e na Secção 4 é vista a rotulação de junções de políedros. Estes problemas foram escolhidos por serem interessantes, representativos e por ilustrarem os conceitos básicos do modelo adotado para visão. Finalmente, a Secção 5 conclui o trabalho com uma visão crítica do modelo adotado.

2. Organização de sistemas de visão por computador.

A extração de informação útil (descrição da cena) a partir de imagens não é uma tarefa fácil. Imagens de cenas naturais são um caos incrível. Os dados disponíveis presentes nas imagens são as intensidades de luz. Estas intensidades dependem dos objetos da cena (o que é bom, pois a informação que se deseja extrair refere-se aos objetos) mas também (e principalmente) da iluminação. Descontinuidades da intensidade de luz podem ser tanto devido à descontinuidade do objeto quanto devido à iluminação e mesmo a oclusão de um objeto por outro. Colocar ordem neste caos certamente não pode ser feito num único passo.

Um paradigma usado para organizar sistemas de visão por computador é quebrar o processamento numa sequência de passos. Neste modelo básico, a entrada de cada passo (ou módulo) é a saída do passo anterior. A entrada do primeiro módulo é a imagem e a saída do último módulo é a descrição desejada. Cada passo de processamento define um "nível". Quanto mais inicial for o passo, mais baixo é dito ser o nível. O que caracteriza os níveis baixos de processamento ("visão de baixo nível") é que eles lidam diretamente com a imagem ou com versões transformadas desta. Os níveis altos manipulam estruturas não-pictóricas que representam, de forma abstrata, os dados de entrada.

Como a estrutura de cada nível é a saída do nível precedente, os passos do processamento podem ser organizados de forma puramente sequencial. No modelo básico é possível distinguir três passos de processamento: extração de características, segmentação e reconhecimento da cena, correspondendo aos níveis baixo, intermediário e alto.

O extrator de características num sistema de visão por computador procura inferir, a partir da imagem, as superfícies dos objetos presentes na cena. A saída deste passo pode ter a forma pictórica, ou seja, uma matriz de valores, onde cada valor representa uma propriedade da superfície (por exemplo, a direção do vetor normal à superfície). A imagem resultante deste processo é chamada "imagem paramétrica", "imagem intrínsica" ou, mesmo, "esboço

2 1/2 D". É comum, também, o extrator produzir não uma única imagem paramétrica, mas várias, cada uma com informação a respeito de uma propriedade de interesse (profundidade, movimento, etc.).

O nível seguinte (ou acima) é o que divide a imagem em objetos que podem ser reconhecidos individualmente (segmentos). A segmentação produz uma lista de segmentos e atributos. A saída neste nível é, portanto, não-pictórica. Num sistema de visão onde os objetos são poliedros (por que não?) a segmentação poderia produzir uma relação de polígonos (lados dos poliedros), atributos que definem os pclígonos (número de lados, tamanho dos lados, e a orientação do plano do polígono).

O passo final do processo de visão combina e agrupa os objetos extraidos do passo anterior em objetos complexos e produz a descrição da cena. No nosso exemplo, a descrição da cena poderia consistir na identificação dos poliedros (cubos, pirâmides, etc...), suas localizações e orientações e relacionamentos existentes entre os diversos poliedros (por exemplo, "pirâmide A está apoiada em cubo B").

Cada nível usa uma base de conhecimento com informação específica para o nível. Esta base de conhecimento é constituída por modelos que descrevem o universo ao nível do processamento. Por exemplo, no nível mais alto, a base de conhecimento contém a descrição dos objetos que se espera encontrar na cena e o relacionamento entre estes objetos.

