



PALAVRAS CHAVES/KEY WORDS
PROCESSAMENTO DE IMAGENS
VISÃO POR MÁQUINA - MODELOS DE PERCEPÇÃO VISUAL
PROCESSAMENTO DE INFORMAÇÃO BIOLÓGICA

AUTORES
AUTHORS

AUTORIZADA POR/AUTHORIZED BY
Dr. Marco Antonio Raupp
Diretor Geral

AUTOR RESPONSÁVEL
RESPONSIBLE AUTHOR
Antonio M.V. Monteiro

DISTRIBUIÇÃO/DISTRIBUTION
 INTERNA / INTERNAL
 EXTERNA / EXTERNAL
 RESTRITA / RESTRICTED

REVISADA POR / REVISED BY
Flávio Roberto D. Velasco

CDU/UOC
681.3.019

DATA / DATE
Agosto 1988

TÍTULO/TITLE	PUBLICAÇÃO Nº PUBLICATION NO INPE-4638-TDL/337
	PROCESSAMENTO DE IMAGENS DE SATÉLITE USANDO ESTRUTURAS SIMBÓLICAS DO BAIXO NÍVEL DA VISÃO
AUTORES/AUTHORSHIP	Antonio Miguel Vieira Monteiro

ORIGEM
ORIGIN
PG/DPI

PROJETO
PROJECT
FRH/CAP

Nº DE PAG.
NO OF PAGES
120

ULTIMA PAG.
LAST PAGE
103

VERSÃO
VERSION

Nº DE MAPAS
NO OF MAPS

RESUMO - NOTAS / ABSTRACT - NOTES

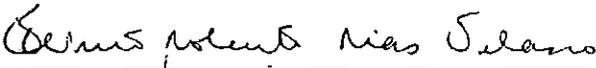
No processamento automático de imagens de satélites, tarefas como a segmentação das imagens em regiões significativas, e mesmo a classificação das cenas, têm sido tradicionalmente executadas, usando como entrada a imagem "bruta" gerada pelos satélites (já consideradas a restauração e operações de eliminação de ruídos). Os resultados obtidos, em muitos casos, não têm sido muito precisos, especialmente quando comparados com a eficiência de fotointerpretes (humanos) para as mesmas tarefas. Os "ganhos" em resolução espacial dos satélites mais recentes têm aumentado ainda mais o problema. O propósito desta dissertação, é a adoção de modelos correntes de percepção visual humana como paradigma de organização para sistemas de processamento de imagens de satélites. Em particular, usamos o ESBOÇO PRIMÁRIO ("PRIMAL SKETCH"), como definido por David Marr, para que seja o ponto inicial para as tarefas de processamento. Esta dissertação mostra alguns experimentos usando o ESBOÇO PRIMÁRIO para executar segmentação em imagens de satélites, discriminação de texturas e aplicado ao mapeamento geológico de regiões.

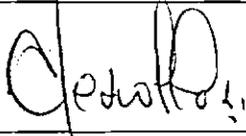
OBSERVAÇÕES / REMARKS

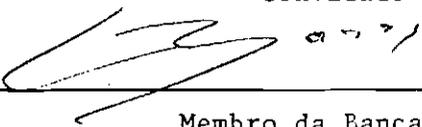
Dissertação de Mestrado em Computação Aplicada aprovada em 10 de junho de 1988.

Aprovada pela Banca Examinadora
em cumprimento a requisito exigido
para a obtenção do Título de Mestre
em Computação Aplicada

Dr. Nelson Delfino d'Ávila Mascarenhas 
Presidente

Dr. Flávio Roberto Dias Velasco 
Orientador

Dr. Clézio Tozzi 
Membro da Banca
-convidado-

Dr. Gérald J. F. Banon 
Membro da Banca

Candidato: Antônio Miguel Vieira Monteiro

São José dos Campos, 10 de junho de 1988

Todos os homens têm, por natureza, o desejo de conhecer; o prazer causado pelas sensações é prova disso, já que, mesmo para além da sua utilidade, elas nos agradam por si mesmas, e, mais do que quaisquer outras, as sensações visuais. Com efeito, não só para agir, como até mesmo quando não nos propomos efetuar qualquer ação, preferimos, por assim dizer, a vista a tudo o resto. A causa disso está em que a vista é, de todos os sentidos, aquele que nos permite adquirir um maior número de conhecimentos, para além de nos permitir descobrir um maior número de diferenças.

Aristóteles. *Metafísica*. Livro I - cap I

ABSTRACT

In automatic satellite image processing, tasks such as segmentations of the image into meaningful objects and even classification have been commonly performed using as input the raw satellite image (except for restoration and other cosmetic operations). In many cases the results obtained have not been very good, specially when compared with the performance of (human) photointerpreters in the same tasks. The increase in spatial resolution of recent satellites has added to the problem rather than improved it. The proposal of this dissertation is the adoption of current models of human visual perception as the paradigm to organize processing systems for satellite images. In particular it is proposed that the "PRIMAL SKETCH", as defined by David Marr, be the starting point of the processing instead of the raw image. This dissertation shows some experiments using the "PRIMAL SKETCH" to perform segmentation of satellite imagery and discusses the applicability of the method.

Para Ari, Emilce, Quequel e Graze, minha família.

AGRADECIMENTOS

Primeiro e principalmente, gostaria de agradecer a Flávio Roberto Dias Velasco, quem através de seu entusiasmo, seu incentivo, seu objetivismo, em resumo, sua efetiva orientação, tornou possível a realização deste trabalho.

A Ricardo Cartaxo, chefe do Departamento de Processamento de Imagens (DPI), quem garantiu a infraestrutura necessária para a conclusão desta dissertação.

A "Seu" Alcir e "Dona" Dulce, meus "pais" em São José dos Campos, para quem qualquer agradecimento ainda seria pouco.

A Janete e a Angela, secretárias do DPI, que têm me "suportado" heroicamente durante estes anos.

A todo pessoal do DPI, pelo ambiente saudável que proporcionam, tornando agradável o dia a dia. Em especial a Luciano, que tem estado comigo desde 1985 e com quem aprendi muito de Processamento de imagens.

Ao Eugênio e ao Fábio, que garantiram o funcionamento da PC-VISION acabando com o estoque de "freon" do Instituto.

Finalizando, ao Celso, pela disposição de ajudar em qualquer momento e, da qual muitas vezes abusei, e ao Cláudio, quem juntamente com Celso, ajudou bastante com o software básico da PC-VISION.

SUMÁRIO

	<u>Pág.</u>
LISTA DE FIGURAS	
<u>CAPÍTULO 1 - INTRODUÇÃO</u>	1
1.1 - Alguns problemas com o processamento digital de imagens de satélites	1
1.2 - Visão e PDI	2
1.3 - A questão da representação	4
1.4 - Sobre este trabalho	6
<u>CAPÍTULO 2 - COMO DEFINIR A INFORMAÇÃO VISUAL</u>	9
2.1 - A visão como uma tarefa de processamento de infor mação	9
2.2 - A abordagem de David Marr para o problema da visão	18
2.3 - O processamento de imagens de satélites usando es truturas simbólicas primárias	25
<u>CAPÍTULO 3 - COMO CRIAR A REPRESENTAÇÃO SIMBÓLICA INI CIAL</u>	27
3.1 - A organização modular do processamento primário da informação visual	27
3.2 - A natureza de representação proposta	29
3.3 - A análise de restrições	31
3.4 - Alguns princípios gerais	37
3.5 - Como detectar e representar as variações de inten sidade	40
3.6 - Como gerar o ESBOÇO PRIMÁRIO BRUTO	54
<u>CAPÍTULO 4 - O ESBOÇO PRIMÁRIO BRUTO PARA O PROCESSAMEN TO DE IMAGENS DE SATÉLITES</u>	61
4.1 - O imageamento por satélite	61
4.2 - O tamanho dos operadores ∇^2G	63
4.3 - As primitivas simbólicas	65
4.3.1 - O segmento cruzamento por zero (SCZ)	66
4.3.2 - A Barra (BARRA)	68
4.4 - A interligação de canais	69

	<u>Pág.</u>
<u>CAPÍTULO 5 - OS EXPERIMENTOS E RESULTADOS OBTIDOS</u>	75
5.1 - 1ª Experiência : A utilização de ESBOÇO PRIMÁRIO para detecção automática de lineamentos geológicos	75
5.2 - 2ª Experiência : A segmentação primitiva de uma cena	80
5.3 - 3ª Experiência : A discriminação de texturas	87
<u>CAPÍTULO 6 - CONCLUSÃO</u>	97
REFERÊNCIAS BIBLIOGRÁFICAS	100

LISTA DE FIGURAS

	<u>Pág.</u>
1.1 - Cena original, esboço e primitivas de Roberts ...	3
2.1 - Distribuição dos fotorreceptores da retina	9
2.2 - Espiral de Fraser	10
2.3 - " Ascending and Descending", litogravura de Escher com detalhe de estrutura fisicamente irrealizável	11
2.4 - Esquema perceptual proposto por Gibson	13
2.5 - Esquema perceptual de Brunswick	15
2.6 - Esquema Gestaltista	17
2.7 - Randons-dot Stereogram's	20
3.1 - Evidência da importância da continuidade	32
3.2 - O triângulo de Kaniza	33
3.3 - Organizações perceptivas de conjunto de pontos ..	36
3.4 - Representação esquemática da idéia do <u>ESBOÇO PRI</u> <u>MÁRIO</u>	39
3.5 - Esquema simplificado para detecção de bordas	40
3.6 - Respostas de células ON-CENTER a estímulos da luz	43
3.7 - Perfil de intensidade e suas derivadas em algumas direções	46
3.8 - Definição de contraste e orientação para slope ..	51
3.9 - (a) Imagem original. (b) Saída do operador $\nabla^2 G$ aplicado a imagem em (a). (c) Mostra os segmentos cruzamento por zero para a escala definida pelo operador em (b)	53
4.1 - Perfil de intensidade isolado	64
4.2 - Em (a) urso de brinquedo, em (b) representação dos SCZ's da cena com $W = 8$ pixels, em (c) as BARRAS.	72
4.2 - Em (d) SCZ's, BARRAS e BORDAS para a cena com dois canais $\nabla^2 G$ de $W = 6$ e 8 pixels interligados, e em (e), parte do arquivo gerado	73
5.1 - (a) área TM estudada. (b) O EPB completo para a cena'	78
5.1 - (c) O mapa de feições para 45 graus obtido manual mente. (d) O mapa obtido automaticamente	79
5.2 - O ESB com $\nabla^2 G$ de $W = 6$ e 8 pixels	81
5.3 - (a) Um SCZ. (b) Uma BARRA. (c) Uma BORDA	82
5.4 - Algumas primitivas no espaço	85

	<u>Pág.</u>
5.5 - (a) Imagem TM original. (b) EPB. (c) Imagem segmentada sobreposta a original	86
5.6 - As imagens originais, retiradas de Brodatz (1968) de (a) veio, (b) rãfia, (c) cortiça e (d) mica ..	89
5.7 - Primitivas para (a) veio, (b) rãfia, (c) cortiça e (d) mica	91
5.8 - Distribuições de TAMANHO e ORIENTAÇÃO para (a) veio e (b) rãfia	92
5.9 - Distribuições de TAMANHO e ORIENTAÇÃO	93
5.10 - Distribuições de TAMANHO e ORIENTAÇÃO	94
5.11 - Descrição com ∇^2G de $W = 8$ pixels para (a) cortiça e (b) mica	94
5.11 - Distribuições de TAMANHO e ORIENTAÇÃO em (c) cortiça e em (d) mica	95

CAPÍTULO 1

INTRODUÇÃO

1.1 - ALGUNS PROBLEMAS COM O PROCESSAMENTO DIGITAL DE IMAGENS DE SATÉLITES

A habilidade de analisar e interpretar a informação visual do mundo tem sido, durante todo o processo evolutivo do homem, uma auxiliar indispensável à sobrevivência e à atitude investigativa sobre seu ambiente. Quando a tecnologia de computadores digitais já gerava máquinas com alguma velocidade e capacidade de armazenamento, uma de suas primeiras aplicações foi o processamento de dados visuais. Estes dados são obtidos através de algum tipo de sistema de imageamento (sensor + digitalizador) que capta a energia refletida pelos objetos de uma dada cena. Esta é medida via uma função bidimensional, $I(x, y)$, sobre uma região determinada, denominada a imagem da cena (Mascarenhas e Velasco, 1984), onde o valor de I para cada (x, y) corresponde a intensidade da luz refletida por um ponto ou conjunto de pontos da cena, e é o que chamamos de intensidade ou brilho para cada elemento da imagem. Assim é que, nos anos 60, com o programa espacial americano e a possibilidade de implementar os algoritmos de processamento de dados visuais em computadores digitais, que, na prática, tem início o que chamamos de Processamento Digital de Imagens (PDI), cuja base estava na análise e manipulação de imagens por computador (Gonzalez e Wintz, 1977; Mascarenhas e Velasco, 1984).

A partir daí, muito se formulou e se fez na área de Processamento de Imagens. Construiu-se uma noção formalizada, sob o ponto de vista matemático, e muitas técnicas e resultados foram e são até hoje discutidos. As técnicas de PDI foram expandidas para o uso em várias aplicações, da medicina aos experimentos militares. Por muito tempo, no

entanto, elas permaneceram compartilhando de idéia comum. Elas se orientavam para métodos capazes, em última análise, de realçar a informação visual obtida pelos dispositivos de imageamento para facilitar a análise e a interpretação visual de uma especialista. Esta orientação estava diretamente ligada aos primeiros momentos do processamento digital de imagens e a tecnologia dos primeiros imageadores.

É fácil observar, quando tratamos com imagens geradas por satélites, que os sistemas de imageamento vêm apresentando cada vez mais qualidade na "reprodução" das cenas, com resoluções espaciais cada vez menores (Veja, por exemplo, a banda pancromática do SPOT, com 10 x 10 m).

Isto indica uma tendência, num tempo não tão remoto, que a análise de cenas usando imagens geradas por satélite, poderá ser feita de maneira semelhante a análise de fotografias aéreas. O especialista analisa a cena usando somente seu próprio aparato visual, sem a necessidade de tratamentos adicionais.

Desta forma os sistemas de processamento de imagens que desenvolvemos devem evoluir. Já não é suficiente "tratar as imagens", é necessário "compreender as cenas", para que possamos interagir com dados de fontes pictóricas ou não. É necessário dotar nossos sistemas de uma maior capacidade de descrição e análise, assegurando precisão a facilidade às decisões do especialista.

1.2 - VISÃO E PDI

Construir sistemas com maior capacidade de descrição das cenas, demanda dotar máquinas com alguma capacidade visual. Dado o interesse industrial e militar na realização de algumas tarefas, uma área de estudos que buscava dar alguma capacidade visual a máquinas, foi naturalmente

suportada, e enquadrada como estudos em "Visão".

Os trabalhos de L.G. Roberts poderiam ser considerados os primeiros em Visão por Máquinas (Barrow e Tenenbaum, 1981). Roberts trabalhou em um domínio específico e simplificado, conhecido como "Block World" (Mundo de blocos), composto de sólidos limitados por polígonos planos regulares denominados poliedros regulares. A partir da cena original, Roberts construía um "esboço" usando as linhas de contorno que definiam o sólido. Este trabalho era facilitado pela distribuição uniforme da luz sobre a cena. Isto determinava que descontinuidades no "brilho" dos pontos que formavam a Imagem da cena demarcassem os limites dos objetos, possibilitando a extração dos "contornos" destes objetos com grande veracidade. Trabalhando sobre este "esboço", ele decompunha a Imagem original em seus elementos primitivos formadores, dentro do contexto definido pelo seu domínio de trabalho (polígonos planos regulares). Veja Fig. 1.1.

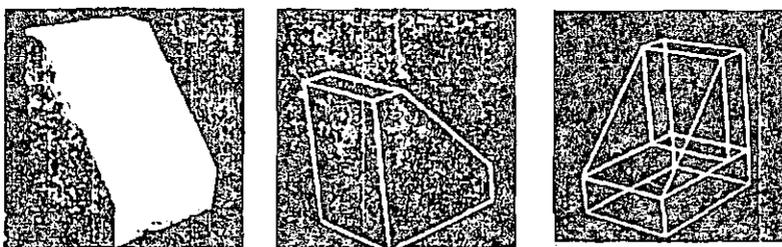


Fig. 1.1 - Cena original, esboço e primitivas de Roberts.

É importante notar que Roberts não trabalhou sobre a cena original e sim sobre um tipo de descrição, representada pelo "esboço", o qual delineava por linhas os "contornos" encontrados para o objeto na cena (Barrow e Tenenbaum, 1981).

