

1. Publicação nº <i>INPE-4290-PRE/1161</i>	2. Versão	3. Data <i>Agosto 1987</i>	5. Distribuição <input type="checkbox"/> Interna <input checked="" type="checkbox"/> Externa <input type="checkbox"/> Restrita
4. Origem <i>DPI</i>	Programa <i>ANIMA</i>		
6. Palavras chaves - selecionadas pelo(s) autor(es) <i>VISÃO BIOLÓGICA REPRESENTAÇÃO VISUAL INTERNA</i> <i>VISÃO POR COMPUTADOR PROCESSAMENTO VISUAL</i>			
7. C.D.U.: <i>621.376.5</i>			
8. Título <i>ASPECTOS E APLICAÇÕES DO PROCESSAMENTO PRIMÁRIO DA INFORMAÇÃO VISUAL</i>		10. Páginas: <i>11</i>	
		11. Última página: <i>10</i>	
9. Autoria <i>Antonio Miguel Vieira Monteiro</i> <i>Flávio Roberto Dias Velasco</i>		12. Revisada por <i>Nelson Mascarenhas</i> <i>Nelson D. A. Mascarenhas</i>	
Assinatura responsável 		13. Autorizada por  <i>Dr. Marco Antonio Raupp</i> <i>Diretor Geral</i>	
14. Resumo/Notas <i>Este artigo analisa uma proposta de organização para sistemas que tratam dados visuais. Baseado nos sistemas visuais biológicos, o processamento inicial (primário) da informação visual é discutido. O "Esboço Primário" como uma possível representação simbólica de cenas é apresentado e uma possível aplicação em imagens de satélite é analisada.</i>			
15. Observações <i>Trabalho submetido para apresentação no 4º Simpósio Brasileiro de Inteligência Artificial, de 13 a 16 de outubro de 1987, Uberlândia, MG.</i>			

ASPECTOS E APLICAÇÕES DO PROCESSAMENTO PRIMÁRIO DA INFORMAÇÃO VISUAL

Antonio Miguel Vieira Monteiro
Flávio Roberto Dias Velasco
Ministério da Ciência e Tecnologia - MCT
Instituto de Pesquisas Espaciais - INPE
Departamento de Processamento de Imagens - DPI
Caixa Postal 515 - São José dos Campos - SP - Brasil

1 _ INTRODUÇÃO

O processamento de dados visuais é uma atividade antiga (uma das primeiras aplicações de computadores digitais). Estes dados obtidos por um sistema de imageamento qualquer (sensor + digitalizador) são analisados e manipulados por computadores. Ao conjunto de técnicas usadas para processar dados visuais, analisar e manipular imagens por computador, chamamos processamento digital de imagens (PDI). Esta área iniciou-se, praticamente, nos anos 60, com o programa espacial americano, e vem até hoje sendo discutidos e formalizados seus conceitos.

No entanto, para ir do tratamento de imagens para a compreensão de cenas, é necessário entender os mecanismos que tornam possível ao especialista interpretar uma cena. Devemos dotar nossa máquina de uma capacidade visual tal que desempenhe bem a função do especialista.

Vamos procurar entender "habilidade" ou "capacidade" visual como a característica de captar e interpretar a informação visual do mundo transmitida via energia luminosa. Esta habilidade nos permite dizer "como são" e "onde estão" os objetos que vemos no mundo.

Seguindo Marr (1982), caracterizaremos a "visão" como uma tarefa de processamento de informação que pode ser dividida em módulos específicos e separáveis funcionalmente, sendo cada um realizador de uma tarefa de processamento de informação, lidando, então com representações e processos.

Estas idéias conduziram a uma "abordagem representacional" para a "visão" onde, em resumo, a preocupação está em definir o tipo da informação derivada do mundo através do aparelho visual e qual ou quais as representações envolvidas (Brady, 1982; Marr, 1978, 1982). Esta abordagem é um legado de David Marr (1982).