O conhecimento necessário para o processamento pode ou não ter uma representação própria. Nos níveis mais baixos principalmente, o conhecimento é muitas vezes embutido no próprio processo. Por exemplo, no processo que infere a distância de um objeto a partir da disparidade estereoscópica, é suposto que as superfícies são não-transparentes e que as superfícies são, em geral, suaves, o que implica que distâncias entre pontos vizinhos são próximas. Como veremos na Secção 3, este modelo de superfície está implícito no algoritmo que computa a disparidade para cada ponto

A imagem de entrada, do ponto de vista da informação que contém, é extremamente ambígua e ruidosa. É necessário, para se chegar a uma interpretação única, a exploração das regularidades do mundo externo. O processo de visão é muito mais um processo de inferência do que uma simples redução de dados. Este fato é responsável pelas ilusões óticas no sistema visual humano. O que enxergamos ao olhar uma cena depende não só da imagem formada na retina como das expectativas que o sistema visual tem da cena observada nos diversos níveis de percepção. Ilusões óticas podem ser construídas simplesmente frustrando estas expectativas ou então construindo mais que uma interpretação que as satisfaz.

No processamento de baixo nível, as operações executadas têm, em geral, caracter local. O valor calculado para cada ponto depende só da vizinhança do ponto. Uma consequência deste tipo de processamento é que a operação pode ser aplicada simultaneamente a todos os pontos da imagem, em paralelo. Esta estratégia tem o potencial de permitir que os imensos volumes de dados representados pelas imagens sejam processados com a presteza necessária. A organização matricial de processadores, onde cada processador é responsável por um ponto e está conectado ao ponto e à vizinhança deste, faz com que o tempo de processamento independa do tamanho da imagem. Embora o número de processadores seja enorme (igual ao número de pontos da imagem), cada um dos processadores pode ser extremamente simples.

A segmentação de uma imagem paramétrica pode ser feita de diversas formas. É possível agrupar características semelhantes e vizinhas para formar regiões que correspondem aos segmentos. Linhas, por exemplo, podem ser formadas a partir de pequenos pedaços de borda; planos podem ser construídos agrupando pontos com mesma orientação espacial. Muito mais que a extração de características, a segmentação pode ser guiada por objetivos, ou seja, pelos segmentos que se espera encontrar.

Outra técnica usada na segmentação é o casamento de formas ("template matching"). Nesta técnica, de padrões (formas) são armazenados e comparados com a imagem, um a um. Ao segmento é dada a interpretação que corresponde ao melhor casamento. Interpretações dadas a segmentos próximos podem ser verificadas para a determinação de consistência; as interpretações incompatíveis com todas as interpretações vizinhas são descartadas. Neste caso, à semelhança do que acontece na extração de características, temos um processamento local de compatibilidades. Da mesma forma, este processamento pode ser aplicado em paralelo a todos os segmentos da imagem.

Na segmentação, alguns processos como a de verificação de consistência ("constraint satisfaction") e mesmo o casamento de formas podem ser executados em paralelo. Outros, como agrupamentos de regiões ("region growing"), devem ser feitos sequencialmente, pelo menos em parte.

O processo de reconhecimento da cena pode ser entendido, também, como um processo de casamento. O conhecimento do domínio de aplicação (o modelo do universo em questão) é armazenado como um grafo ("rede semântica"). Os nós da rede representam os objetos da cena e os arcos representam relações entre os objetos. O processo de reconhecimento consiste na extração de um grafo (objetos e relações) da imagem e no casamento do grafo com o grafo armazenado. Este processo é equivalente ao isomorfismo de sub-grafos com algumas complicações adicionais, uma vez que objetos e relações podem faltar (e sobrar) no grafo extraído da imagem.

Ao contrário dos níveis baixo e intermediário, onde o processamento é (ou pode ser) feito em paralelo, no nível mais alto o processamento tem natureza preponderantemente sequencial.

No modelo exposto, o fluxo de controle e de informação entre os diversos níveis de processamento é de baixo para cima ("bottom-up") A interação entre os níveis é minimizada e limita-se às estruturas geradas pelos níveis inferiores e

passadas aos níveis superiores (imagens paramétricas e lista de segmentos). Como um todo, o sistema é dirigido pelos dados, embora dentro de cada nível o processamento possa ser guiado por objetivos. Este modelo é adequado para a apresentação dos conceitos de visão computacional mas não tão apropriado como um modelo operacional geral. Uma série de problemas podem ocorrer. Decisões feitas num nível mais baixo são irreversíveis num nível mais alto. Por exemplo,depois que uma imagem é segmentada, ela não pode ser "re-segmentada" no nível do reconhecimento da cena.