Desde Roberts muito se avançou em proposições para Sistemas de Visão. Em comparação com o "Block World"

(domínio restritivo e limitado), o tratamento de cenas naturais, do mundo físico real, era muito mais complexo. Condições de iluminação, textura, movimento, são fortes complicadores para geração de Sistemas de Visão por Máquina, para ambientes irrestritos.

Vários problemas foram resolvidos considerando domínios específicos. No entanto, cada novo problema merecia nova solução. É só em meados dos anos 70, que o foco das pesquisas em sistemas de visão começa a se definir mais estreitamente ligado as habilidades visuais, e não às restrições impostas por um domínio específico, ligado a uma aplicação particular (Brady, 1982). O objeto de estudo passou a ser a exploração do processo visual perceptivo.

Neste trabalho vamos aplicar alguns conceitos e teorias desenvolvidos nos estudos de Visão, que buscam caracterizar nossas "habilidades visuais". Isto nos leva a um primeiro e primordial problema: identificar quem são e de terminar como são representados os dados visuais.

1.3 - A QUESTÃO DA REPRESENTAÇÃO

As imagens adquiridas através de satélites de recursos naturais, chamaremos de IMAGEM INTENSIDADE (In), dado que relacionam valores da luz refletida com os pontos formadores das superfícies da cena. Esta é a primeira "representação" da cena. Tradicionalmente é com esta representação que as tarefas usuais de PDI têm trabalhado.

Uma cena possui duas fontes distintas de complexidade envolvidas no seu processo de geração: a estatística e a estrutural (Tanenbaum et alii, 1980). A complexidade de estatística está ligada às variações de valor da intensidade de luz refletida pelos pontos da cena, variações no "brilho", originadas de fenômenos aleatórios, como alguns

ruídos e mesmo variações devido a fenômenos físicos de grande dificuldade de modelagem detalhada (o que permitiria sua correção), como o "espalhamento atmosférico", que atinge as imagens geradas por satélites. A complexidade estrutural está ligada às variações no "brilho" originadas de algum modo deterministicamente a partir da estrutura física da cena, como a orientação local das superfícies e a reflectância das mesmas.

As cenas reais têm sempre ambos os tipos de complexidade. Historicamente, métodos estatísticos têm sido mais usados. Modelos estatísticos descrevem bem a complexidade estatística, para a complexidade estrutural são necessários modelos estruturais (Tenenbaum et alii, 1980). Uma descrição estrutural para as cenas, envolve representações onde as "estruturas" físicas da cena como orientação, reflectância, distância relativas às superfícies visíveis da cena estejam explicitamente apresentadas.

Vamos olhar o exemplo da Segmentação. Esta é basicamente a tarefa de isolar, numa cena imageada, seus elementos constituintes de significância interpretativa. Tradicionalmente usamos métodos estatísticos para esta tarefa, quase uma consequência natural da representação. Isto nem sempre leva a resultados satisfatórios. A invalidade das suposições estatísticas sobre a realidade física da cena e o uso de testes estatísticos com significância somente estatística, podem levar a resultados sem correção. O mais importante, no entanto, é que a modelagem puramente estatística leva a uma representação inadequada, porque é uma descrição inadequada. Isto dificulta os níveis subsequentes do processo global de compreensão de uma cena. Com isto deslocamos o eixo de atenção do objeto de estudo, a cena, para os sistemas de controle do fluxo da informação, tornando-os mais complexos para adequá-los a uma representação inapropriada.

Isto fica claro no problema da segmentação, com o uso da tecnologia de sistemas especialistas. Têm-se buscado melhorar a tarefa de "segmentar" sofisticando o controle do fluxo da informação. Basicamente, a partir de um conhecimento do domínio armazenado, procura-se dirigir a "segmentação". Os resultados são interessantes e tem-se melhorado a precisão. Entretanto, a opção foi sofisticar o controle porque a representação de trabalho é a IMAGEM INTENSIDADE.

É bom ressaltar que a questão não é negar o papel das técnicas estatísticas na análise de imagens, mas que, se usamos somente esta abordagem, estamos ignorando a natureza física da cena e do processo de imageamento, estaremos, pois, "forçados" a basear nossa interpretação em suposições "ad hoc" e, quase sempre, inválidas fisicamente. O que devemos é procurar uma descrição que nos leva a uma representação tal que as leis e os relacionamentos que governam os objetos no mundo real sejam inerentes à estrutura da representação utilizada e as suas operações de criação (Fischler, 1978). A IMAGEM INTENSIDADE é então uma representação "pobre" da cena.

1.4 - SOBRE ESTE TRABALHO

Torna-se claro que para evoluir nossos sistemas de Processamento Digital de Imagens, uma das tarefas fundamentais é trabalhar uma nova e mais "rica" representação para as cenas. Este trabalho busca experimentar um modelo de representação para dados visuais em tarefas de Processamento Digital de imagens de Satélites.

A preocupação desenvolvida aqui não está centrada na eficiência computacional do modelo implementado (em relação ao tempo de processamento) e sim em verificar a possibilidade de operar com imagens de satélites a partir de

uma descrição simbólica e estrutural da cena, e não mais a partir da IMAGEM INTENSIDADE.

Para isto, a interseção com os estudos em VI
SÃO foi o caminho percorrido. Mais precisamente, com aquele
les estudos onde o objeto é a exploração do processo visual
perceptivo do sistema visual mais completo e complexo que
existe, o sistema visual humano.

CAPÍTULO 2

COMO DEFINIR A INFORMAÇÃO VISUAL

2.1 - A VISÃO COMO UMA TAREFA DE PROCESSAMENTO DE INFORMAÇÃO

Falamos em "habilidade visual". Seria interessante examinar melhor este termo. Vamos procurar entendê-lo como a característica que possuímos de captar a informação visual do mundo transmitida via energia luminosa. Mais precisamente, pelas variações da energia luminosa refletida pelas superfícies dos objetos. Esta habilidade inclui também a análise e interpretação destas variações captadas que nos permite dizer "como são" e "onde estão" os objetos que vemos no mundo. Vamos considerar o sistema visual mais complexo e completo conhecido, que realiza as tarefas de análise e interpretação mais eficiente e confiavelmente, o sistema visual humano (SVH).

No SVH, uma rede de aproximadamente 250 milhões de minúsculos receptores sensíveis a luz (fotorreceptores) estão dispostos em uma superfície interna ao globo ocular. Eles formam o plano visual da retina. Estes fotorreceptores são de dois tipos, cones e bastonetes, e estão distribuídos na retina como mostra a Fig. 2.1.

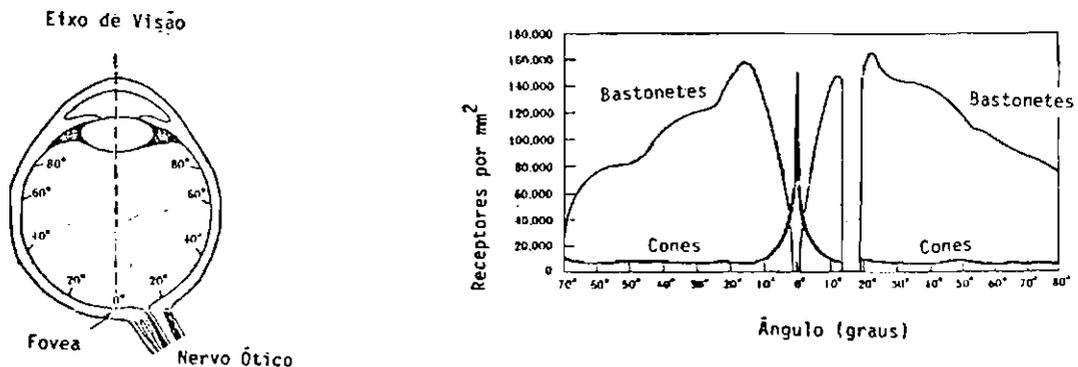


Fig. 2.1 - Distribuição dos fotorreceptores na retina

São estes cones e bastonetes que captam as variações de energia luminosa, provenientes das superfícies dos objetos em uma cena. Estas variações determinam individualmente padrões de comportamento na rede de fotorreceptores. Poderíamos ser tentados a pensar que, reunidos estes padrões no plano visual da retina, ali estaria a imagem que "vemos" do mundo. No entanto, observe a Fig. 2.2, denomina da a espiral de Fraser. Nela o que vemos indica uma espiral. Se acompanhados com um dos dedos as linhas circulares que formam a espiral, é fácil verificar que, na realidade, temos circulos concêntricos organizados de maneira a permitir a "ilusão" de vermos uma espiral.

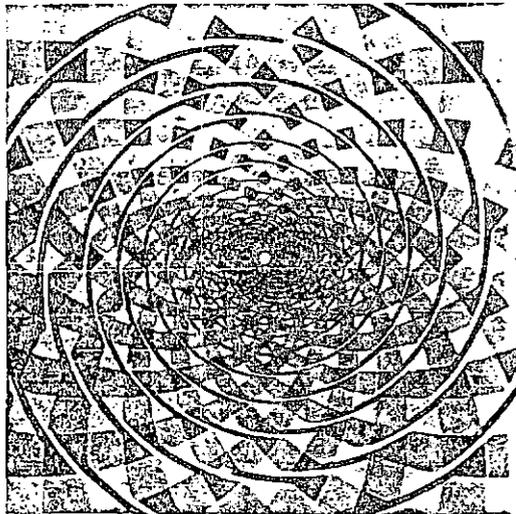


Fig. 2.2 - Espiral de Fraser

Observe agora a Fig. 2.3, uma litogravura de M.C. Escher. Nela existe uma construção tridimensional fisicamente irrealizável. No entanto, ao observador, parece possível aquela cena. Se pegamos elementos individuais que formam a escada, nada existe de problemático com eles. Individualmente eles são fisicamente realizáveis. Montando estes elementos a construção torna-se impossível. Existe um conflito claro entre as características individuais e a interpretação global da cena, pois temos a "ilusão" de que a construção é possível.

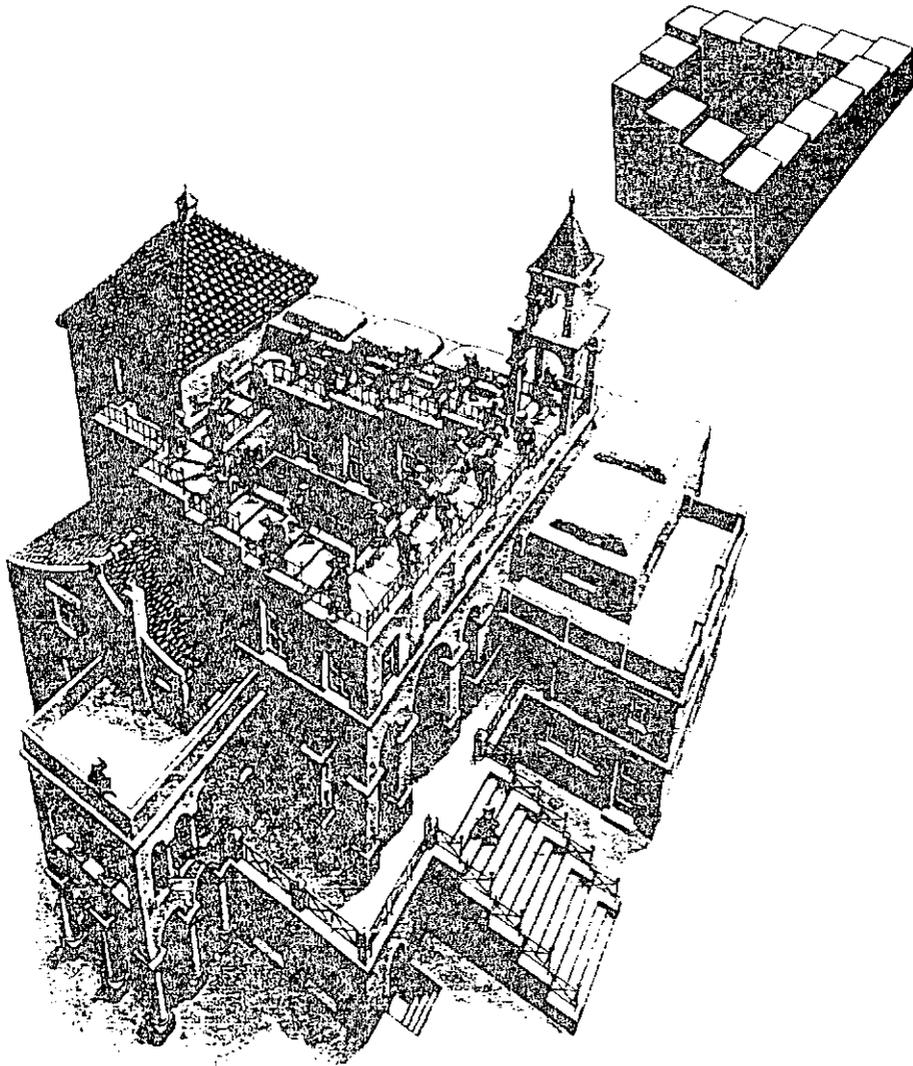


Fig. 2.3 - "Ascending and Descending", litogravura de Escher com detalhe de uma estrutura fisicamente irrealizável.

O certo, então, é que existe uma diferença entre a energia captada pelos fotorreceptores da retina e a informação real sobre a cena. É este o fato que gera um conceito muito importante. São diferenciadas a percepção que temos de uma cena e a cena fisicamente observada. De alguma forma, quando "vemos", estamos percebendo mais do que as va

riações da energia captada. Na verdade, de algum modo as variações de energia estão codificadas, e são estes códigos que nos dão a percepção da cena. Isto conduz a uma proposição mais geral, indentificando como fase primária e básica para realização da experiência visual, a geração de uma descrição simbólica da cena observada. Sobre esta descrição, trabalhariam processos visuais que nos levariam à percepção da cena global. Gerar uma descrição simbólica qualquer sempre implica uma tarefa de processamento de informação. Dessa forma, estão, a experiência visual é inerentemente uma tarefa de processamento de informação.

Diversas propostas, sobre como se organizam as excitações retiniais de forma a gerar descrições consistentes do ambiente, têm sido feitas ao longo dos anos. Algumas destas propostas são de interesse apresentar. Elas são as primeiras tentativas em formalizar uma teoria de percepção visual, que observa, de certa forma, a caracterização simbólica da informação visual.

A teoria da " Ótica ecológica " de Gibson .

O interesse central de Gibson orientava-se para o problema de como um observador conseguia adquirir informação sobre a posição, o tamanho, a forma e os movimentos dos objetos e uma cena. Ele entendia que para todo aspecto ou propriedade do mundo de fenômenos (mundo físico visível), existia um fluxo de energia variável, atuando nos fotorreceptores do aparelho visual, com o qual aquela propriedade ou aquele aspecto poderia corresponder. Um experimento psicofísico poderia ser realizado de forma a comprovar a relação entre a energia captada e a propriedade (Vickers, 1979). Este estímulo teria então uma validade "ecológica" . A percepção para Gibson é, então, explicada em termos simplesmente, das relações estímulo-resposta. O SVH captaria padrões diretamente, sem precisar de mediadores, através

das respostas aos estímulos de luz. De acordo com Gibson, es tão, toda a informação necessária para a descrição da cena, estava contida na imagem gerada pelos padrões. Estes padrões seriam definidos pela iluminação ambiente, ao sensibilizar os fotorreceptores da retina. Veja Fig. 2.4.

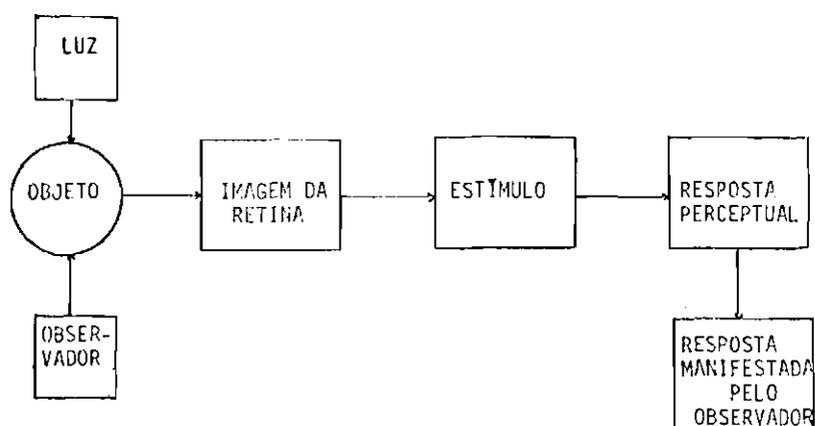


Fig. 2.4 - Esquema perceptual proposto por Gibson

O funcionalismo probabilístico de Brunswich

Na abordagem de Gibson, as fontes de informação visual normalmente operam de modo "geometricamente" pre ciso, determinando a percepção da cena direta e automáticamente. Para Brunswich, a percepção não é uma simples relação estímulo-resposta. Os estímulos são insuficientes, ambíguos, não confiáveis e muitos deles são perdidos. Isto dificulta a percepção direta e automática, como propõe Gibson. Devido a essas características dos estímulos, um sistema sensorial (no caso o visual) é obrigado a agir rapidamente. Para isso, ele é "forçado" a depender de uma variedade de fontes de in for ma ç ã o de maneira parcial, e de informação sensorialmente poten cial mente confusa. Considerado isto, para Brunswich, o SVH providencia a nós, somente "indícios" da natureza do objeto que estamos vendo. Para ele, a função primária do SVH é con du zir a estabilização as relações estímulo-resposta no am

biente, apesar das variações dos estímulos.