Derivar uma representação de forma em uma única etapa, seria certamente impossível além de inapropriado. A idéia, então, foi criar uma "sequência de representações" (a "abordagem representacional" para a visão). O processamento primário da informação visual (Early vision processing) consistiria basicamente em derivar estas representações. A sequência de representações proposta por Marr (1978), fica assim definida:

(1) IMAGEM INTENSIDADE -->

(2) ESBOÇO PRIMÁRIO (PRIMAL SKETCH) -->

(3) ESBOÇO 2 1/2 DIMENSIONAL -->
(2 1/2 DIMENSIONAL SKETCH)

(4) MODELO DE REPRESENTAÇÃO 3D

A IMAGEM INTENSIDADE é a representação inicial da cena, obtida pelos olhos ou por um dispositivo de imageamento que relaciona valores de intensidade da luz refletida com os pontos formadores das superfícies da cena;

O ESBOÇO PRIMÁRIO (PRIMAL SKETCH) é a primeira representação proposta por Marr. É, já, uma representação baseada em uma descrição simbólica da cena que traz informações explícitas sobre as variações na IMAGEM INTENSIDADE que representam eventos físicos reais na cena, tratando da distribuição geométrica e da organização destes elementos na cena observada. Ela trabalha sobre a imagem bidimensional de entrada, não fazendo considerações sobre a natureza espacial dos objetos.

O ESBOÇO 2 1/2 DIMENSIONAL (2 1/2 DIMENSIONAL SKETCH), conhecido como o 2 1/2 D, é a representação que traz as propriedades das superfícies visíveis da cena sob um sistema de coordenadas centrado no observador, traz informações explícitas sobre a orientação das superfícies, sobre a reflectância, sobre os contornos para eventuais descontinuidades e uma série de informações que podem aqui ser representadas como textura, movimento, transparência, etc, vindas dos módulos especializados existentes no baixo nível da visão.

O MODELO DE REPRESENTAÇÃO TRIDIMENSIONAL (3D) descreve formas e a organização destas no espaço sob um sistema de coordenadas centrado no próprio objeto, e contém as características das superfícies já obtidas no 2 1/2 D.

A proposta deste trabalho é considerar o ESBOÇO PRIMÁRIO tal como definido em Marr (1976, 1982) como base para operações de processamento de imagens,

tais como a segmentação ou detecção de bordas. Para isto, a secção 2 descreve a detecção e representação das variações de intensidade, que são responsáveis pelas primitivas simbólicas da representação da cena. Estas devem demonstrar um alto grau de correspondência com as estruturas físicas da cena observada. Na secção 3, o ESBOÇO PRIMÁRIO BRUTO é descrito, e no passo seguinte, na secção 4, o ESBOÇO PRIMÁRIO COMPLETO é apresentado sumariamente. Finalmente, a secção 5 apresenta a conclusão e um melhor detalhamento de nossa proposta de trabalho.

2 - DETECÇÃO E REPRESENTAÇÃO DAS VARIAÇÕES DE INTENSIDADE

Para gerar o ESBOÇO PRIMÁRIO (PRIMAL SKETCH) três estágios centrais são necessários. O primeiro é detectar e representar as variações de intensidade na matriz de entrada para várias escalas. O segundo é trabalhar na interligação entre estas primitivas geradas para as diversas escalas, criando o ESBOÇO PRIMÁRIO BRUTO (RAW PRIMAL SKETCH). O terceiro é trabalhar os processos seletivos e de grupamento sobre as primitivas representadas, criando o ESBOÇO PRIMÁRIO COMPLETO (FULL PRIMAL SKETCH) ou ESBOÇO PRIMÁRIO (PRIMAL SKETCH).