Uma outra estratégia é ter um fluxo de controle de cima para baixo ("top-down").(O fluxo de informação, neste caso, flui nos dois sentidos.) O reconhecimento da cena estabelece os objetivos de mais alto nível (por exemplo: reconheça lados do cubo) até chegar a tarefas simples que podem ser executadas diretamente na imagem (por exemplo: extraia elementos de borda). Caso estas tarefas não possam ser realizadas, os processadores responsáveis por elas transmitem o sinal de fracasso aos níveis superiores. Estes decidem se tentam novas alternativas ou se transmitem que a tarefa não pode ser realizada. Deste modo, a estratégia descendente garante que as interpretações encontradas para os níveis mais baixos são compatíveis com as dos níveis mais altos. Além disso, uma vez decididos os objetivos dos níveis superiores, a busca nos níveis inferiores é feita com uma certa diretividade. Uma estratégia puramente descendente tende, contudo, a ser ineficiente pois, de certa forma equivale a uma enumeração exaustiva das possibilidades.

Alguns pesquisadores têm sugerido que, ao invés de uma estratégia puramente ascendente ou puramente descendente, fosse adotada uma estratégia "oportunística", ou seja, algumas vezes ascendente e outras vezes descendente, dependendo dos dados disponíveis no decorrer do processamento.

3. Visão Estereoscópica.

A visão estereoscópica é o cálculo da distância de objetos através do uso da visão binocular. O fato de termos

dois olhos não serve não só para ter um de reserva no caso da perda de um deles mas também como um instrumento para avaliação de distâncias, ou seja, visão tridimensional. O que torna a visão estereoscópica possível é o fenômeno da paralaxe: um ponto na cena é projetado em posições diferentes das retinas dos dois olhos. Esta diferença (disparidade) varia conforme a distância do objeto. Supondo os olhos fixados no infinito, a disparidade é tanto maior quanto mais próximo estiver o objeto. Conhecida a disparidade é um problema fácil de trigonometria calcular a distância do objeto.

O problema geral da visão estereoscópica pode ser resolvido achando, em um par de imagens, quais pontos correspondem a um mesmo objeto. Antes de se definir um algoritmo para resolver este problema, deve-se determinar qual a informação usada neste processo. Isto equivale a decidir em que nível do modelo deve ser resolvido o problema da estereoscopia. Embora seja viável resolver o problema da visão estereoscópica tanto no nível baixo quanto no intermediário e mesmo no alto, descobriu-se que, no caso da visão humana, o casamento é feito no baixo nível, sem informação adicional dos níveis superiores. Isto é demonstrado através de "estereogramas de pontos aleatórios" inventados por Bela Julesz.

Um estereograma é um par de imagens formadas por pontos pretos e brancos aleatoriamente distribuídos. As imagens são idênticas, com excessão de um quadrado interno que, num dos estereogramas, é deslocado algumas poucas posições para a direita e, no outro, para a esquerda. O espaço resultante do deslocamento é preenchido aleatoriamente com pontos pretos e brancos. Os estereogramas apresentam uma textura uniforme e, sem um escrutínio cuidadoso, são indestinguíveis um do outro. Quando olhados através de um estereoscópio, que faz com que cada olho enxerge uma única imagem do par, o quadrado que foi deslocado parece estar em plano diferente que o resto da imagem. Parecerá estar mais próximo caso o quadrado tenha sido deslocado para o centro (na direção da outra imagem do par); parecerá estar mais longe em caso contrário. Os deslocamentos aparentes são exatamente os que os olhos

perceberiam se o quadrado estivesse em plano diferente do resto do estereograma.