A idéia básica do pensamento de Brunswick (Vickers, 1979) é a seguinte. A focalização inicial da cena fornece alguma propriedade física. Esta propriedade, em função de diferenciadas interferências (ruídos), deriva diversos efeitos (não completamente previsíveis) sobre os fotorreceptores. Isto gera uma série de estímulos. Estes estímulos serão considerados como "indícios" que têm um certo grau de confiabilidade em relação à natureza do objeto visto. O grau de confiabilidade é dado pela correspondência entre a propriedade e a validade "ecológica" do estímulo (no sentido de Gibson). Temos uma série de estímulos, e o sistema deve, então, selecionar qual deles procede. Para isso, o sistema vai procurar "pesar" os estímulos. Aquele estímulo, que gerar uma resposta perceptual mais indentificada com a verdade da cena observada, será o determinador da percepção que se tem da cena. Diz-se deste estímulo, que ele tem uma resposta perceptual de alta validade "funcional". Daí o termo funcionalismo probabilístico para a teoria de Brunswick. Variando-se o grau para a utilização do indício" (estímulo) , o sistema procura adquirir uma certa constância na percepção. Isto corresponde a estabilizar as variações dos estímulos, provocadas pelo ambiente físico. Desta forma são necessários processos mediadores (intermediários) entre os estímulos e percepção da cena. São estes processos os responsáveis para a convergência da informação de estímulos em uma única descrição da cena.

Este é um importante achado de Brunswick. A percepção já não é mais um processo direto. É necessário, de alguma forma, manipular informação sensória, ou seja processar informação. Brunswick, entretanto, não detalhou os processos intermediários. Veja Fig. 2.5.

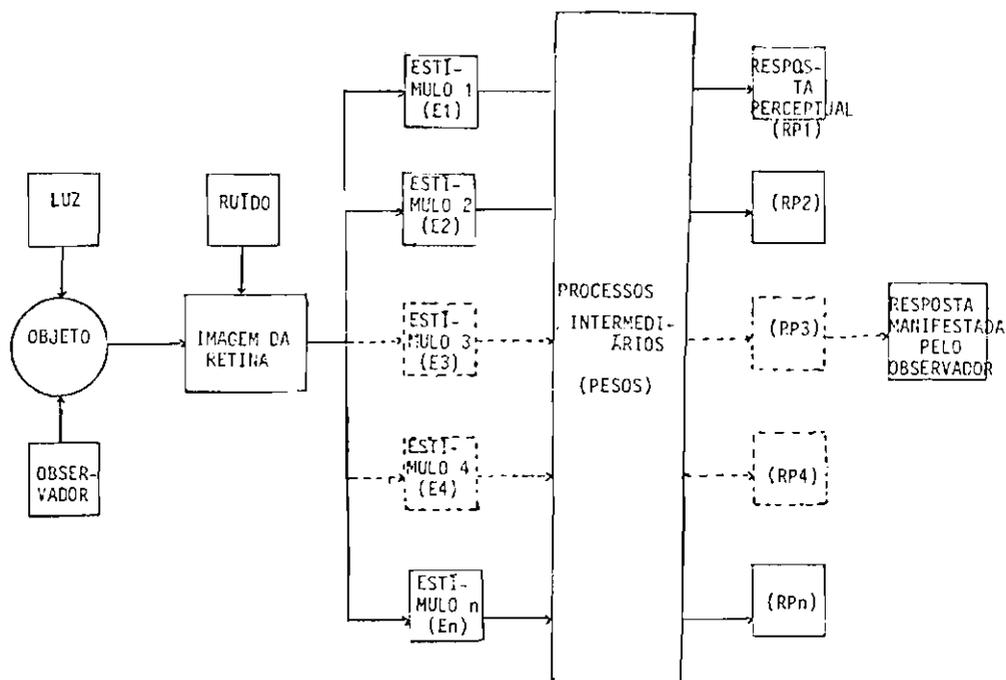


Fig. 2.5 - Esquema perceptual de Brunswick

A teoria Gestaltista

A teoria de Gestalt apareceu na década de 20 com alguns psicólogos alemães, principalmente Wertheimer, Kholer e Koffka. O interesse dos Gestaltistas estava com a estruturação dos estímulos. Seria esta estruturação que nos levaria à percepção das Gestalts (formas) e à compreensão visual da cena.

Como Brunswick eles reconheciam a existência de processos intermediários na percepção. Os Gestaltistas enfatizavam as relações internas na imagem gerada pelos estímulos visuais, e não o papel da experiência ou do aprendizado. Acreditavam que as características intrínsecas aos elementos, determinavam quais estímulos da rede seriam selecionados e agrupados, formando uma estrutura visual perceptiva.

De acordo com os Gestaltistas, a imagem visual seria transformada em um campo de forças tridimensional com certas propriedades dinâmicas. Este sistema buscaria um estado de equilíbrio estacionário, redistribuindo as suas forças componentes. Esta redistribuição seguiria uma tendência a estabilidade, regularidade e simplicidade. Esta suposição se baseava na observação de que as configurações (estruturas) em uma cena, para sua efetiva percepção, caminhavam sempre na direção da forma que melhor demonstrava estas características. Esta foi chamada a Lei de Pregnância para a organização dos estímulos visuais.

Os Gestaltistas avançam mais um passo no esclarecimento do processo visual perceptivo. Os processos intermediários começam a ser desvendados. Mesmo que não tenham conseguido formalizar precisamente suas "leis" e "princípios" de organização para os dados perceptuais, determinando seus ambientes de aplicação e suas possíveis restrições, eles abriram uma grande porta. As formas (Gestalts), podiam ser obtidas já no nível mais básico ou processamento visual. Não nos preocupando com aspectos particulares, a percepção de configurações numa cena seria direta, estabelecida por leis e princípios que regulam as relações entre estímulos. Veja Fig. 2.6.

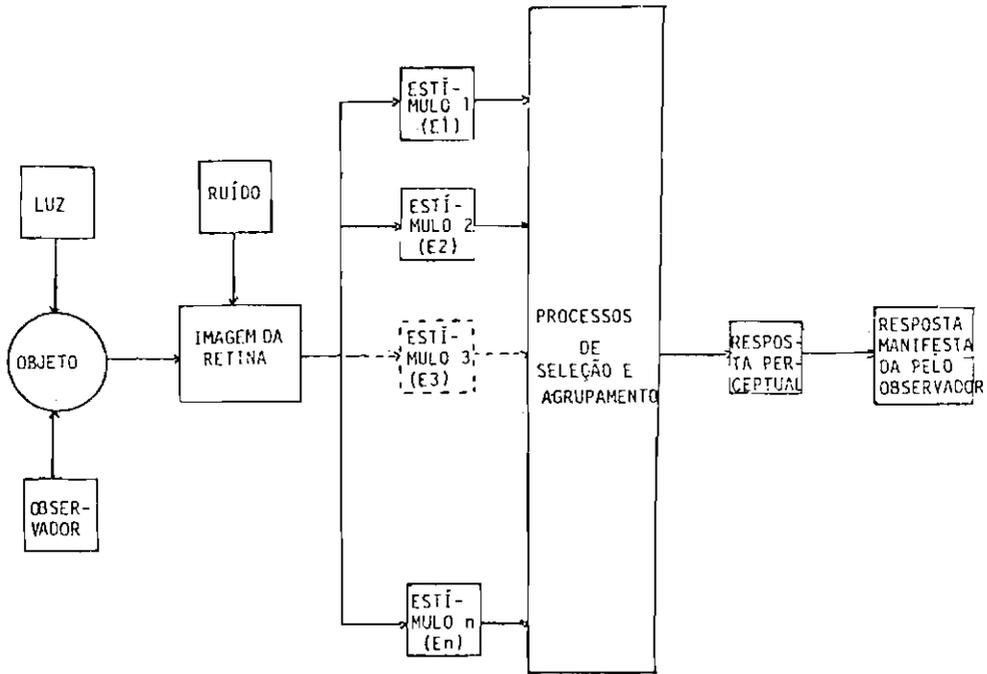


Fig. 2.6 - Esquema Gestaltista

Outras propostas são, ainda, os sistemas detectores de funções (Feature-detecting systems) baseados em evidências neurofisiológicas, e a abordagem do problema da percepção segundo a teoria da informação de Shannon e Weaver (Vickers, 1979). No entanto, creio que Gibson, Brunswick e a Gestalt determinam bem o procedimento de compreender a visão como uma tarefa de processamento de informação simbólica. Gibson, com os padrões, determina a utilização de símbolos no processo descritivo das cenas, Brunswick, com seu modelo, determina que uma cena só é percebida, quando manipulamos estes símbolos de alguma. E a Gestalt propõem como manipular símbolos para uma efetiva percepção da cena.

Para os sistemas atuais de processo digital de imagens (PDI) e sistemas de visão por máquina, a descrição inicial da cena, é, em geral, uma matriz, onde cada ponto representa o valor da intensidade de luz refletida por uma á

rea da superfície imageada. Uma matriz de "brilhos" é a entrada. Isto é uma descrição muito "pobre" para as complexas tarefas que estes sistemas devem executar.

Se queremos melhorar nossos sistemas, é preciso ter bem claro que a base da experiência visual é uma descrição da cena, que seja "rica" em explicitar propriedades daquela cena. No SVH, os estímulos com validade "ecológica" determinam um tipo de correspondência física com a cena. Assim, para os sistemas automatizados devemos pensar que tipo de descrição devemos usar, para basear nossas propostas de representação para os dados visuais. Algumas idéias têm sido formuladas nesta direção. Vamos analisar uma delas, que trabalha modelos computacionais baseando-se na realidade biológica do aparato visual humano.

2.2 - A ABORDAGEM DE DAVID MARR PARA O PROBLEMA DA VISÃO

A abordagem de Marr pode ser inserida no que se tem chamado de "teorias de representação da mente", enfoque que concebe a "mente" como tendo acesso a sistemas internos de representação para os dados perceptivos (Marr, 1982). Para ele, o problema é a *percepção visual, o objeto de seu estudo é a visão*", identificada primeira e principalmente como uma complexa tarefa de processamento de informação, que não pode ser pensada como um único processo. Considerando nossa capacidade de estimar "como são" e "onde estão" os objetos no mundo, de alguma forma devemos ser capazes de representar esta informação. O estudo da "visão" não deveria somente extrair de uma cena observada os vários aspectos do mundo por nós utilizados mas, principalmente, *inquirir sobre a natureza da representação interna da informação visual e dos processos pelos quais "capturamos" essa informação*, tornando-a disponível para nossas decisões no nível interpretativo.

Procurando respostas nesta direção, sua abordagem liga-se aos sistemas visuais biológicos, em particular ao sistema visual humano (SVH).

Desta forma, os diversos trabalhos em fisiologia, neurofisiologia, e psicofisiologia ligados à percepção visual influenciaram suas idéias. Estes trabalhos começavam a demonstrar que questões antes encaminhadas sob um enfoque psicológico, que atentava para um "conhecimento", poderiam ser clareadas e até completamente explicadas, por experimentos em neurofisiologia ou fisiologia, por exemplo. Estes experimentos indicavam um trabalho a nível dos componentes do sistema visual sobre a informação "capturada" pelos fotorreceptores da retina, já no nível mais baixo, primário, da experiência visual. No entanto, os sucessos destes achados experimentais acabaram por levar a uma abordagem restritiva, na medida em que se passou a buscar neurônios que "respondiam" pela realização de uma determinada tarefa. Um exemplo, seriam os "Bug detectors" de Barlow (Marr, 1982) que seriam células no sistema visual de uma rã que conseguiram "avaliar" a presença ou não de um inseto no seu campo visual, executando assim, primitivamente, uma tarefa de reconhecimento. Muito se procurou por "detectores" de diversos tipos e para diversas tarefas, mas o que se fazia era descrever o comportamento das células, não explicar este comportamento.

Não sabendo *como* nem *porque* um dado trabalho era realizado pelas células, não se podia caracterizar seu aspecto funcional, através, por exemplo, de um programa de computador e, assim, dotar máquinas com aquela capacidade visual.

Outros experimentos, como os "Random-dot Stereogram's" de Bela Julesz (Marr, 1978, 1982; Poggio, 1984), demonstravam que determinadas análises, como a estereoscopia, ou seja, a habilidade de "ver" o relevo das cenas, podia

ser realizada somente com base no deslocamento dos pontos , sem nenhum outro nível adicional de informação (Poggio, 1984), o que parecia demonstrar a existência de módulos in dependentes trabalhando na percepção (Marr, 1976). De certa forma, este conceito facilitaria o estudo, pois seria pos sível pensar a percepção visual como a tarefa realizada por módulos especializados, os quais poderiam ser tratados sepa radamente. Veja Fig. 2.7.

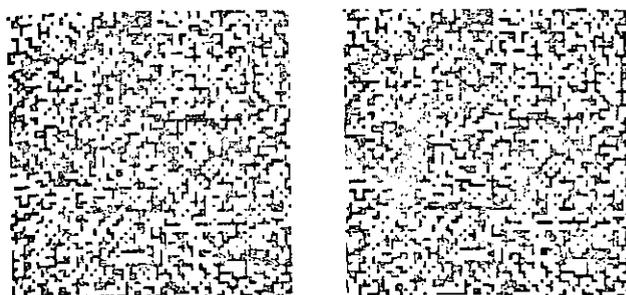


Fig. 2.7 - Random-dot Stereogram's. Se olhar com seu olho direito a imagem da esquerda e, com seu olho esquerdo a imagem da direita, durante um dado intervalo de tempo, poderá perceber um quadrado em profundidade, formado pela fusão das duas imagens.

Esta idéia é fundamental para a compreensão dos estudos de Marr, pois ele podia, então, caracterizar a "visão" como uma tarefa de processamento de informação que podia 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.

Abordando sistemas que processam informação , Marr (1977, 1982) identificava três níveis de caracterização necessários. Primeiramente, o sistema avaliado, no caso o módulo escolhido, deveria ser identificado por uma Teoria Computacional, ou seja, a tarefa realizada pelo módulo deveria ser caracterizada através do mapeamento de um tipo de informação

para outro, isto é, uma transformação de uma dada representação para outra. Deviam, ainda, serem apresentadas as propriedades deste mapeamento, e demonstrando, neste nível, a adequação da representação e das propriedades para a tarefa.

Na definição das propriedades reside a etapa crítica da Teoria Computacional do módulo. Restrições adicionais deverão ser feitas para que se limite o resultado encaminhando-o para uma solução única, sem ambiguidades. As restrições não podem estar ligadas às situações de contexto das aplicações, já que o sistema visual humano é de propósito geral. Deste modo, as restrições deveriam ser gerais abordando as restrições físicas do mundo visível. Isto determina que o conhecimento delas é de valor permanente e pode ser acumulado e, sendo uma vez isolado, será a única informação adicional incorporada ao projeto dos processos que realizam a tarefa.

Estando precisamente definida a "Computação" que o módulo realiza, uma segunda fase seria a preocupação com a representação e com o algoritmo para aquela Teoria. Qual ou quais os tipos de representações serão usadas para entrada e saída do módulo e qual ou quais algoritmos serão usados para as transformações, são as preocupações deste nível. Para uma mesma Teoria Computacional podemos ter diversas possíveis representações para o tipo de descrição proposto por ela. A escolha adequada é fundamental para simplificar o processo de transformação e os níveis subsequentes do processamento.

Após estas duas etapas, podemos partir para a Implementação, o nível onde as representações e algoritmos escolhidos serão realizados fisicamente. Novamente diversas e diferentes implementações podem ser feitas, no entanto a tarefa realizada, descrita pela Teoria Computacional, é a mesma.

Dessa forma, Marr conseguia se abstrair dos neurônios e dos "chips" dos computadores digitais e pensava somente sobre a tarefa a ser realizada. Pensava no tipo de informação a ser mapeada e os tipos de restrições gerais que poderiam ser usadas. Tentava esclarecer os módulos básicos, de mais baixo nível, do sistema visual. Dava-lhes uma Teoria consistente sob o ponto de vista explicativo de *o que e porque* acontece, e procura entender que informação se fazia necessária representar para a tarefa estudada.