Os métodos comumente usados para detectar as variações de intensidade incorporam três operações essenciais :

1 - A Imagem Intensidade (I) é "suavizada" por algum operador;

2 - A imagem (I) "suavizada" é diferenciada por um operador derivativo de primeira ou segunda ordem;

3 - A saída da imagem diferenciada é utilizada, buscando-se nela "picos" (extremos positivos ou negativos) ou "cruzamentos por zero" (valores onde existe a transição entre o positivo e negativo).

A "suavização" da imagem tem dois propósitos principais :

(a) Reduzir o efeito do ruído na detecção da variação;

(b) Definir a resolução, ou escala, ou canal, para o qual a variação de intensidade é detectada.

A diferenciação deve ser efetuada, pois variações significativas no perfil de intensidades levam a "picos" na primeira derivada ou a cruzamentos por zero (zero-crossings) na segunda derivada. A importância de trabalhar sobre estas características é que propriedades como tamanho, posição, etc., são mais fáceis de extração. O problema é como "suavizar" a

Imagem Intensidade e que operador diferencial usar. Das investigações psicofísicas e neurofisiológicas muito subsídio foi dado para solução deste problema. Campbell e Robson através de experimentos, concluíram que o caminho da informação visual incluía um conjunto de "canais". Estes eram seletivos para orientação e frequência espacial das variações de intensidade na entrada, o que levou a definição de um filtro gaussiano para realizar a suavização necessária (Marr, 1982; Hildreth, 1983).

A detecção das variações passa, então, a ser o problema de encontrar cruzamentos por zero na segunda derivada da intensidade na direção apropriada. Para sinais unidimensionais, sob certas restrições, um teorema de B. F. Logan prova ser possível reconstruir um sinal a partir de seus cruzamentos por zero (Marr, 1982; Hildreth, 1983). Apesar da dificuldade de expandir o teorema para duas dimensões, o caso de sistemas de visão, a "riqueza" de informação do cruzamento por zero fica estabelecida. Além disso, mecanismos fisiológicos simples podem ser propostos para detecção de cruzamentos por zero pelo aparelho visual humano (Frisby, 1979; Marr, 1982).

Marr e Hildreth (1980) mostraram que, na condição de variação linear, os CZ são detectados e precisamente localizados pelos valores zero do operador Laplaciano (LAPLACIANO), agindo sobre a imagem filtrada com a gaussiana bidimensional (G). Combinando os dois operadores, o Laplaciano e a Gaussiana, temos um operador, LAPLACIANO(G), agindo sobre a imagem $I(x,y)$ e os valores zero em sua saída são representativos das variações de intensidade existentes, se a condição de variação linear é mantida para o perfil de intensidades nesta cena. A saída cruzamento por zero fica grosseiramente definida por :

$$CZ = \text{LAPLACIANO}(G) * I$$

Para representar as variações de intensidade, a saída da convolução (LAPLACIANO(G) * I) gera linhas de CZs que sempre são possíveis de subdividir em pequenos segmentos, de forma que cada um deles obedece aproximadamente a condição de variação linear. A primitiva simbólica primária será então o SEGMENTO CRUZAMENTO POR ZERO (SEGMENT ZERO_CROSSING), SCZ. Para cada CZ da saída de (LAPLACIANO(G) * I) duas propriedades adicionais devem ser calculadas. Estas propriedades são o CONTRASTE (SLOPE) e a ORIENTAÇÃO LOCAL. O CONTRASTE de um certo CZ está relacionado a magnitude e extensão da variação de intensidade local, detectada por aquele CZ. A ORIENTAÇÃO LOCAL é definida pela direção da tangente ao longo do contorno definido pelos CZs da imagem. Para o seu cálculo podemos determinar o gradiente na saída do operador LAPLACIANO(G).

Estes segmentos serão a representação simbólica mais primitiva das variações de intensidade que estão ocorrendo para uma superfície em uma determinada escala. Os segmentos orientados CZ marcam a transição entre a análise analógica e simbólica da cena, e são os precursores das primitivas que descreverão o ESBOÇO PRIMÁRIO BRUTO (EPB).