A experiência de Julesz demonstra, que o sistema visual humano é capaz de visão estereoscópica sem necessidade de informação adicional de níveis superiores. Uma vez definidas as saídas (disparidades ponto a ponto), conhecidas as entradas (par de imagens estéreo e nenhuma informação auxiliar) o problema fica definido a nível de especificação. O passo seguinte é definir um algoritmo que resolve o problema e o último passo é definir uma realização (implementação) para o algoritmo, no equipamento ("hardware") disponível (no caso da visão humana o "hardware" consiste de rede de neurônios).

O algoritmo proposto por Marr e Poggio (1976) para visão estéreo baseia-se num modelo simples das superfícies imageadas. Além de resolver o problema, o algoritmo apresentado tem uma série de características que o tornam um modelo interessante para computações realizadas na visão de baixo nível. Estas características são: o algoritmo usa só informação local para calcular a disparidade em cada ponto; o algoritmo é uma soma de processos cooperativos e paralelos; cada processo (que atua em cada ponto) é extremamente simples; as operações são iteradas até convergir (em geral um pequeno número de vezes).

O modelo sobre o qual o algoritmo de Marr e Poggio está baseado é composto de duas restrições naturais:

1. qualquer ponto numa superfície tem uma única posição num dado instante de tempo;

2. variações da distância de pontos pertencentes a uma mesma superfície são suaves; descontinuidades ocorrem infrequentemente e só nas bordas da superfície.

As consequências deste modelo para a imagem são duas. A um ponto da imagem corresponde uma única disparidade, ou seja, cada ponto de uma imagem do par estereoscópico pode ser casado a um única ponto da outra imagem do par. Pontos

vizinhos da imagem têm disparidades semelhantes uma vez que tendem a pertencer a uma mesma superfície.

O algoritmo aceita como entrada um par estereoscópico (I, I') de imagens e produz, como saída, uma coleção de "imagens-disparidade" D(di), i=1, ... n, tal que um ponto (x, y) de D(di) é igual a 1 se a disparidade do ponto (x, y) for igual di e igual a zero caso contrário. Em outras palavras, D(di)(x,y) = 1 se I(x,y) = I(x, y+di) e D(di)(x,y) = 0 em caso contrário. Note que o deslocamento é segundo a coordenada y, ou seja, no mesmo plano em que estão alinhados os planos óticos do par estereoscópico.

O algoritmo é extremamente simples. Valores iniciais são atribuídos às images disparidades D(di). O valor de D(di)(x,y) é modificado dependendo dos valores de D(dj)(x,y), para dj próximo de di (por exemplo, |dj - di| (= 1) e de valores de D(di)(x',y') para (x',y') próximos de (x,y) (por exemplo, |x'- x| + |y' - y| (= 1). O procedimento de atualização dos valores de D(di) é repetido até que um critério de convergência seja satisfeito (por exemplo, o procedimento não resulte em modificações de valor).

Precisamos, pois, definir a atribuição de valores iniciais a D(di)(x,y) e a regra de atualização de valores. A regra deve incorporar as restrições do modelo de superfície adotada. A atribuição de valores iniciais é feita do seguinte modo:

1. D(d)(x,y) = 1 se I(x,y) = I'(x,y+d);

2. D(d)(x,y) = 0 em caso contrário.

Após esta atribuição inicial de valores, poderemos ter D(d)(x,y) = 1 e D(d')(x,y) = 1, para d diferente de d', o que contraria a nossa hipótese de cada ponto ter uma única disparidade. Quanto à segunda restrição, a atribuição dos valores iniciais resulta em superfícies suaves (contínuas) para os pontos que tenham disparidades corretas e valores aleatórios para pontos com disparidades incorretas. A regra

de atualização procura sanar estes dois problemas. Num iteração k, o valor de D(k,di) é dado por:

1. D(0,d)(x,y) = D(d)(x,y);

2. D(k,d)(x,y) = 1 se D(k-1,d) + D(k-1,d)(x',y') - D(k-1,d')(x,y) > T;

onde (x',y') são vizinhos espaciais de (x,y) e d' é uma disparidade próxima de d;

3. D(k,d)(x,y) = 0 em caso contrário.

A regra de atualização favorece a persistência de valores para pontos (x,y) cujos vizinhos (x',y') com a mesma disparidade d tenham o mesmo valor (0 ou 1). O valor T é um limiar que deve ser determinado empiricamente.