A preocupação constante de Marr (1978) estava com o tipo de informação internamente utilizado e, portanto, com a representação desta informação. Devido a diversidade de propósitos dos diferenciados sistemas visuais biológicos, seria improvável que estes usassem a mesma representação para a informação visual. O provável seria que cada um tivesse uma ou mais representações, que seriam as mais adequadas a realização de seus propósitos. Estas idéias conduziram a "*abordagem representacional*" para a "visão" onde, em resumo, a preocupação estaria 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) .

Os contornos definitivos para o trabalho de Marr foram fortemente influenciados por trabalhos em neurologia clínica. Em especial os trabalhos de Elizabeth Warrington e Taylor (Marr, 1982), que descreviam as capacidades e limitações de pacientes que haviam sofrido lesões cerebrais, e que sugeriam duas coisas:

- a - A representação de formas de objetos e a representação dos seus atributos como nome, uso, finalidade estavam armazenadas em lugares diferentes e portanto eram tarefas diferentes;
- b - A experiência visual podia determinar uma descrição

interna da forma do objeto, mesmo quando este objeto, mesmo quando este objeto não era reconhecido, isto é, não havia uma compreensão de seus atributos.

Evidenciou-se, para Marr, que a crença corrente dos pesquisadores em Visão por Máquinas, de que a tarefa de reconhecimento requeria algum tipo de conhecimento de ordem superior, não estava de todo correta. Mesmo sob difíceis condições de iluminação e observação da cena, formas podiam ser determinadas somente pelos processos básicos do aparelho visual. Isto o conduziu a pensar o propósito da "visão" como sendo o de construir uma descrição das formas e das posições (arranjo no espaço) dos objetos de uma cena observada. A "visão" não fazia somente isso, ela também poderia dizer sobre a iluminação, sobre a cor, sobre o movimento dos objetos e outras situações, mas pelo menos no primeiro nível descritivo do aparelho visual o trabalho central seria derivar uma representação das formas dos objetos e de sua organização na cena (Marr, 1976, 1978, 1982; Marr e Hildreth, 1980).

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 proposta por Marr ficaria assim de finida (Marr, 1978):

- (1) IMAGEM INTENSIDADE
 !
 !
- (2) ESBOÇO PRIMÁRIO
 (PRIMAL SKETCH)

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

!
!

(4) MODELO DE REPRESENTAÇÃO 3D onde,

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 deste elementos na cena observada. Ela trabalha sobre a imagem bidimensional de entrada, não fazendo considerações sobre a natureza espacial dos objetos. É a fonte de informação mais básica para todos os posteriores estágios do processamento visual.

O *ESBOÇO 2.1/2 DIMENSIONAL (2 1/2 DIMENSIONAL SKETCH)*, conhecido como 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, no nível da retina.

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 caracte

rísticas das superfícies já obtidas no 2.1/2 D.

É importante observar que, nesta abordagem, a interpretação 3D da cena é construída hierarquicamente através de informações disponíveis nas representações propostas, que se constroem extraído informações de uma imagem bidimensional. Dessa forma, muito trabalho é feito no baixo nível do processamento da informação, sem que se recorra a grau de "conhecimento" ou a relacionamentos com situações já encontradas. O processamento se dá sobre a *IMAGEM INTENSIDADE* de entrada. Dela se extrai informação, em geral suficiente, para que os processos conhecidos como de mais alto nível (o reconhecimento de objetos numa cena, por exemplo) não sejam complicados, uma vez que a informação necessária as suas decisões está acessível e pronta para ser usada.

2.3 - O PROCESSAMENTO DE IMAGENS DE SATÉLITES USANDO ESTRUTURAS SIMBÓLICAS PRIMÁRIAS

Até este ponto estivemos conduzindo uma análise sobre os trabalhos em visão ligados a capacitação de máquinas para a realização de efetiva experiência visual. Ao fazer a intersecção destes estudos em Visão com os Sistemas de Processamento de Imagens tradicionais, mostra-se com clareza os limites destes quando usando uma representação baseada em uma descrição da cena que considera simplesmente a intensidade da luz refletida pelos objetos que é "capturada" pelos dispositivos de imageamento.

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 ten

tativa 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 maneira uma SEGMENTAÇÃO primitiva da cena. 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.

O ESBOÇO PRIMÁRIO é a fonte de informação básica para os processamentos posteriores. Vamos trabalhar com esta hipótese para realizar tarefas de processamento de imagens de satélites de recursos naturais. Não trataremos com a IMAGEM INTENSIDADE e sim com uma nova representação da cena obtida a partir dela. Para isso, geramos o modelo proposto com uma implementação visando esta aplicação. Nossa imagem de trabalho passa a ser um conjunto de estruturas simbólicas primárias. Estas estruturas são as obtidas no nível mais baixo do processamento dos dados visuais.

Em vista desta disposição, no capítulo 3 discutimos aspectos gerais da proposta de Marr, para no capítulo 4 estabelecer a representação que usamos neste trabalho.

CAPÍTULO 3

COMO CRIAR A REPRESENTAÇÃO SIMBÓLICA INICIAL

3.1 - A ORGANIZAÇÃO MODULAR DO PROCESSAMENTO PRIMÁRIO DA INFORMAÇÃO VISUAL

O sistema visual humano (SVH) é uma "máquina" que realiza descrições eficientes (confiáveis) e rápidas do mundo físico visível a partir de imagens deste. Nos sistemas biológicos como nos sistemas de visão por máquina, os processos de descrição iniciam com uma matriz grande e volumosa, onde cada ponto define uma medida da quantidade de luz refletida pelas superfícies dos objetos em um dado ambiente. Em ambos os sistemas, objetiva-se recuperar as propriedades físicas destas superfícies, o que levaria a identificação dos objetos e a possibilidade de compreensão da cena observada. Este objetivo é amplo e complexo para que seja atingido numa única etapa, num processo. É necessário, então, compreender que processos estão sendo executados, que tipo de informação manipulam e como eles se organizam para realizar o processamento global de informação visual.

Para os sistemas de visão por máquina propostos nos anos 60 e, até aproximadamente 1973, prevalecia a escola Top-Down de organização. Basicamente a idéia era que considerado o sistema visual como um "reconhecedor" (identificador dos objetos de uma cena e de seus relacionamentos espaciais), ele deveria ser organizado de maneira que o objeto a ser reconhecido, dirigia o fluxo das informações. Esta informação contextual, interferia diretamente sobre os processos operando no nível da matriz de entrada (a IMAGEM INTENSIDADE).

Evidências psicofísicas e neurofisiológicas levaram a uma reação a escola Top-Down (Marr, 1976; 1978;

1982). A tarefa da visão era complexa e agigantada, então um sistema de visão deveria se organizar como uma coleção de pequenas sub-partes, módulos visuais. Estes seriam aproximadamente independentes, organizando-se com um alto grau de funcionalidade (no sentido de executar uma tarefa específica) e com pequena interação entre módulos vizinhos. Isto garantia que adaptações e/ou modificações pudessem existir sem alterações profundas na estrutura do sistema. Observe que o SVH desenvolveu-se no curso da evolução natural.

Pensando em sistemas de visão por máquina, esta característica atende a requisitos importantes para um bom projeto de software. "Quebrar" um grande processamento em funções que teriam um alto grau de coesão e baixo acoplamento, facilitaria expansão, manutenção, implementação e testes do sistema.

No entanto, existe ainda um problema. Como evidenciar este isolamento relativo dos processos visuais? Isto é, como identificar os módulos visuais? A idéia foi procurar identificar funções realizadas isoladamente pelo aparato visual, identificar e tentar isolar nossas "habilidades visuais". Por exemplo, a percepção de movimento, a percepção de forma, de textura, de profundidade, etc. Identificado um possível processo, parte-se para verificar, apoiando-se em uma proposição teórica formulada a partir de evidências experimentais, se retirado todo tipo de informação, exceto aquela suposta suficiente para o processo realizar sua função, ele ainda trabalha bem. Acontecendo isto, este processo requisita pouca ou nenhuma interação com outras partes do processamento global da informação visual. Desta forma ele pode ser entendido como realizando sua tarefa em relativo isolamento. Sendo isolável, constituir-se-a num módulo visual, o qual age prioritariamente sobre a informação disponível na cena observada.

Estas idéias conduziram ao princípio da organização modular para sistemas de visão (Marr, 1976; 1982). Nele os módulos visuais identificados se agrupam no que ficou estabelecido como sendo o processamento de nível baixo da informação visual (Early vision processing, Marr, 1976). É este o processamento mais básico, o PROCESSAMENTO VISUAL PRIMÁRIO, na tarefa complexa de compreensão global de uma cena.

3.2 - A NATUREZA DA REPRESENTAÇÃO PROPOSTA

Os módulos visuais devem interagir de alguma maneira, facilitando e conduzindo as tarefas ditas de nível mais alto nos sistemas de visão. A saída do sistema visual é uma descrição única da cena e, portanto, as saídas dos módulos têm que interagir. É necessário notar que estes módulos aqui tratados operam sobre a informação disponível e suas saídas deveriam refletir um estreito relacionamento com as propriedades das superfícies dos objetos do mundo físico visível. Sendo assim, definir as entradas e saídas dos módulos é estudo fundamental para seu entendimento.

A Imagem Intensidade é invariavelmente nosso padrão de entrada. A sua estrutura reflete confusamente as propriedades físicas das superfícies na cena. Ela é uma entrada analógica e a variação nas medidas que ela representa compõem a informação que ela carrega. Temos trabalhado na idéia de visão como um processamento de informação simbólica. Sendo assim, naturalmente deve existir um passo onde o dado analógico de entrada é diretamente convertido em informação simbólica. O que deve ser observado, porém, é que não pode haver perda de informação que seja relevante na execução de tarefas posteriores no momento desta conversão. Essencialmente, devemos ter uma entrada que represente os elementos na Imagem Intensidade que tem propriedades físicas identificáveis e ocupem uma localização definida nas super

fícies dos objetos pertencentes a cena.

Nossa representação inicial será um conjunto de primitivas simbólicas, que podem ser derivadas direta e repetidamente da Imagem Intensidade. Estas primitivas tem uma lista de atributos físicos como orientação, tamanho, posição, etc., que as determinam unicamente. A parte crítica é que *estas primitivas devem responder pela realidade física das superfícies observadas*. A importância desta ligação é que, para determinação das estruturas das superfícies, e finalmente dos objetos da cena, inferências serão feitas sobre os dados simbólicos, descritores da Imagem Intensidade de entrada. Estas inferências serão tão mais confiáveis quanto mais significativas fisicamente forem as primitivas geradas.

É o mundo físico visível quem deve determinar algumas características de natureza geral, refletindo o comportamento das medidas das intensidades da luz projetada na retina ou captada por uma câmera de TV. Estas características vão levar a indícios determinantes para a definição da estrutura da representação primária da tarefa visual global. O conhecimento a ser agregado no processamento da informação visual neste nível, não é, então, aquele conseguido com a familiaridade com a cena, mas sim suposições gerais de como se comportam as superfícies dos objetos no mundo físico visível quando refletindo luz.

Antes de estabelecer quais serão as primitivas simbólicas e a representação visual primária da informação é necessário analisar o comportamento das superfícies no mundo físico segundo algumas definições. Estas definições, de alguma forma geram restrições. Analisar estas possibilidades restritivas é ponto crítico da definição dos processos visuais e da representação proposta por Marr. As restrições devem ser naturais, refletindo uma situação permanente. Desta maneira, estando relacionadas com o mundo físico visível,

o conhecimento proporcionado por elas é acumulável e pode portanto se integrar à definição do processo ou da representação que este gera ou modifica.

3.3 - A ANÁLISE DE RESTRIÇÕES

Quando investigando a natureza da representação primária da informação visual, o conjunto de restrições necessárias a sua determinação deve estar ligado à natureza geral das funções de reflectância das superfícies no mundo. Orientando-se por este raciocínio, a idéia é buscar suposições que delineam o comportamento das superfícies, ou melhor, das funções de reflectância das superfícies visíveis. Essas suposições serão nossas *asserções físicas básicas*, que podem gerar *restrições de caráter geral para os processos e representações visuais*. Vamos apresentá-las e discutí-las.

1 - *A existência de superfícies*

O mundo físico visível pode ser conderado como composto de superfícies "macias", que têm uma dada função de reflectância (aquela que define como a energia luminosa reflete sobre a superfície) cuja estrutura espacial pode ser elaborada. Para o mundo visível, *é próprio então se falar em superfícies ao invés de objetos, nos níveis iniciais do processamento visual*.

2 - *A organização hierárquica*

A estrutura espacial relativa à função de reflectância de cada superfície é definível e tem uma certa organização. Tomemos um exemplo. Olhando o pêlo de uma onça, seria possível detectar vários níveis organizativos. Estando bastante próximo dela, o pêlo seria identificado por cada fio que o compõem, onde cada um teria sua própria função de reflectância e seria o nível mais fino. Um nível acima,

os fios se agrupam em conjuntos, definidos uma nova estrutura homogênea. Nesta nova estrutura, não mais cada fio é identificado e sim o conjunto destes situados próximos e paralelos um do outro e que tem características comuns (como a cor, por exemplo). Este conjunto determina uma nova superfície que tem sua própria função de reflectância. Subindo mais de nível, se afastando da onça, grupos demarcando formas de terminam nossa visão de pêlo da onça. Não se identifica mais cada fio, nem pequenos conjuntos deles, e sim conjuntos maiores, gerados a partir dos menores, que determinam novas superfícies. A função de reflectância se organiza hierarquicamente, de forma que a cada nível de sua estruturação as informações importantes no mundo visível passam a existir.

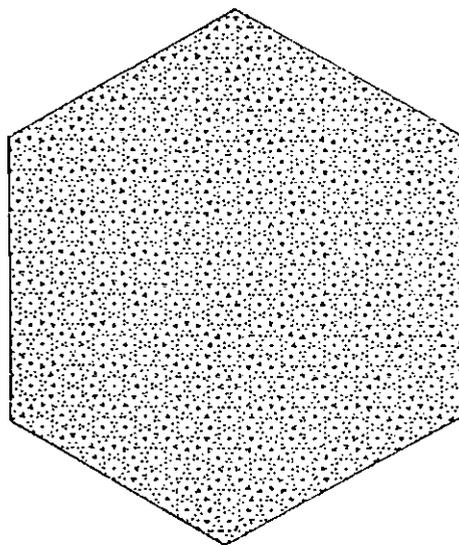


Fig. 3.1 - Evidência da importância da continuidade

5 - Continuidade de descontinuidades

As superfícies existem no mundo (segundo a primeira asserção) e, logo, elas têm limites. Isto leva a descontinuidades em profundidade ou na orientação das superfícies. Assumindo que estas descontinuidades são "suaves"

para quase todo ponto ao longo das superfícies analisadas , geramos uma restrição que trabalha para nossa compreensão do mecanismo de contornos subjetivos. A Fig. 3.2 exemplifica essa idéia. Aparentemente o sistema visual toma as variações em profundidade como indício visual tão importante, que as faz explícitas, mesmo onde não há evidência visual direta de sua existência. Como isso não temos uma "quebra" brusca das superfícies, pois representadas estas possíveis descontinuidades elas são um tipo de primitiva e obedecem a *asserção de continuidade*.

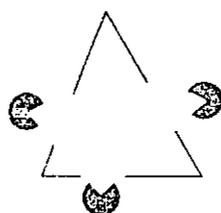


Fig. 3.2 - O triângulo de Kaniza

Sendo assim, as primitivas simbólicas a serem usadas na representação inicial, devem ser capazes de apresentar explicitamente as características físicas importantes para os diversos níveis hierárquicos que compõem a superfície visível final. Isto significa que quando vemos o pêlo da onça como o "vemos", as informações dos níveis inferiores foram capturadas e contribuíram decisivamente para a percepção que estamos tendo. Desta forma, a estrutura espacial da função de reflectância das superfícies é sempre gerada por um número de diferentes processos, e cada um deles opera para diferentes níveis, isto é, diferentes escalas. As primitivas simbólicas da representação inicial devem captar essa noção de construção hierárquica da cena.

3 - *Similaridade*

A função de reflectância é relativamente independente para cada escala. Como vimos, a organização hierárquica tende a definir novas superfícies pelo relacionamento dos elementos encontrados numa hierarquia inferior. É natural, então, que características observadas numa dada escala sejam visualmente mais similares para outras naquela mesma escala. Cada escala gera primitivas para a representação inicial, e estas devem embutir as propriedades físicas do elemento observado. É isso que caracteriza fisicamente a superfície naquela escala, é essa informação que a primitiva carrega para as outras escalas. Estas propriedades podem ser o tamanho, a cor, a orientação, o contraste local, etc. Então cada primitiva gerada para uma superfície numa escala tende a ser mais similar para outras com propriedades semelhantes na mesma escala, e não para qualquer outra em escala diferente ou com propriedades muito diferenciadas (primitivas dissimilares). Desta forma, as primitivas definidas carregam uma informação que pode facilitar o entendimento da estrutura espacial da superfície nos níveis hierárquicos mais altos. Esta asserção aposta no processamento simbólico, quando propriedades locais das primitivas são comparadas e inferem novas e mais complexas organizações.