Uma possível descrição do SCZ seria:

(SEGMENTO (posição x y)
 (orientação o)
 (contraste s))

3 - GERANDO O ESBOÇO PRIMÁRIO BRUTO (EPB)

Depois de detectadas e descritas as variações de intensidade existentes na matriz de entrada, temos um novo problema. As variações foram determinadas para cada canal separadamente, através de operadores LAPLACIANO(G) de tamanhos diferentes. Contudo, a suposição de localização espacial para as variações de intensidade detectadas e representadas para uma dada escala, vai garantir que se um segmento cruzamento por zero (SCZ) está presente para uma escala, então ele deve estar presente em qualquer outra escala de tamanho maior. Isto é, se para uma imagem I filtrada por operadores LAPLACIANO(G) de tamanhos W_1 , W_2 e W_3 onde $W_1 < W_2 < W_3$, um SCZ definido para $I(W_1)$ deve estar presente em $I(W_2)$ e $I(W_3)$, na mesma posição. Se isto não acontece, duas são as razões possíveis :

(1) Duas ou mais variações de intensidade para o canal menor estão interagindo no canal maior;

(2) Dois fenômenos físicos diferentes estão produzindo variações de intensidade no mesmo local do espaço mas em escalas diferentes.

Situações do primeiro tipo podem ser detectadas pela existência nas escalas menores de SCZ muito próximos. As situações do segundo tipo são mais delicadas. Se os SCZs definidos para os dois canais, as duas escalas, têm a mesma posição e orientação, então a localização do segmento pode não conter informação suficiente para que sejam isolados os fenômenos físicos ocasionadores do problema. Na prática, essa situação é muito rara, em geral acontece um deslocamento relativo do segmento de um canal para o outro (Marr, 1982). Simplificadamente podemos dizer que se os SCZs de canais independentes e adjacentes coincidem, então eles devem formar um elemento descritivo único. Caso não coincidam, eles pertencem a superfícies distintas ou são gerados por fenômenos físicos diferentes e não devem ser juntados. Fica claro então que precisaremos de no mínimo dois canais, dois operadores

LAPLACIANO(G), para estabelecer uma realidade física mínima para a representação, isto é, as primitivas geradas a partir da combinação serão fisicamente significativas.

Para compatibilizar a interligação dos canais, é necessário definir configurações especiais de SCZ nas escalas menores. Estas configurações devem tornar explícitas a relação de proximidade deste SCZ neste canais, tornando possível sua identificação para as escalas maiores. Assim, novos elementos descritivos além do SCZ são necessários. Para cada canal, além dos SCZ, alguns tipos de arranjos deste devem se constituir em elementos primitivos de descrição. Estes elementos são os BLOBS e as BARRAS. As TERMINAÇÕES podem ser outro tipo descritivo elementar.

Um BLOB é um pequeno contorno fechado, com uma extensão espacial definida, observado na saída (LAPLACIANO(G) * I). Em vez de descrever o contorno como um SCZ, ele passará a ser descrito como um BLOB.

Uma BARRA consiste de dois SCZ paralelos, separados por uma distância adequadamente definida (Hildreth, 1980), e que tem comprimentos similares. Na noção de paralelismo está embutida uma possibilidade de alguma diferença na orientação dos SCZ, os SCZ não necessitam ter a mesma orientação para serem ditos paralelos (Critérios tanto para limiares em definição de BLOBS quanto de BARRAS podem ser encontrados em Hildreth, 1980). O descriptor associado com esses dois SCZ estará localizado no centro deles.

As TERMINAÇÕES, grosseiramente, seriam os pontos terminais de SCZ. A definição precisa das TERMINAÇÕES é ainda um problema em aberto. É certo, no entanto, que de alguma forma a indicação de descontinuidade existe representada (Hildreth, 1980, 1983).