O algoritmo proposto resolve satisfatoriamente estereogramas de pontos aleatórios com densidades de pontos que vão de 50% (pontos pretos e brancos equiprováveis) até 10% (pontos pretos são 10% do total de pontos). Este último caso mostra a capacidade de interpolação do algoritmo. Além de resolver o problema, o algoritmo proposto apresenta diversas propriedades semelhantes a propriedades do sistema visual humano tais como a interpolação ("filling-in") e a histerese (persistência da percepção tridimensional mesmo quando as imagens são afastadas uma da outra e, portanto, o casamento não é mais possível).

4. RECONHECIMENTO DE POLIEDROS.

Esta secção apresenta alguns resultados de sistemas destinados à compreensão de cenas simples compostas por poliedros. O problema ilustra uma tarefa típica de visão de nível intermediário. A técnica principal, satisfação de restrições ("constraint satisfaction"), é bastante usada não só na análise de cenas mas também em outros campos da Inteligência Artificial.

O trabalho pioneiro na análise de cenas compostas de poliedros é devido a Roberts (1965). O sistema implementado por Roberts aceita como entrada uma imagem de uma cena e produz, como saída, um desenho da mesma cena, segundo um pcnto de vista diferente. O sistema de Roberts interpreta a cena em termos de (somente) três poliedros. Poliedros mais complexos são decompostos nos três poliedros primitivos transformados por escala, translação e rotação.

A identificação de primitivos é feita através do casamento de características "topológicas" (invariantes sob as transformações) tais como faces, linhas e vértices com as mesmas características extraídas dos primitivos. Uma vez que um primitivo (transformado) é encontrado, ele é retirado da cena e as linhas que se tornaram visiveis são desenhadas. O processamento continua até que todos os os objetos sejam reconhecidos

O sistema desenvolvido por Roberts é baseado em dois princípios:

1. o sistema possui um modelo dos objetos a serem extraídos, armazenados na forma de primitivos;

2. o sistema executa uma pesquisa exaustiva na imagem, procurando características que possam ser provenientes dos primitivos.

O primeiro princípio é um tema constante em todos os níveis de processamento. O segundo princípio pode implicar num tempo excessivo de processamento, caso a imagem sob exame seja complexa.

Huffman e Clowes (1971), trabalhando independentemente, chegaram a uma mesma proposta para evitar a busca exaustiva de todos os possíveis casamentos. A proposta é baseada nos possíveis tipos de linhas e junções que podem aparecer numa imagem de poliedros.

Supondo que não há sombras na cena (o que pode ser conseguido com iluminação conveniente), as linhas que

aparecem na imagem podem ser linhas de contorno ou internas. Estas útimas podem corresponder a concavidades ou convexidades. O primeiro passo é rotular as linhas da imagem; as linhas de contorno são rotuladas com uma seta de tal forma que o objeto fique à direita ao se percorrer a linha na direção indicada pela seta; as linhas côncavas recebem o rótulo "-" e as linhas convexas o rótulo "+". A rotulação das linhas corresponde, na realidade, a uma interpretação do papel das linhas no poliedro e é o que se quer produzir ao fim da análise.

Examinando os possíveis tipos de vértices que podem estar presentes numa imagem de um poliedro onde os vértices são formados por junções de três planos e não mudem de tipo devido a pequenas variações na posição do observador, Huffman e Clowes chegaram à conclusão que só quatro tipos de vértices eram possíveis: L, T, flexa e garfo.

Como cada linha numa imagem pode ter quatro rótulos possíveis, as quatro junções básicas podem ser rotuladas de 208 formas diferentes (3 x 64 + 16). Fisicamente, contudo, só 18 rotulações são possíveis: 3 para flexa, 4 para T, 5 para garfo e 6 para L.