4 - *Continuidade*

Considerando que as primitivas da representação inicial estão sobre o critério da similaridade e possuem uma organização hierárquica, elas terão sempre uma organização no espaço. Elas vão estar arranjadas em linhas e é possível que criem padrões mais complexos que aquilo que individualmente representam. A tendência é de organizarem-se sempre definindo contornos "macios" para as superfícies. Nós somos bastante sensíveis à continuidade no espaço, e isto pode ser observado na Fig. 3.1, onde diversas organiza

ções coexistem "disputando" a percepção final da cena. Note que todas elas no entanto, são caracterizadas pela continuidade no arranjo dos elementos básicos formadores da imagem.

6 - Continuidade de fluxo

A última asserção trata de um ponto crucial para visão, o movimento. O sistema visual humano tem extrema mobilidade. O observador se movimenta, so objetos no mundo se movimentam, os olhos se movimentam. Considerando rígidas as superfícies, o movimento de partes destas superfícies que estão próximas uma da outra são semelhantes. Disto resulta que o campo médio de velocidades dos elementos representantes de imagem analisada deve variar uniformemente. Se a direção do movimento é descontínua para mais que um ponto ao longo de uma linha representada, então um limite para o objeto se encontra presente.

É interessante relacionar o tratamento dado por Marr (1982) as asserções físicas e os trabalhos dos psicólogos "gestaltistas", principalmente Wertheimer, Kohler e Koffka. Essa corrente sustentava que só havia alguma organização perceptiva a nível visual se o estímulo provocador fosse variável. Um campo de estímulos homogêneo produziria muito pouca informação perceptual. Casado a esta determinação, experiências demonstravam que a habilidade de perceber padrões era limitada pela nossa capacidade de processar as unidades do padrão isoladamente.

Veja Fig. 3.3, por exemplo. Em (a) não é possível determinar os 36 pontos presentes num único olhar sobre a figura. É necessário contá-los, isto é, definir algum processo para trabalhar as unidades formadoras da figura. A tendência inicial é dizer que se vê uma coleção de pontos com uma certa estrutura espacial. A figura em (b) já não será tratada como uma coleção de pontos, a descrição tenderá

a identificar três colunas de pares de pontos, ou ainda três linhas de pequenos "quadrados". Na figura em (c), onde pontos cheios e vazios coexistem, os padrões são ainda mais distinguíveis. Assim uma descrição pode ser a de que temos ali duas colunas de pares de pontos vazios nos extremos com uma coluna de pares de pontos cheios no centro.

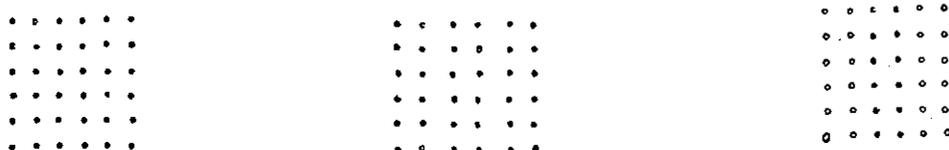


Fig. 3.3 - Organizações perceptivas de conjunto de pontos.

Estes fatores identificam uma necessidade de segregar estímulos, para que tenhamos uma percepção mais eficiente. No entanto, como não temos a capacidade de discriminar todos os estímulos físicos presentes, o campo perceptivo tende a ser organizado em um número limitado de todos ou unidades, em formas, constituindo a Gestalt (Forgus, 1971). Para os gestaltistas, a lei básica a governar a organização perceptiva é aquela que trata da organização dos elementos primitivos. Estas primitivas sempre tenderiam a se agrupar, definindo a forma que seria a mais estável ou que provocaria menos tensão. Esta lei é a chamada *lei de pregnância*, onde a forma de pregnância ou "boa forma" é aquela que constitui a mais simples estrutura estável (Forgus, 1971).

Esta definição se cerca de muitos termos mal definidos ou não esclarecidos. Para trabalhar sobre ela, vários critérios foram formulados e se constituíram em leis perceptivas para a escola gestaltita. Estas buscavam explicar o funcionamento do princípio da pregnância. Estas leis constituíram-se nas leis intrínsecas, governadas principalmen

te pelas formas (gestalts) criadas a partir dos elementos primitivos, e que, ao operar, são independentes da experiência (de conhecimento cognitivo). A idéia central destas leis é verificar como o observador é levado a perceber uma certa forma na direção de maior pregnância. As principais leis definidas foram a lei de fechamento, a lei da boa continuidade e direção, a lei da proximidade e a lei de similaridade (Forgus, 1971; Vickers, 1979).

Os psicólogos da Gestalt determinaram a existência destas leis de comportamento perceptivo, mas não trabalharam na explicação dos mecanismos através dos quais elas poderiam ser realizadas. Marr (1976, 1982) acredita nas leis de gestaltistas mas vai além. Ele verifica a validade física destas com as asserções básicas que propõe. Modificando-as não na assência e sim na substância, passa a tratá-las como resultado das restrições físicas impostas as superfícies pelo mundo visível. Seu conjunto de leis é agora fisicamente plausível, podendo assim se trabalhar na identificação e explicação dos processos geradores de agrupamentos de primitivas, que criam a percepção das Gestalts. A análise de restrições que faremos em algumas de nossas aplicações do modelo, estará vinculada às asserções físicas e a um entendimento sobre esta ótica das leis Gestaltistas para percepção.

3.4 - ALGUNS PRINCÍPIOS GERAIS

O estudo das asserções físicas básicas feito em 3.3, traz uma informação fundamental para a formulação de uma proposta de representação visual primária. O elemento na Imagem Intensidade de entrada que carrega a informação física da cena é a variação de intensidade. São elas as primitivas analógicas do SVH, e são elas que determinam uma variedade de organizações espaciais, aproximadamente independentes para as diferentes escalas. Essas organizações são responsáveis em refletir a estrutura das superfícies visíveis das cenas analisadas. As variações na matriz de entra

da trazem consigo importantes e necessários indícios sobre a estrutura física da cena. Estes indícios necessitam ser capturados e se fazerem presentes na representação primária da imagem. Precisa estar codificado nesta representação a informação física que as "variações" carregam. A proposição é que isso seja feito com um conjunto de primitivas simbólicas, criadas a partir das variações na matriz de entrada (informação analógica) e das restrições impostas pelo mundo visível geradas a partir das asserções físicas básicas.

A representação que contém essas idéias foi chamada por Marr (1979) de PRIMAL SKETCH. Chamaremos neste trabalho de ESBOÇO PRIMÁRIO e esta representação.

O ESBOÇO PRIMÁRIO é uma representação bidimensional e centrada na retina (retinocêntrica), isto é, centrada no observador. Ela fica completamente determinada pelos itens abaixo:

1 - Consiste de primitivas simbólicas do mesmo tipo geral, que se apresentam sobre um conjunto de diferentes escalas;

2 - As primitivas são geradas primeiro pela análise das variações nos valores de intensidade da matriz de entrada (variações de intensidade), então representando estas, formando símbolos diretamente a partir daí. Operando sobre estes símbolos como processos seletivos e de agrupamento, símbolos novos são gerados, denominados símbolos de maior escala e estes tendem a refletir as estruturas das escalas maiores na imagem. Assim, de maneira recursiva vai se construindo a estrutura das superfícies que "vemos";

3 - As primitivas obtidas possuem parâmetros associados a elas, são seus atributos. Estes atributos é que determinam a significância física daquela primitiva. São

grandezas como tamanho, posição, orientação, etc., e a precisão destas medidas determina o quanto mais fácil e mais confiável será a recuperação da informação sobre a geometria básica das superfícies visíveis.

Para gerar o ESBOÇO PRIMÁRIO três estágios centrais são necessários. O primeiro é detectar e representar as variações de intensidade na matriz de entrada para as várias escalas. O segundo é trabalhar na interligação entre estas primitivas geradas para as diversas escalas, criando o ESBOÇO PRIMÁRIO BRUTO. O terceiro é trabalhar os processos seletivos e de agrupamento sobre as primitivas representadas, criando o ESBOÇO PRIMÁRIO COMPLETO (EPC), a proposta de representação primária para dados visuais. Veja Fig. 3.4.

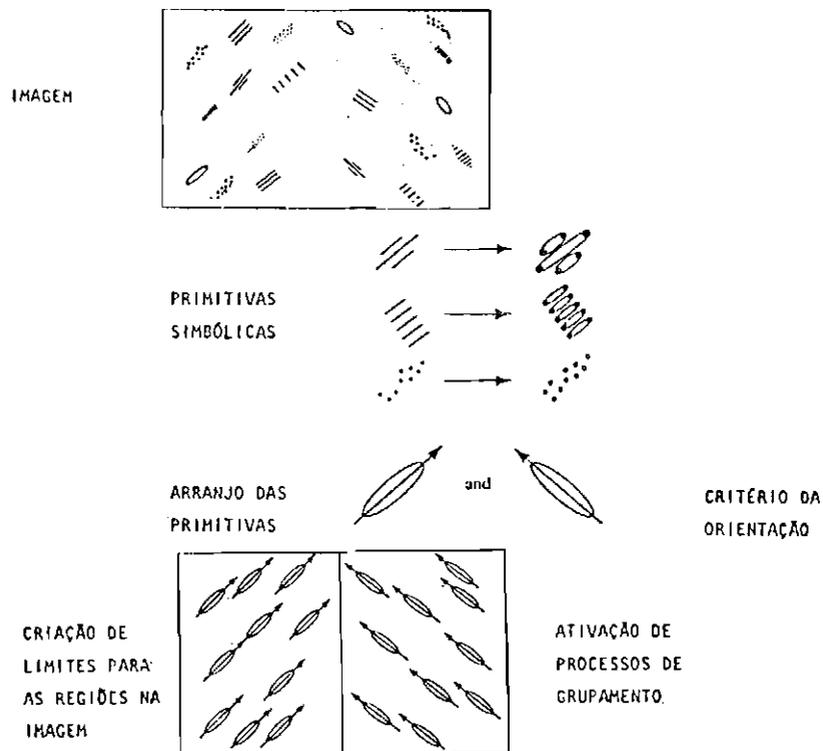


Fig. 3.4 - Representação esquemática da idéia do ESBOÇO PRIMÁRIO.

Observe que o ESBOÇO PRIMÁRIO COMPLETO de uma cena, já define primitivamente formas, regiões e posições . Os processos de seleção e agrupamento levam naturalmente a uma percepção primitiva da cena observada. Na implementação realizada, a noção de ESBOÇO PRIMÁRIO COMPLETO está ligada a uma aplicação. O EPC aqui conta apenas com uma ou algumas das muitas estruturas possíveis na cena original. Estas estruturas perceptivas, determinadas e isoladas do ESBOÇO BRUTO, serão o produto final do processamento para os experimentos realizados. Veremos isto com mais cuidado no capítulo 4.

3.5 - COMO DETECTAR E REPRESENTAR AS VARIAÇÕES DE INTENSIDADE DE

O estudo para detecção de variações de intensidade, desde muito está às propostas de detecção de "bordas" em imagens digitais. O termo "bordas" tem, no entanto, um forte significado físico, ele faz pensar nos limites reais dos objetos na cena. Se temos uma "borda" ela deveria corresponder fielmente a um limite físico existente na superfície do objeto observado.

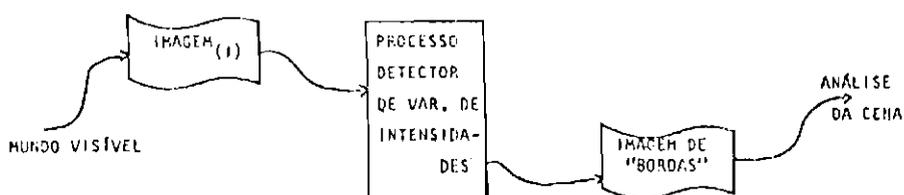


Fig. 3.5 - Esquema simplificado para detecção de bordas

Simplificadamente, o esquema adotado para detectar "bordas", tem sido implícita ou explicitamente do tipo apresentado na Fig. 3.5. Neste esquema não existe distinção entre a variação de intensidade detectada e aquilo que intuitivamente chamamos "bordas" da cena.

Existem pelo menos quatro (4) fatores afetando diretamente os valores da intensidade de luz refletida: a geometria das superfícies visíveis, a reflectância destas, a iluminação sobre a qual se encontra a cena e a posição do observador em relação às superfícies observadas. Sendo desta forma, algumas variações detectadas podem ser causadas por um destes fatores, outras causadas por outro fator e ainda teríamos aquelas cuja causa é uma combinação de fatores. Identificar a propriedade física que motivou a variação é tarefa essencial na linha que vem sendo proposta. O coração da percepção visual é a capacidade de inferir a partir das estruturas de uma imagem a estrutura física, a realidade, do mundo visível. Para isso não podemos confundir variações de intensidade com "bordas" na cena. As variações de intensidade serão buscadas e representadas de forma a criar as primitivas simbólicas básicas da representação.

Como, então, detectar as variações de intensidade?

Os métodos comumente usados 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 (I) diferenciada é utilizada, buscando-se nela "picos" (extremos positivos ou negativos) ou "zero-crossings" (valores onde existe a transição entre o positivo e negativo).

A "suavização" da imagem (I) 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 intensidade levam a "picos" na primeira derivada ou a cruzamentos por zero ("zero-crossings") na segunda derivada. 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 (Marr, 1982; Hildreth, 1983). Iniciou-se então uma série de estudos que culminaram com um modelo quantitativo da estrutura deste canais em humanos (Marr, 1982).

A idéia básica define que para cada ponto (P) no campo visual (plano visual da retina) existem quatro (4) filtros de tamanhos diferentes analisando a informação. Estes filtros, os canais, foram rotulados de N, S, T e U em ordem crescente de tamanho (Existe evidência de que um canal menor que o N está presente, para nossa análise no momento isto é irrelevante, no entanto para mais informações consulte referências em Marr (1982)). Da fisiologia era conhecido que a organização espacial dos campos receptivos das células na retina é, em geral, circularmente simétrico, com uma região central "excitada" e a região em sua volta "inibida". Considerando estes fatos algumas células foram denominadas ON-CENTER, quando tinham o centro de seus campos receptivos excitados por uma pequena fonte de luz, e outras denominadas OFF-CENTER, quando inibidas sob o mesmo critério (Fig. 3.6).

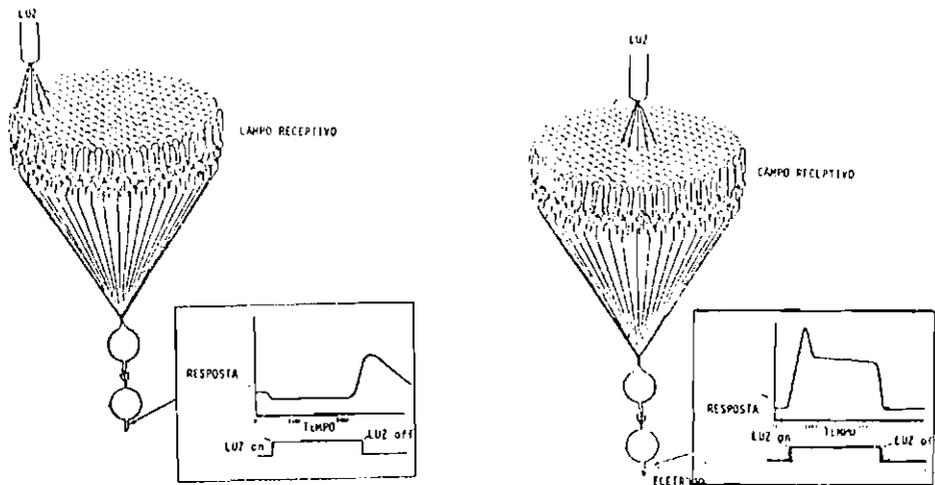


Fig. 3.6 - Resposta de células ON CENTER a estímulo de luz.

Supondo que a organização observada era o resultado da sobreposição entre uma pequena região excitada e outra grande região inibida, cada uma das regiões descritas por uma gaussiana, modelou-se o campo receptivo destas células como sendo uma função DOG, diferença de duas distribuições gaussianas, definida pela equação (Marr, 1976, 1982):

$$\text{DOG}(\sigma_e, \sigma_i) = [1/(2\pi)^{1/2}\sigma_e] \exp(-x^2/2\sigma_e^2) - [1/(2\pi)^{1/2}\sigma_i] \exp(-x^2/2\sigma_i^2)$$

onde:

σ_e - desvio padrão para a região excitada.