O ESBOÇO PRIMÁRIO BRUTO (EPB) é, então, a representação definida por este conjunto de elementos simbólicos descritivos, obtidos para as diversas escalas que se combinam. Podemos então ter os seguintes tipos de descritores para as primitivas simbólicas do EPB :

(SEGMENTO (posição x y)
 (orientação local 0)
 (contraste s))

(BLOB (posição x0 y0 x1 y1)
 (contraste médio s)
 (orientação local 0)
 (comprimento l)
 (largura w))

(BARRA (posição x y)
 (contraste s0 s1)
 (orientação média 0)
 (comprimento l)
 (largura w))

4 - O ESBOÇO PRIMÁRIO COMPLETO

Até aqui descrevemos como obter primitivas básicas, que realizam a ponte entre o processamento analógico e o simbólico da informação visual. Estas primitivas tem uma relação direta com as estruturas físicas do mundo visível e, por isso, têm a capacidade de representar as superfícies contidas nele mais fielmente. A organização destas primitivas, determinam estruturas que vão refletir diretamente um "esboço" das superfícies dos objetos na cena. Mas um "esboço", um desenho tracejado da cena, representa adequadamente aquela cena, principalmente quando as primitivas que geram o "esboço" tem significância física relevante no contexto da cena observada. Surpreendentemente, "esboços" e imagens são muito similares, mesmo que aparentemente um "esboço" da cena não pareça ter informações suficientes para sua descrição. As operações locais com as primitivas descritas, montando os elementos da imagem em "contornos", mesmo que estes sejam abstratos (não existam fisicamente), precede o acesso a aplicação de conceitos de nível alto para a identificação da cena. O problema passa a ser, então, o de agrupar símbolos sob alguns critérios. Estes processos de agrupamento agindo nas primitivas geradas pelo EPB determinarão a organização espacial da imagem e produzirão um seu "esboço" que é bastante representativo da realidade física visível.

Os processos de grupamento têm dois requisitos básicos :

(1) Processos seletivos, onde se evita que primitivas não similares se combinem.

(2) Processos de grupamento e discriminação, onde as primitivas similares são combinadas gerando primitivas "maiores", ou ainda construir limites entre

grupos de primitivas que diferem sobre algum critério (o processo discriminativo).

Considerando estes fatos, o produto final do ESBOÇO PRIMÁRIO pode ser uma sequência de "quadros", onde as primitivas organizadas segundo critérios de seleção e de grupamento representam as partes constituintes da cena. Assim, o ESBOÇO PRIMÁRIO é uma segmentação, ainda que primitiva, da cena observada. Além disso ele fornece indícios confiáveis para os processos posteriores de reconhecimento e interpretação dos objetos presentes na cena.

5 _ CONCLUSÃO E PROPOSTA

A abordagem de David Marr, largamente motivada por estudos do sistema visual humano, pode ainda ser sustentada somente por argumentos computacionais. Sendo assim, ela se torna bastante relevante para estudos sobre o processamento primário da informação visual em sistemas de visão por máquina. Como ele tratou principalmente do problema representacional, seus estudos são muito significativos se queremos ir dos sistemas convencionais de processamento digital de imagens para sistemas que compreendem cenas. Isto porque um grande problema nos sistemas convencionais foi, e é, a "pobreza" de informação que a representação inicial das imagens nestes sistemas carrega.

Além disso, quando tratamos com imagens geradas por satélites, os sistemas de imageamento vem apresentando cada vez mais qualidade na "reprodução" das cenas. Isto talvez possa indicar, num tempo não tão remoto, que a análise de cenas usando imagens geradas por satélites, será feita de maneira similar a do especialista, quando este analisa imagens de fotografia aérea, usando somente seu próprio aparato visual.