Uma observação adicional é que uma linha só pode ter um único rótulo. Uma linha não pode, por exemplo, começar convexa numa junção e acabar côncava noutra junção. Isto implica que rótulos de junções vizinhas devem ser compatíveis no sentido que a linha comum às duas junções deve ter o mesm rótulo.

Há poliedros que não é possível rotular, ou seja, é impossível achar rótulos para todas as junções de uma forma consistente. Neste caso o objeto é impossível de existir. Pode acontecer, também, de um poliedro admitir mais que uma rotulação. Um exemplo disto é a figura de um cubo, que admite quatro rotulações consistentes: uma que corresponde ao cubo flutuando no ar, e outras três que correspondem ao cubo repousando ou preso a um plano por um de seus lados. Finalmente, é possível ter um poliedro que pode ser rotulado consistentemente embora não possa existir fisicamente.

Waltz (1975) estendeu o trabalho de Huffman e Clowes para considerar também sombras e rachaduras ("craks") formadas pelo justaposição de lados de dois poliedros. Neste caso, há cerca de 50 modos de rotular uma linha e 10 tipos diferentes de juncões. O número de rótulos combinatoriamente possíveis para as juncões cresce enormemente. A junção do tipo garfo passa a ter aproximadamente 125.000 rótulos diferentes. Destes, contudo, só aproximadamente 500 são realizáveis fisicamente, ou seja, 0.4%. Mesmo assim, o número de rótulos por junção faz com que a estratégia de busca exaustiva se torne proibitiva. A solução encontrada po Waltz foi eliminar (filtrar) as rotulacões inconsistentes antes de executar a busca exaustiva. O algoritmo proposto por Waltz apresenta semelhanças interessantes com os algoritmos cooperativos usados na visão de baixo nível, uma vez que usa informação local, é potencialmente paralelo e é iterado até convergir.

Vimos, na discussão sobre o método de Huffman e Clowes, que uma linha não pode mudar de rótulo ao longo de sua extensão. Assim, numa junção qualquer, um rótulo que não é compatível com nenhum rótulo de uma junção vizinha pode ser descartado por ser impossível. Um rótulo é compatível com um rótulo vizinho se a linha comum às duas juncões têm o mesmo rótulo. O algoritmo da filtragem de Waltz é o seguinte:

1. Atribua a todas as juncões todos os rótulos fisicamente possíveis;

2. Para toda junção:

    2.1. para cada junção vizinha examine e elimine os rótulos inconsistentes;

3. Se nenhum rótulo foi eliminado, pare;

4. Vá para 2.

O algoritmo de Waltz produz, para cada junção, um conjunto de rótulos que são localmente consistentes. Isto

não significa que os rótulos são globalmente consistentes. É possível ter, ao final da filtragem, conjuntos de rótulos vázios para todas as juncões, o que caracteriza uma figura impossível. O que acontece mais frequentemente com a aplicação do algoritmo de Waltz é que, ao final, um único rótulo sobrevive para cada junção, ou então muitos poucos rótulos para cada junção.

A idéia da filtragem de Waltz constitui uma ferramenta poderosa e é aplicavel a outros problemas além da rotulação de juncões. Ela permite reduzir drasticamente o número de possibilidades e evita, muitas vezes, a explosão combinatória em um espectro muito grande de problemas.

O algoritmo de Waltz para a rotulação de imagens exige, contudo, que a imagem seja dividida em segmentos que possam receber rótulos individualmente. No caso de imagens de poliedros estes segmentos são as juncões e as linhas. Especialmente em imagens ruidosas como as de cenas naturais, este é um requisito difícil de preencher. Como fazer então neste caso? Uma possibilidade é não restringir a uma única segmentação da imagem mais sim considerar todas as segmentacões "razoáveis". A tarefa de eliminar as segmentacões errôneas e que dão origem, possivelmente, a figuras impossíveis, fica então com o próprio algoritmo de Waltz aplicado a todas as segmentacões. É fácil de ver que este enfoque é impraticável, em geral, devido ao número muito grande de segmentacões possíveis.