σ_i - desvio padrão para a região inibida.

Os canais psicofísicos N, S, T e U, corresponderiam provavelmente às células caracterizadas por essas funções DOGs (Marr, 1982).

Consideradas estas questões, podemos voltar ao problema como "suavizar" a imagem. A definição de canais visuais sugere que as variações de intensidade ocorrem em diversas escalas, dadas pelos tamanhos dos canais. O processo

de "suavização" deve considerar este fato. Tratando-se com imagens digitalizadas, "suavizar" significa achar um filtro digital que responda a essa característica quando convolvendo a imagem. O filtro deve ser "macio" (um passa-baixa) e limitado no domínio da frequência, isto é, sua variância Δw deve ser pequena. Isto considera a possibilidade de trabalhar escalas, filtros com tamanhos diferentes podem ser definidos.

Das asserções físicas (1) *existência de superfícies* e (2) *organização hierárquica* analisadas em 3.3, podemos considerar que para cada escala, os fatores físicos do mundo visível que conduzem a variações na imagem intensidade são localizados no domínio do espaço. Isto gera a *restrição física de localização espacial*, que como consequência dita que cada ponto na imagem filtrada origina-se a partir de uma média de seus vizinhos. Desta forma o filtro deve ser localizado no domínio de espaço, isto é, sua variância Δx também deve ser pequena. Existe então um requisito claro. O filtro deve ser localizado tanto no domínio da frequência quanto no domínio do espaço e esta é uma situação conflitante. A distribuição que otimiza a relação entre Δw e Δx , segundo Marre Hildreth (1980), é a gaussiana.

$$G(x,y) = e^{-\frac{x^2 + y^2}{2\sigma^2}}, \text{ Gaussiana bidimensional.}$$

Ora, cada membro na definição do campo receptivo das células era uma gaussiana, e o filtro a ser proposto baseado em critérios físicos (localização espacial) chegou a definição de uma gaussiana como operador de "suavização".

Agora o problema é qual operador diferencial usar. Toda vez que ocorre na Imagem Intensidade uma variação significativa encontra-se:

- (1) Um "pico" na primeira derivada;
- (2) Um cruzamento por zero na segunda derivada direccional.

Detectar variações pode passar então a ser o problema de encontrar cruzamentos por zero na segunda derivada da intensidade na direção apropriada. Para sinais unidimensionais, sobre 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, a "riqueza" de informação dos cruzamentos por zero fica estabelecida. Além disso, mecanismos fisiológicos simples podem ser propostos para detecção de cruzamento por zero pelo aparelho visual humano (Frisby, 1979; Marr, 1982).

Os cruzamentos por zero fornecem, então, um caminho para se ir de uma representação analógica da imagem, como a matriz de valores de intensidade para uma representação simbólica discreta. A idéia passa a ser definir uma função, cuja saída sejam os cruzamentos por zero (CZ) da imagem intensidade filtrada. Então teriamos:

$$f(x,y) = D^2 (G(x,y) * I(x,y)) \quad \text{onde:}$$

D^2 é o operador derivativo de segunda ordem;

$G(x,y)$ é a gaussiana bidimensional, cujo desvio padrão define a escala e opera a "suavização";

$I(x,y)$ é a Imagem Intensidade;

$f(x,y)$ é a saída cruzamento por zero para a escala definida.

Das regras da convolução temos:

$$f(x,y) = D^2 G(x,y) * I(x,y)$$

Desta forma as variações de intensidade nas diversas escalas podem ser calculadas convolvendo a Imagem Intensidade, $I(x,y)$, com um operador D^2G e procurando pelos cruzamentos por zero na saída desta operação. Mas D^2 é um operador direcional, qual seria a orientação associada com ele? Veja Fig. 3.7.

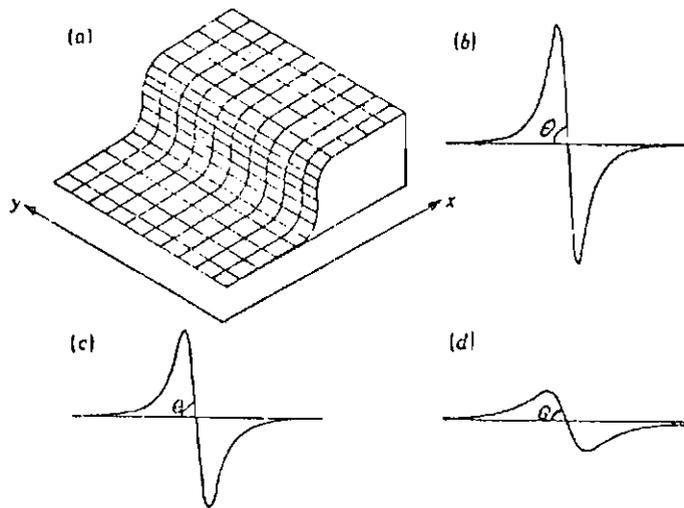


Fig. 3.7 - Perfil de intensidade e suas derivada em algumas direções.

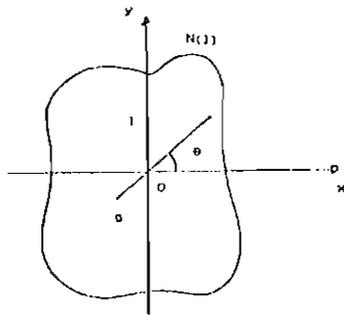
A orientação escolhida para a derivada deve coincidir com a orientação formada localmente pelos CZ obtidos da imagem. Esta definição se deve a motivação pela qual estamos buscando detectar as variações de intensidade e representá-las. Elas devem corresponder às propriedades físicas do mundo visível como variações na reflectância, na iluminação, na orientação das superfícies ou na posição do observador. Sendo estas propriedades, para cada escala, localizadas espacialmente e contínuas (derivado das asserções físicas) é possível associá-las com uma direção tal que projeta uma certa orientação na imagem, definida pela orientação da linha gerada pelos cruzamentos por zero. Observando a Fig. 3.7, a variação mais provável de representar um fato

físico é aquela mostrada em (b), onde a derivada foi feita segundo o eixo X, isto é, perpendicular a orientação definida pela linha cruzamento por zero. É nessa direção que a inclinação (θ), no cruzamento por zero, é menos "macia". A condição de inclinação (θ) máxima para a segunda derivada é interessante na medida que confere maior credibilidade a evidência de que a variação detectada, tem correspondente na estrutura física da cena. Verificar sob que condições a direção escolhida para a derivada coincide com a direção de inclinação (θ) máxima, é importante para a fundamentação da proposta. Marr e Hildreth (1980) argumentam com um teorema para definir esta condição.

Teorema:

Seja l um segmento de reta aberto sobre o eixo Y e que contém a origem O , Suponha que $f(x,y)$ é contínua, duplamente diferenciável e que $N(l)$ é uma vizinhança bidimensional aberta de l . Assuma que $\partial^2 f / \partial^2 x = 0$ em l . Então, se $\partial f / \partial y = 0$ é constante em $N(l)$, a inclinação da segunda derivada feita perpendicularmente a l é maior que a inclinação desta derivada feita ao longo de qualquer outra linha que cruza a origem O .

Prova:



Considere o segmento de reta $\Omega = (r \cos \theta, r \sin \theta)$ para θ fixado e valores de r suficientemente pequenos

para que esteja na vizinhança de 1, em $N(1)$.

$\Omega \rightarrow (r \cos\theta, r \sin\theta)$ e:

$$x = r \cos\theta$$

$$y = r \sin\theta$$

$$(\partial^2 f / \partial \Omega^2)_{r,\theta} = ?$$

$$(\partial f / \partial \Omega) = (\partial f / \partial x \cdot \partial x / \partial \Omega) + (\partial f / \partial y \cdot \partial y / \partial \Omega)$$

$$\partial f / \partial x = \partial x / \partial r + \partial x / \partial \theta = \cos\theta + \partial x / \partial \theta = 0, \quad \theta \text{ fixo.}$$

$$\partial f / \partial y = \sin\theta$$

fazendo $\partial f / \partial \Omega = \alpha$ temos:

$$(\partial^2 f / \partial \Omega^2) = (\partial \alpha / \partial x \cdot \partial x / \partial \Omega) + (\partial \alpha / \partial y \cdot \partial y / \partial \Omega)$$

como $\partial f / \partial y = 0$ ao longo de 1, vamos ter:

$$(\partial^2 f / \partial \Omega^2)_{r,\theta} = (\partial^2 f / \partial x^2) \cos^2\theta$$

Então $(\partial^2 f / \partial \Omega^2)_{r,\theta}$ tem inclinação máxima para $\theta = 0$, uma linha perpendicular a 1. Isto estabelece que a orientação da linha CZ é perpendicular a orientação na qual os CZ's têm máxima orientação.

A esta condição definida pelo teorema, denomina-se *condição de variação linear*, e é quem determina que a função variação de intensidade próxima e paralela a linha de cruzamentos por zero definida na imagem é localmente linear. Esta condição é aproximadamente verdadeira em imagens "suavizadas", amaciadas com filtros passa-baixas por exemplo (Marr e Hildreth, 1980; Hildreth, 1980, 1983).

Já temos como detectar CZ's usando operadores derivativos direcionais, no entanto convoluções são relativamente custosas computacionalmente e, se reduzidas, dimi

nuem o custo computacional. Se o operador direcional não for imprescindível, ele pode ser substituído por algum operador derivativo de segunda ordem que seja independente da orientação. Este operador é o Laplaciano (∇^2).

$$\nabla^2 = \partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$$

Temos que verificar sob que condições será possível esta substituição. Novamente Marr e Hildreth (1980) mostraram que, na condição de variação linear, os CZ's são detectados e precisamente localizados pelos valores zero do operador Laplaciano (∇^2), agindo sobre a imagem filtrada com a gaussiana ($G * I$). Combinando os dois operadores, o Laplaciano e a Guassiana, temos um operador $\nabla^2 G(x,y)$ 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. O operador $\nabla^2 G$ fica definido pela equação (1):

$$\nabla^2 G(r) = 1/\sigma^2 [r^2/\sigma^2 - 2] \exp(-r^2/2\sigma^2) \quad (1)$$

$$e: r^2 = x^2 + y^2$$

onde r é a distância radial a partir da origem, dada a simetria circular do operador.

A saída cruzamento por zero fica grosseiramente definida por:

$$CZ = \nabla^2 G * I = \emptyset$$

Cada valor de σ , desvio padrão da gaussiana G , define uma escala, um canal, semelhante aqueles canais sugeridos pelos experimentos psicofísicos. O tamanho do operador é definido pelo diâmetro da região central, denominado W , e podem ter uma relação direta com os tamanhos encontrados para os canais psicofísicos. Para isso basta achar o valor de r que

faz $\nabla^2 G = 0$, dado que $W = 2r$.

É interessante, neste ponto, relembrar que da fisiologia veio um modelo quantitativo para a estrutura dos campos receptivos de células na retina, este era a função DOG definida anteriormente (página 32). O operador $\nabla^2 G$ é o limite de uma função DOG. Pode-se provar, que sobre certas condições relacionadas a razão entre os desvios padrão das regiões inibida e excitada (σ_i/σ_e), a melhor aproximação para o operador $\nabla^2 G$ é dada por uma função DOG (Marr e Hildreth, 1980). Sendo assim, se o sistema visual utiliza uma DOG, estão implicitamente ele faz uso da suposição de que a condição de variação linear é aproximadamente satisfeita.

Utilizando o operador $\nabla^2 G$ sobre a imagem, conseguimos detectar as variações de intensidade com significância física para uma dada escala. Precisamos, agora, identificá-las e representá-las. É isto que vai nos gerar as primitivas simbólicas primárias da representação inicial da imagem. Para representar essas variações, a saída da convolução ($\nabla^2 G * I$) gera linhas de CZ's 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 o SEGMENTO CRUZAMENTO POR ZERO (SCZ). O que é um SCZ? Para cada CZ da saída de ($\nabla^2 G * I$) duas propriedades adicionais devem ser calculadas. Estas propriedades são o CONTRASTE e a ORIENTAÇÃO LOCAL. O CONTRASTE de um certo CZ está relacionado a magnitude da variação de intensidade local, detectada por aquele CZ. A ORIENTAÇÃO LOCAL é definida pela direção da tangente ao longo do contorno definido pelos CZ'S da imagem. Para obter estes atributos, podemos determinar o gradiente na saída da convolução [$\nabla^2 G * I$]. O CONTRASTE será dado pelo módulo do gradiente e a ORIENTAÇÃO LOCAL pela sua direção.

Veja Fig. 3.8. Nela em (a) temos o operador unidimensional $D^2G(x)$, em (b) temos definido um perfil de intensidade e em (c) o resultado da convolução de (a) com (b).

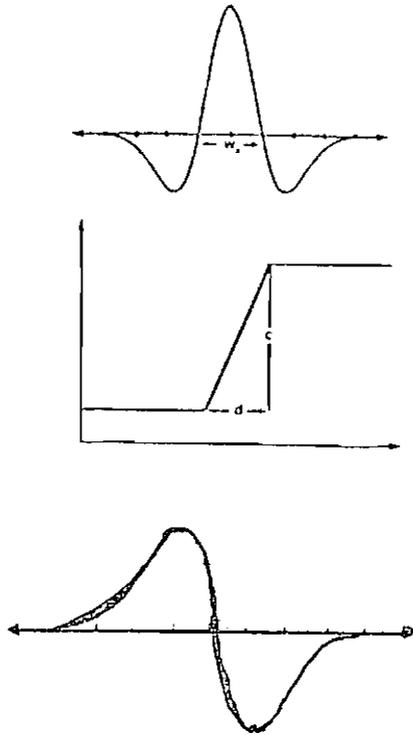


Fig. 3.8 - Definição de contraste e orientação para Slope.

Chamamos a grandeza vetorial $S(i)$ de SLOPE do CZ definido para a escala determinada pelo operador $\nabla^2G(x)$. O SLOPE unidimensional fica definido pela equação (2) (Hil_{dreth}, 1980).

$$S(id) = c \exp(-d^2/8\sigma^2) \quad (2)$$

onde:

- c é o contraste;
- d é a extensão da variação
- σ é o desvio padrão da Gaussiana que define a escala.

. No caso bidimensional, temos a equação (3):

$$S(2d) = cw \exp(-d^2/4\sigma^2) \quad (3)$$

onde:

w é o tamanho da máscara bidimensional.

A idéia do SLOPE é bastante importante quando buscamos interligar escalas, canais, diferenciados. Isto será tratado mais detalhadamente no capítulo 4.

Conhecidas as propriedades podemos falar sobre o segmento cruzamento por zero (SCZ). Cada um deles descreve um pedaço do contorno obtido em $(\nabla^2 G * I)$, para o qual o valor do CONTRASTE e da ORIENTAÇÃO LOCAL podem ser considerados uniformes. Os SCZ's vão carregar estas informações e mais duas informações sobre sua posição e seu tamanho na imagem. 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, SCZ's, marcam a transição entre a análise analógica e simbólica da cena, e são os precursores das primitivas que descreverão o ESBOÇO PRIMÁRIO BRUTO.

Veja a Fig. 3.9. Esta mostra um possível tipo de resultado para a descrição das variações de intensidade em uma escala. A largura, W, do operador $\nabla^2 G$ é de 8 pixels. Em (a) temos a imagem original. Em (b) aparecem os contornos CZ obtidos na saída do operador. Em (c) os SCZ's são mostrados.