Para imagens adquiridas através de satélites de recursos naturais, temos estas representadas por matrizes inteiras na memória de computadores digitais, onde cada inteiro simboliza um valor de intensidade de luz refletida para um conjunto de pontos da cena e baseadas em um certo sistema de coordenadas para sua localização espacial, definidas como $I(x,y)$. Uma tarefa comum e de uso diário é a tentativa de um especialista em separar regiões significativas, como áreas de plantação de trigo ou cana, áreas de determinada formação geológica, etc., sobre a imagem recebida, que é o trabalho chamado de SEGMENTAÇÃO e CLASSIFICAÇÃO. Tradicionalmente temos usado métodos estatísticos para estas tarefas, porque nossa descrição da cena é representada por uma Imagem Intensidade do tipo $I(x,y)$, e isto nem sempre ocasiona resultados satisfatórios.

A abordagem de Marr, colocando no centro das tarefas da "visão" no seu nível mais baixo, a descrição das formas e posições dos objetos na cena, é de alguma forma uma SEGMENTAÇÃO primitiva. Em sua abordagem, ele usa uma descrição inicial baseada em informações bidimensionais, o caso de imagens de satélites, para gerar o ESBOÇO PRIMÁRIO, que é uma representação baseada em descrição estrutural e simbólica da cena, gerada a partir da Imagem $I(x,y)$ de entrada independente do domínio de aplicação.

Em função destas hipóteses, nossa proposta é procurar um melhor entendimento destas idéias, adaptando-as aos nossos objetivos e a natureza das imagens que tratamos, já que não objetivamos construir sistemas de visão de propósito geral, e sim resolver problemas de análise e interpretação de cenas geradas por satélites, de maneira automática e confiável. Vamos experimentar esta proposta representativa na tentativa de solucionar problemas como a segmentação e detecção de bordas, ou pelo menos contornar dificuldades relacionadas a eles quando usando a abordagem estatística tradicional.

Além disso, entendendo que o processo de compreensão das cenas não é somente cognitivo, para evoluir nossos sistemas de PDI para sistemas de compreensão, é inevitável que passemos a pensar em descrições estruturais e simbólicas para nossas Imagens.

REFERÊNCIAS BIBLIOGRÁFICAS

Brady, M.; 1982, Computational approaches to image understanding. Computing surveys, vol 14, n1, mar, p. 3 - 71

Frisby, J.P.; 1979, Seeing. Ilusion, brain and mind. Oxford University Press, Oxford.

Hildreth, E.C.; 1980, Implementation of a theory of edge detection. MIT, AI Lab, Tech. Rep 579, apr.

Hildreth, E.C.; 1983, The detection of intensity changes by computer and biological vision systems. Computer vision, graphics, and image processing, vol 22, p. 1 -27

Hildreth, E.C.; 1985, Edge detection. MIT, AI Lab, Tech. Rep 858, sep.

Marr, D; Ullman, S; Poggio, T.; 1976, Bandpass channels, zero-crossings, and early visual information processing. J. Opt. Soc. Am., vol 69, n 6, jun, p. 914 - 916

Marr, D.; 1976, Early processing of visual information. Phil. Trans. R. Soc. London, B - 275, oct, p. 483 - 519

Marr, D.; 1977, Artificial intelligence - A personal view. Artificial Intelligence, vol 9, p. 37 - 48

Marr, D.; 1978, Representing a visual information - a computational approach. Computer vision systems, ed. Hanson, A.R; Riseman, E.M., Academic Press, New York, p. 61 - 80

Marr, D.; 1980, Visual information processing: the structure and creation of visual representations. Phil. Trans. R. Soc. London, B - 290, p. 199 - 218

Marr, D.; Hildreth, E.C.; 1980, Theory of edge detection. Proc. Royal Soc. London, B - 207, p. 187 - 217

Marr, D.; 1982, Vision. W. H. Freeman and company, San Francisco.

Poggio, T.; 1984, Vision by man and machine. Scientific American, vol 250, apr, p. 106 - 116