Não só a aplicação do algoritmo de Waltz a todas as segmentacões é impraticável (embora possível) como é um modo pouco inteligente de atacar o problema. Se enumerarmos todas as segmentacões, verificaremos que cada segmentação difere muito pouco de várias outras segmentacões. Em suma, se aplicarmos o algoritmo a todas as segmentacões, estaremos na maior parte, fazendo trabalho redundante.

O problema da aplicação do algoritmo de Waltz para imagens com segmentações ambíguas foi estudado em Velasco e Rosenfeld (1979) e Mota e Velasco (1986). Nestes trabalhos

são definidos novos operadores de filtragem que podem ser aplicados a segmentacões ambíguas.

5. CONCLUSÃO

Neste trabalho procuramos mostrar como sistemas de visão por computador são organizados, bem como ilustrar alguns conceitos usados nesta organização. Dissemos que a tarefa de visão é dividida em níveis de processamento onde cada nível tem caracteríticas e preocupacões próprias. A grosso modo, três níveis podem ser distinguidos: a visão baixo nível (extração de característica), a visão de nível intermediário (segmentação) e visão de alto nível (reconhecimento da cena). Dependendo do domínio específico a que o sistema de visão se aplica, a tarefa pode ser dividida em mais níveis e mesmo outras divisões são possíveis (por exemplo, um único nível responsável pela segmentação e reconhecimento).

É importante explicitar, para cada nível, o modelo do universo no qual o processamento está baseado. Embora o domínio possa ser o mesmo, o modelo varia dependendo do nível. No nível mais baixo, o modelo inclue restricões e propriedades das superfícies dos objetos imageados. No nível intermediário, o modelo descreve como os segmentos são compostos pelas características básicas extraídas na visão de baixo nível. Finalmente, no nível mais alto, o modelo indica como os objetos da cena são formados pelos segmentos e como os objetos se relacionam entre si.

Os problemas de visão examinados neste trabalho (visão estereoscópica e rotulação de juncões) ilustram os níveis baixo e intermediário da visão. Eles deixam claro o papel dos modelos, exemplificam algumas técnicas de processamento paralelo e cooperativo, satisfação e propagação de restricões, comumente usadas.

6. BIBLIOGRAFIA

(1) Agin, G.J., "Computer vision systems for industrial inspection and assembly", Computer, vol. 13 (11), pp. 11-20, maio 1980.

(2) Clowes, M. B., "On seeing things", Artificial Intelligence 2, pp. 79-112, 1971.

(3) Huffman, D. A., "Impossible objects as nonsense sentences Machine Intelligence 6, 295-323, 1971.

(6) Grimson, W.E.L.,"A computer implementation of a theory of human stereo vision", MIT A.I. Lab. Memo 510, 1980

(5) Marr, D. e Poggio T., "A computational theory of human stereo vision", Proc. R. Soc. Lond. 204, B. 301-328.

(6) Marr, D. e Poggio, T., "Cooperative computation of stereo disparity", AI Memo 364, Artificial Intelligence Lab., Massachusetts Institute of Technology, Cambridge, 1976.

(7) Mota, F. A. e Velasco, F.R.D., " A method for the analysis of ambiguous segmentations of images", lEEE Trans. on Pattern Analysis and Machine Intelligence, vol. PAMI-8, pp. 778-759, novembro 1986.

(8) Roberts, L. G., "Machine perception of three-dimensional solids", in Optical and electro optical information processing, ed. J.T.Tippett et al., 159-197, Cambridge, MIT Press, 1965.

(9) Velasco, F.R.D., "Computer vision and image understanding", relatório técnico TR-85-09, Wang Institute of Graduate Studies, junho 1985.

(10) Velasco, F.R.D. e Rosenfeld, A., "The application of relaxation to ambiguous waveforms", IEEE Trans. Syst. Man and Cybernetics, vol. SMC-9, pp. 420-428, agosto 1979.

(11) Waltz, D., "Understanding line drawings of scenes with shadows", em The Psychology of Computer Vision,

P.H. Wiston, ed. pp. 19-91, New York, McGraw Hill, 1975.