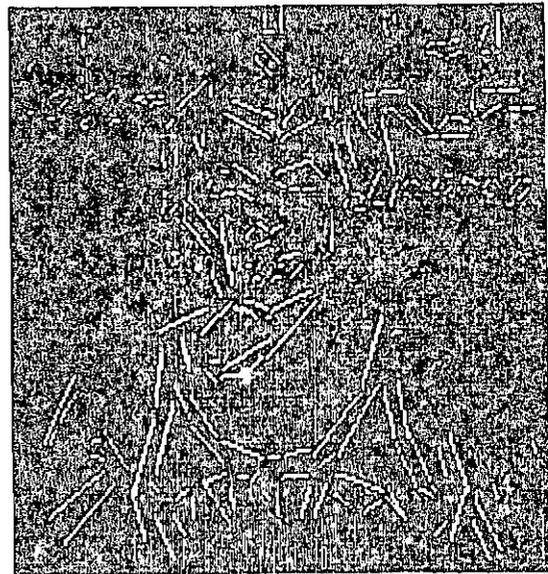
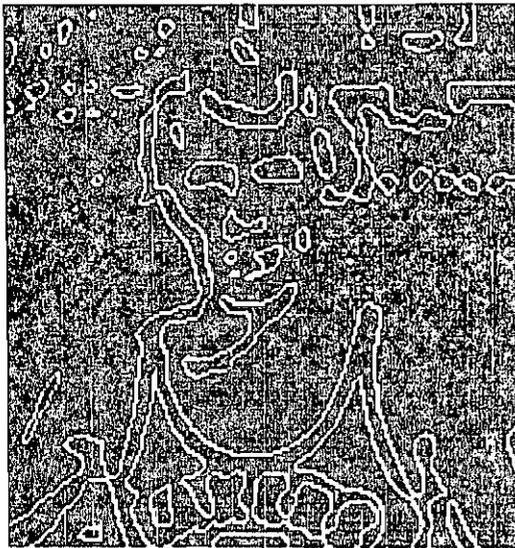


Fig. 3.9 - (a) Imagem original. (b) Saída do operador $\nabla^2 G$ aplicado a imagem em (a). (c) Mostra os segmentos cruzamentos por zero para a escala definida pelo operador em (b)

Um tipo de descrição possível seria:

(SEGMENTO	(posição	p)
	(comprimento	t)
	(orientação	o)
	(contraste	c))

Outros parâmetros poderiam aparecer como mais um atributo daquele segmento. Bastaria uma clara definição destes em relação às propriedades físicas com explicitam, e a forma de obtê-los a partir da imagem de intensidade. Por exemplo, o tamanho é definido como o número de pixels que forma um SCZ. O tamanho traz embutido uma noção da escala hierárquica em que se está trabalhando.

3.6 - COMO GERAR O ESBOÇO PRIMÁRIO BRUTO

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 $\nabla^2 G$ de tamanhos diferentes. Sabemos que as variações ocorrem para as diversas escalas, isto é, para os diversos canais, no entanto devemos ter uma única descrição da cena. Sendo assim, de alguma forma as primitivas simbólicas obtidas para cada canal devem se combinar.

A priori no entanto, não existe nenhuma razão para justificar que estes canais se interliguem através de suas primitivas. Recorrendo às asserções físicas, é possível verificar a existência de uma restrição natural imposta às superfícies nas cenas. 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 mai

or. Isto é, se para uma imagem I filtrada por operadores $\nabla^2 G$ 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 então 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 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 $\nabla^2 G$, para estabelecer uma realidade física mínima para a apresentação, isto é, as primitivas geradas a partir da combinação serão fisicamente significativas.

Consideradas estas questões, de uma certa maneira, a estrutura do mundo físico visível restringe a geometria (o arranjo espacial) dos SCZ de diferentes canais. Is

to levou a definição da *coincidência espacial* para os SCZs, enunciada da seguinte forma (Hildreth, 1980):

Se um SCZ está presente para um conjunto de canais independentes ∇^2G , e os SCZs tem a mesma posição e orientação em cada canal, então o conjunto de segmentos SCZ indicam a presença de uma variação de intensidade na imagem de entrada que é devida a um único fenômeno físico: uma variação na reflectância, na iluminação da cena, na posição do observador ou na orientação da superfície, e podem ser representados unicamente.

Verificar esta correspondência para os SCZ de diversos canais leva a algumas ponderações. A primitiva que temos até agora foi definida em 3.5, como sendo segmento cruzamento por zero. Nas escalas menores a detecção é mais ruidosa e muitos são os segmentos descritos. Grosseiramente estas escalas definem o que "vemos". Nas escalas maiores, mais "largas", também detectamos e representamos as variações, no entanto a "realidade física" é modificada para redução de ruído. Este é um efeito causado pela maior "suavização" provocada pelo operador ∇^2G para essas escalas. Como isto acontece a relação de um para um entre os SCZ pode não ser satisfeita diretamente.

Para compatibilizar a interligação dos canais, é necessário definir configurações espaciais de SCZ nas escalas menores. Estas configurações devem tornar explícitas a relação de proximidade destes SCZ's nestes canais, tornando possível sua identificação para as escalas maiores. Novos elementos descritivos além do SCZ são necessários. Para cada canal alguns tipos de arranjos dos SCZ's devem se constituir em elemento de descrição. Estes elementos são os BLOBS e as BARRAS. As TERMINAÇÕES podem ser outro tipo descritivo elementar. Mas, o que é um BLOB, uma BARRA ou uma TERMINAÇÃO? Vamos tentar defini-los.

Um BLOB é um pequeno contorno fechado, com uma extensão espacial definida, observado na saída ($\nabla^2 G * I$). Em vez de descrever o contorno como um SCZ, ele passará a ser descrito como um BLOB, que tem os seguintes atributos (Propriedades de um BLOB):

- (1) CONTRASTE MÉDIO ao longo dos pontos do contorno;
- (2) ORIENTAÇÃO, especificada com relação ao eixo horizontal, através da orientação da linha que contém o seu centro de gravidade;
- (3) COMPRIMENTO, uma medida da extensão do contorno fechado ao longo do seu maior eixo;
- (4) LARGURA, medida do menor eixo.

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, ou SCZ não necessitam ter a mesma orientação para serem ditos paralelos (Critérios tanto para limites em definição de BLOBS quanto de BARRAS podem ser encontrados em Hildreth, 1980). O descritor associado com esses dois SCZ estará localizado no centro deles e tem os seguintes atributos:

- (1) CONTRASTE dos dois SCZ;
- (2) ORIENTAÇÃO MÉDIA dos dois SCZ;
- (3) COMPRIMENTO, definido pelo tamanho do SCZ;
- (4) LARGURA, definida pela distância entre os dois SCZ

Existem evidências de que, nos estágios iniciais do processamento de informação visual no SVH, os pontos terminais das "bordas" se fazem explícitos, uma vez que eles são um indício muito importante de descontinuidade (Marr, 1982). Estas seriam as TERMINAÇÕES, os pontos terminais de SCZ, grosseiramente. 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 (Hildrth, 1980, 1983).

O ESBOÇO PRIMÁRIO BRUTO (EPB) é a representação que fica definida pela interligação destes elementos simbólicos descritivos, obtidos para as diferenciadas escalas (canais). A interligação de canais é ainda um problema em aberto. Neste estudo estaremos considerando a interligação de canais objetivando determinar para cada SCZ uma evidência de sua realidade física na cena. Supondo válidas a *coincidência espacial* e a *condição de variação linear* sobre nossas imagens de trabalho, buscaremos para a cena uma descrição simbólica única.

Como vimos em 3.5, o SLOPE de um CZ define uma relação entre a magnitude da variação de intensidade e a sua extensão. Na interligação dos canais, uma razão entre os módulos dos SLOPES ($S(\text{canal } 1) / S(\text{canal } 2)$) será feita de maneira a que possamos definir a extensão e intensidade da variação capturada. No novo elemento descritivo, as propriedades de CONTRASTE e LARGURA aparecerão correspondendo às características físicas da variação que ocorreu na cena. Elas "carregam" uma informação relativa a estrutura física da cena observada e são atributos importantes para os processos posteriores de compreensão global da cena.

Possíveis descritores para as primitivas definidas seriam:

(BLOB (posição p1 p2)
 (comprimento t)
 (orientação local o)
 (contraste médio c)
 (largura l))

(BARRA (posição p)
 (comprimento t)
 (orientação média o)
 (contraste c)
 (largura l))

CAPÍTULO 4

O ESBOÇO PRIMÁRIO BRUTO PARA O PROCESSAMENTO DE IMAGENS DE SATÉLITES

4.1 - O IMAGEAMENTO POR SATÉLITE

No aparelho visual humano, o sistema de aquisição é bastante poderoso a nível das resoluções espacial e espectral das imagens que trata. A região mais central da retina, chamada fovea, possui uma extraordinária densidade de elementos fotorreceptores. Na fovea temos aproximadamente 150 mil cones/mm² (Lindsay e Norman, 1977). Nesta região os campos receptivos, definidos pelos agrupamentos de fotorreceptores que captam a luz refletida, podem ser formados até por um único cone, definindo uma resolução espacial muito "fina". No total são aproximadamente 250 milhões de elementos sensíveis à luz em nossas retinas para formação das imagens.

As imagens do mundo visível que podem ser adquiridas diretamente numa escala mais aproximada das imagens captadas pelo nosso aparelho visual chamaremos de imagens de cenas naturais. Por exemplo, as imagens adquiridas através de câmaras de TV. O modelo proposto por Marr está diretamente ligado a análise deste tipo de imagens.

Imagens geradas pelos sistemas de aquisição de satélites não são imagens de cenas naturais, considerando a definição do parágrafo anterior. Uma cena imageada por satélites "varre" uma região de grandes dimensões na superfície terrestre. As imagens geradas têm outra escala visual. O SPOT, por exemplo, um satélite Francês de sensoriamento remoto para cenas de recursos naturais, cobre aproximados 120 Km de superfície no solo. Para criar a imagem destes 120 Km ele conta com 6000 elementos sensores para cada uma de suas

três (3) bandas espectrais. Ele possui, ainda, uma banda pan cromática cobrindo 60 Km. Isto resulta em resoluções espa ciais de 20m e 10m para as bandas espectrais e pan cromáticas respectivamente. A cena gerada está em uma escala visual maior com relação às cenas imageadas pela retina ou por uma câmara de TV. Cada ponto representa um conjunto grande de pontos no solo. Esta cena transformada é a imagem que temos para trabalhar.

Considerando os sistemas de geração de imagens a bordo de satélites como sistemas lineares, o estudo desenvolvido para cenas naturais pode ser utilizado em imagens de satélites. As mesmas suposições e restrições no contexto de cenas naturais ficam validadas para as imagens que desejamos tratar. Só existe uma alteração da escala visual per ceptiva.

A proposta de Marr (1976) restringia a teoria naquele momento a cenas monocromáticas, monoculares e está ticas. As imagens geradas por satélites se enquadram neste contexto. A banda pan cromática dá a característica de mono cromaticidade desejada, além de ser a banda com melhor reso lução espacial. As cenas imageadas são monoculares, conside rando a existência de um único sistema detector a cada ins tante. A característica relacionada a cenas estáticas é es tabelecida por restrição na análise. Em relação a diferença de escalas perceptivas, resolve-se o problema analisando o tamanho dos operadores ∇^2G , de forma a sintozinär adequada mente seus diâmetros (W), para obtermos uma boa detecção das variações ocorrendo na cena.

A utilização do ESBOÇO PRIMÁRIO BRUTO como re apresentação inicial para algumas tarefas de processamento digital de imagens de satélites baseia-se nestas relações e adequações.

4.2 - O TAMANHO DOS OPERADORES V^2G

No tratando com escalas humanas, a definição para o tamanho dos operadores V^2G (escalas ou canais) pode ser auxiliada pela psicofisiologia (Marr, 1982). Para imagens de satélites o estudo torna-se mais empírico. Dois são os parâmetros determinados para a definição das escalas: o número de escalas e a separação entre elas. O número está ligado à aplicação desejada. Nesta implementação, optamos pelo número mínimo possível, duas escalas (canais), devido a restrições do hardware disponível. A separação objetiva caracterizar a organização hierárquica das funções de reflexão das superfícies. É isto que vai permitir construir uma representação onde as estruturas físicas das superfícies imageadas estejam explicitamente apresentadas.

A definição da menor escala é a principal decisão na escolha do conjunto de operadores. O fator primário na escolha do menor operador a ser usado é o nível de ruído do sistema de imageamento sendo usado. O ruído do sistema altera a "sintonia" dos canais. Imagens de satélites são sinais discretizados e ruidosos. Como os ruídos são maiores para as altas frequências, então estabelecendo um limite inferior para os operadores V^2G limitamos as altas frequências espaciais, limitando o ruído (Hildreth, 1983).

O que se deseja detectar e representar são as variações de intensidade que estão ocorrendo na cena. Como o sinal é discreto, uma restrição adicional pode ser estabelecida. Considerando um perfil de intensidade isolado como na Fig. 4.1, temos dois valores associados a ele: o seu contraste (c) e sua extensão (d).

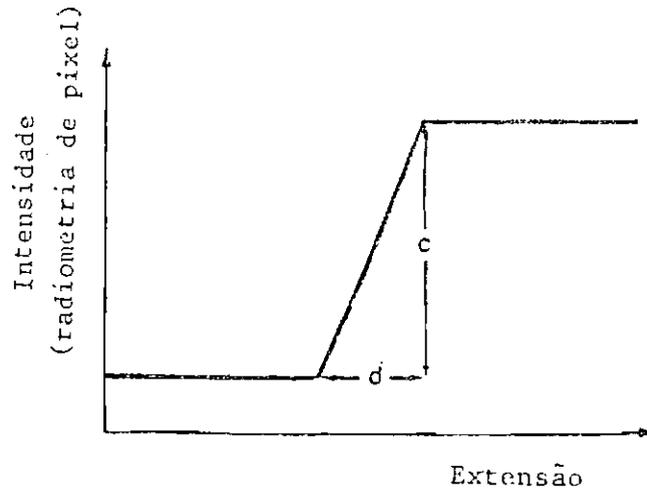


Fig. 4.1 - Perfil de intensidade isolado

Se a extensão (d) da variação é muito maior em comparação com o σ (desvio padrão da Gaussiana que define o operador $\nabla^2 G$), a saída da convolução [$\nabla^2 G * I$] flutua em torno do valor zero, rendendo múltiplos cruzamentos por zero (CZ) ao invés de um único. Considerado aqui o caso prático de imagens discretizadas. Com isto, deriva-se uma detecção muito ruidosa e imprecisa. Trabalhando esta restrição podemos determinar empiricamente o menor canal do conjunto de operadores para um dado sistema de imageamento. Para uma boa detecção, a extensão (d) da variação deve ser próxima ao tamanho do operador usado [$W \approx d$].

De alguns experimentos perceptuais, estabelece-se a proposição de que, na produção do EPB, os canais menores têm um papel mais significativo para a construção da descrição estrutural da cena. Os canais maiores seriam mais importantes para a estereoscopia e a análise de movimento (Hildreth, 1980). Com essa suposição, estabelecemos um conjunto inicial de cinco (5) operadores (canais). O tamanho das máscaras que aproximam estes operadores $\nabla^2 G$ nesta implementação é de [17 x 17] pontos. A opção por este tamanho de máscara baseou-se no tamanho máximo de nossas imagens de trabalho, [256 x 256] pixels, e em uma melhora no tempo do

processamento sem perder qualidade no filtro. Para determinar as máscaras não foi adotada uma aproximação digital para o operador $\nabla^2 G$. Para a definição algébrica do operador:

$$\nabla^2 G(r) = 1/\tau^2 [r^2/\tau^2 - 2] \exp (-r^2/2\tau^2)$$

com

$$r^2 = x^2 + y^2$$

Adotado um tamanho de [17x17] para as máscaras, a partir dos valores assumidos para os desvios padrão das Gaussianas, constituímos os canais. Maiores operadores (Hildreth 1980) e Grimmson (1981) utilizam máscaras de [32 x 32] rendem uma detecção mais confiável na presença dos ruídos ou de imagens pobremente quantizadas, o que deriva uma maior robustez ao processo.

Usamos imagens de tamanho máximo [256 x 256] pixels, também uma restrição do hardware. Para os sistemas de imageamento usados nos experimentos (Landsat TM e câmara de TV), os canais com $W = \{6, 8, 10\}$ mostram-se os mais adequados do conjunto definido por $W = \{4, 6, 8, 10, 12\}$ onde W é o diâmetro dos operadores $\nabla^2 G$, dado em número de pixels.

O menor canal é a resolução mais "fina" do sistema. Isto significa que não é possível distinguir variações de intensidade próximas que não sejam representadas por cruzamentos por zero (CZ) distintos no menor canal. Este dado é importante para a interligação dos canais, quando buscamos gerar a descrição única das cenas.

4.3 - AS PRIMITIVAS SIMBÓLICAS

A partir da definição do conjunto de operadores, é necessário criar a descrição simbólica de cada canal individualmente. No capítulo anterior, três das primitivas

simbólicas básicas da representação foram discutidas: o SCZ, a BARRA e o BLOB. Dado o tipo de imagens que objetivamos tratar, o BLOB foi subtraído como primitiva básica. O BLOB determina pequenos contornos fechados (Hildreth (1980) propõe limiares para sua definição). As dimensões no imageamento por satélites são consideráveis (O SPOT varre aproximadamente 120 Km e o LANDSAT TM 180 Km no solo). As variações de intensidade nestas imagens vão caracterizar transições, em geral, entre grandes superfícies no solo. A convolução da imagem com um operador V^2G leva a definição de contornos abertos e/ou contornos fechados, porém, em geral, de dimensões não reduzidas.

O contorno em linhas da cena é de grande importância para os processamentos posteriores. Para defini-los, os segmentos orientados SCZ's e as BARRAS são suficientes, dada a escala das imagens.

4.3.1 - O SEGMENTO CRUZAMENTO POR ZERO (SCZ)

Na saída da convolução [$V^2G * I$] temos a imagem dos cruzamentos por zero (CZ), definido contornos. O CZ é a representação simbólica atômica para as possíveis variações de intensidade na cena. Então, subdividindo entre contornos em segmentos de reta com uma orientação e tamanho definidos criamos o SCZ, a primitiva simbólica básica da representação.

Nesta implementação o tamanho mínimo para o SCZ é da escolha do usuário, e fica limitado ao tamanho máximo de $3W$ (diâmetro central do operador V^2G). Cada CZ possui duas propriedades: o CONTRASTE e a ORIENTAÇÃO. Para a computação destas propriedades foi utilizado o gradiente na saída da convolução.

$$VICZ = V(V^2G * I)$$

onde:

∇ , é o gradiente;

I_{cz} , é a imagem da saída $\nabla^2 G * I$;

I , é a imagem intensidade.

A aproximação digital usada para o gradiente foi a definida por:

$$V_{Icz} = [(\Delta x_{Icz})^2 + (\Delta y_{Icz})^2]^{1/2} \quad | \quad \text{arc tg}(\Delta y_{Icz}/\Delta x_{Icz})$$

com:

$$\Delta x_{Icz}(i,j) = I_{cz}(i-i,j) - I_{cz}(i-i,j)$$

$$\Delta y_{Icz}(i,j) = I_{cz}(i,j+i) - I_{cz}(i,j-i)$$

O CONTRASTE fica definido pelo módulo do gradiente ($|\Delta I_{cz}|$) e a ORIENTAÇÃO pela fase (arc tg), para cada CZ.

Para avaliar onde quebrar o contorno contínuo de CZ's de modo a definir o SCZ, uma função foi estabelecida. Ela se baseia em dois critérios: a continuidade na orientação e os limites no tamanho. A continuidade na orientação é definida por um limiar:

$$\psi_i = \psi(CZ_i) - \psi(CZ_{viz_i}) \geq \epsilon$$

onde:

ψ_i , é a função de avaliação de orientação em CZ_i ;

$\psi(CZ_i)$, é a orientação do CZ_i ;

$\psi(CZ_{viz_i})$, é a orientação do CZ vizinho de CZ_i ;

ϵ , é o limiar da quebra.

Se $\psi_i \geq \epsilon$, o contorno contínuo é quebrado na aquela CZ e o SCZ é gerado a partir dos CZ's aceitos na busca. Para definir o segmento de reta que representa o SCZ, nes

ta implementação, utilizamos o método do mínimo erro quadrático médio (MMEM), aproximando o conjunto de pontos CZ para um segmento de reta orientado.

O SCZ tem os seguintes atributos:

(1) CONTRASTE, definido pela média dos contrastes dos CZ's que geram o segmento;

(2) ORIENTAÇÃO LOCAL, definida pelas orientações locais de cada CZ que compõe o segmento, em conjunto com a orientação obtida pelo MMEM;

(3) TAMANHO, definido pelo número de CZ's que formam o segmento;

(4) POSIÇÃO, definida pelas coordenadas do ponto médio do conjunto de CZ's.

4.3.2 - A BARRA (BARRA)

A descrição final da cena deve garantir a interligação das diversas escalas (canais), objetivando que um máximo de informação das estruturas físicas da cena se façam presentes. Acontece normalmente que duas ou mais variações de intensidade em um canal menor podem estar interagindo em um canal maior. O reflexo deste fato é um deslocamento do SCZ no canal maior. Isto pode ocasionar um ruído durante os processos de interligação dos canais. Uma mesma variação poderia ter representações diferentes, indicando a falsa existência de duas variações na cena. Para contornar este problema define-se a BARRA.

A BARRA fica definida, nesta implementação, por dois SCZ's cujas orientações diferem de, no máximo, 0.1 rad, os tamanhos de, no máximo, quatro (4) pixels e possuem

uma distância (d) entre seus centros de no máximo $[W - 2]$ (W é diâmetro central do operador $\nabla^2 G$ que define o canal. Estes limiares podem ser alterados. Os valores aqui estabelecidos basearam-se nos sistemas de aquisição e nas aplicações desejadas.

A BARRA possui os seguintes atributos:

(1) CONTRASTE, definido pela média dos contrastes dos SCZ's que a geram;

(2) ORIENTAÇÃO, definida pela média das orientações dos SCZ's;

(3) TAMANHO, definido pelo tamanho do maior SCZ;

(4) LARGURA, definida pelo valor da distância (d) entre os SCZ's;

(5) POSIÇÃO, definida pela média das posições dos SCZ's.

4.4 - A INTERLIGAÇÃO DE CANAIS

Com SCZ's e BARRAS descrevemos os canais separadamente. Cada conjunto de SCZ's e BARRAS descritos apresentam determinadas estruturas da cena, capturáveis para aquela escala definida pelo operador $\nabla^2 G$ aplicado. Para a descrição final, a informação destes canais deve ser integrada. Pretendemos, com a interligação inferir, a partir dos SCZ's e BARRAS, a presença na cena de variações de intensidade que reflitam bordas físicas reais. Considerando válida a *suposição de coincidência espacial*, esta inferência torna-se possível.

Na interligação, uma nova primitiva descrita vai aparecer. A esta primitiva chamamos de BORDA. A BORDA é nossa primitiva simbólica de maior significância física. A BORDA carrega duas propriedades importantes, o CONTRASTE e a LARGURA, que refletem a magnitude e a extensão da variação de intensidade que ela representa. A equação abaixo relaciona o SLOPE (S_i) com a magnitude (c) e a extensão (d) da variação (Veja Fig. 4.1).

$$S_i = cW \exp(-d^2/4ci^2)$$

Para cada SCZ ou BARRA representando possíveis variações reais na cena, a propriedade CONTRASTE pode ser considerada como o módulo do SLOPE ($|S_i|$) para aquela escala. Desta forma, utilizando dois canais, a relação dos SLOPES ($|S_2|/|S_1|$) possibilita o cálculo da extensão (d) e da magnitude (c) da variação detectada.

Para caracterizar uma BORDA, é necessário que tenhamos um SCZ ou uma BARRA no canal menor que encontre um SCZ no canal maior, com orientação diferindo de, no máximo 0.1 rad. Os tamanhos devem ser similares, e a noção de similaridade pode ser trabalhada com a aplicação.

Além do CONTRASTE e LARGURA, a BORDA fica definida pelos seguintes atributos:

(1) TAMANHO, definido pelo média dos tamanhos das primitivas que a geram;

(2) POSIÇÃO, definida pelas coordenadas do ponto médio das posições das primitivas que lhe dão origem;

(3) ORIENTAÇÃO, definida pela média das orientações das primitivas geradoras.

A descrição final única é o ESBOÇO PRIMÁRIO BRUTO (EPB), modificado para aplicações em processamento de imagens de satélites. Na Fig. 4.2 apresentamos em (a) uma cena original, em (b) uma representação para os SCZ's, em (c) uma representação para as BARRAS isoladamente, em (d) uma representação do EPB completa com SCZ's, BARRAS e BORDAS e, em (e), uma listagem do arquivo final gerado, com os descritores para as primitivas simbólicas primárias mostradas em (d).

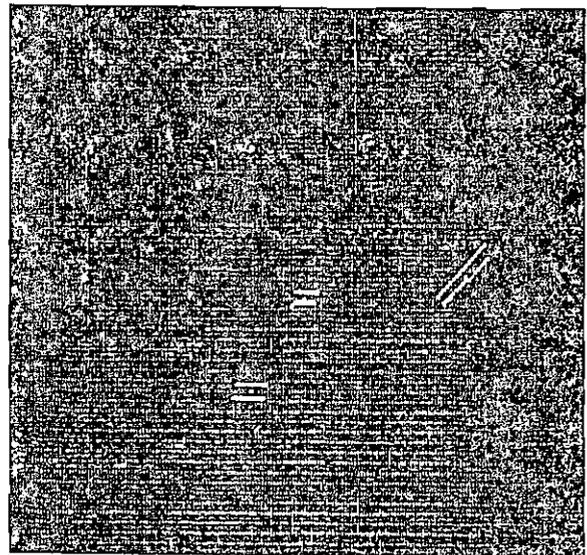
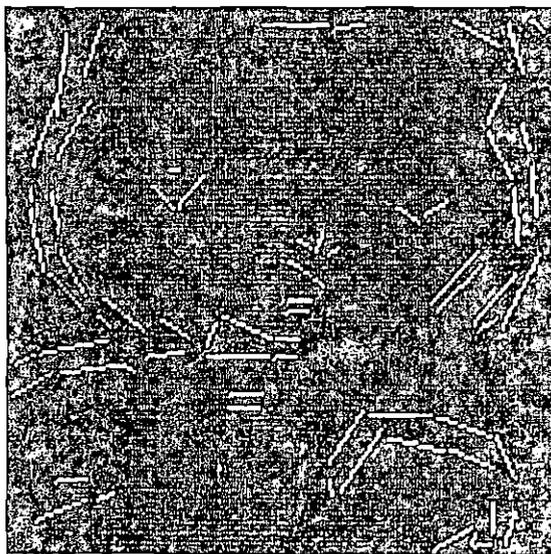
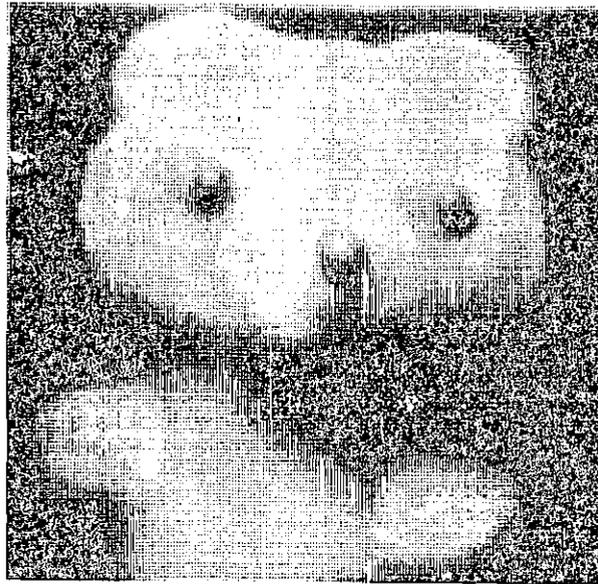
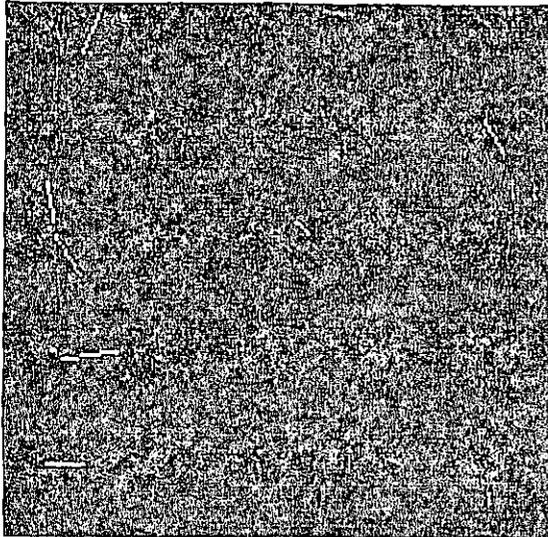


Fig. 4.2 - Em (a) urso de brinquedo (Marra, 1982), em (b) representação dos SCZ's da cena com $W = 8$ pixels, em (c) as BARRAS.



ARQUIVO DE PRIMITIVAS - jrcow6.prl

JANELA

Linha Inicial 124 Nro Linhas 154
 Coluna Inicial 235 Nro Colunas 170

LINHA	COLUNA	TIPO	CONTRASTE	ORIENTACAO	COMPR	LARGO	RETA
C 13	E 90	SCZ	+109	7.14	18	0	E 13 E 90 E 13 E 90
E 23	E 83	SCZ	+10	-0.26	7	0	E 23 E 83 E 13 E 103
E 23	E 104	SCZ	+104	0.16	18	0	E 13 E 99 E 43 E 143
E 33	E 151	SCZ	+201	0.39	18	0	E 13 E 145 E 83 E 157
E 23	E 171	SCZ	+11	0.73	7	0	E 13 E 169 E 43 E 173
E 63	E 168	SCZ	+13	1.19	7	0	E 53 E 168 E 83 E 169
E 73	E 123	SCZ	+7	-0.90	7	0	E 103 E 103 E 43 E 143
E 73	E 173	SCZ	-8	-1.13	6	0	E 83 E 163 E 43 E 103
E 113	E 273	BORDA	+37	-1.14	21	3	E 143 E 203 E 83 E 323
E 133	E 73	SCZ	-5	0.95	10	0	E 113 E 53 E 103 E 103
E 133	E 113	SCZ	+3	1.00	0	0	E 113 E 103 E 143 E 123
E 133	E 168	SCZ	-140	1.22	10	0	E 47 E 157 E 103 E 163
E 143	E 473	SCZ	-5	-0.26	7	0	E 153 E 443 E 133 E 493
E 143	E 173	SCZ	+5	1.02	12	0	E 123 E 171 E 163 E 174
E 153	E 13	SCZ	+8	1.57	10	0	E 103 E 13 E 193 E 13
E 173	E 153	SCZ	-4	1.19	7	0	E 153 E 143 E 103 E 153
E 173	E 176	SCZ	-4	1.57	9	0	E 133 E 176 E 213 E 174

Fig. 4.2 - Em (d) SCZ's, BARRAS e BORDAS para a cena com dois canais V^2G de $W = 6$ e 8 pixels interligados, em em (e), parte do arquivo gerado com as primitivas que descrevem a cena.

CAPÍTULO 5

OS EXPERIMENTOS E RESULTADOS OBTIDOS

5.1 - 1ª EXPERIÊNCIA : A UTILIZAÇÃO DO ESBOÇO PRIMÁRIO PARA DETECÇÃO AUTOMÁTICA DE LINEAMENTOS GEOLÓGICOS

O problema e a solução proposta

A análise geológica estrutural de uma determinada região, utilizando imagens de satélites, envolve a criação de mapas foto-interpretados mostrando as "falhas", "fraturas", "dobramentos", etc, das rochas, na região de interesse. Isto normalmente é feito pelos geólogos usando um produto fotográfico, gerado a partir do imageamento por satélite da área em estudo. O especialista apóia sobre a fotografia um tipo qualquer de papel transparente e, a cada identificação visual de um elemento geológico, ele marca com um segmento de reta que possui a *orientação e extensão* do elemento encontrado.

Estes segmentos de reta representam individualmente as características geológicas encontradas. São conhecidos como as *feições lineares* da área estudada. Um mapa das feições em todas as direções é montado. Para facilitar a interpretação desta informação, é derivado um diagrama a partir do mapa, chamado *roseta de direções*. A roseta plota para cada direção, a frequência com que ocorrem as feições, representantes principalmente de "falhas", "fraturas" ou "dobramentos". O mapa e o diagrama são construídos manualmente. Este é um trabalho tedioso, cansativo e demorado. Devido a estas características, têm sido feito um esforço para automatização destas tarefas (Abrams et al., 1985).

Dadas as características de definição do problema, pensamos em utilizar o ESBOÇO PRIMÁRIO BRUTO (EPB) pa

ra a detecção automática das feições lineares. Nossa proposta baseava-se em duas asserções principais:

(1) A representação proposta (EPB), basicamente detecta e representa variações de intensidade em uma cena. Ora, "falhas", "fraturas" e "dobramentos", aparecem na cena como claras variações na IMAGEM INTENSIDADE que a representa inicialmente. Estes elementos demarcam precisamente regiões de transição no "brilho" dos pontos. Sendo assim, certamente seriam "capturados" para alguma escala ∇^2G ou por um conjunto delas;

(2) A representação usada pelos geólogos na construção dos mapas, aproximava-se bastante da representação para as primitivas do EPB. Em ambos temos, simplificada mente, segmentos de retas orientados e limitados em tamanho.

Considerando as asserções, partimos para a experimentação. Utilizamos uma imagem [256 x 256] pixels, obtida por câmara de TV, a partir do produto fotográfico da banda 5 do LANDSAT TM. A área é uma região do estado de Minas Gerais, Brasil.

A melhor solução foi conseguida com os operadores ∇^2G de $W = 6$ e 10 pixels. Interligando estes canais geramos o EPB da região de estudo. A representação visual das primitivas do EPB, passam a representar o conjunto das feições lineares. O mapa de feições está então obtido.

Existe um interesse dos geólogos na determinação de concentrações de feições em uma determinada direção. Estas feições orientadas definem os *lineamentos geológicos*. É do estudo das feições e lineamentos que se procura caracterizar o terreno. Utilizando a informação de ORIENTAÇÃO de cada primitiva do EPB, podemos isolar, usando um simples fil

tro de direção, as primitivas nas orientações desejadas. Fazendo isto foi possível analisar visualmente a quantitativa mente a potencialidade do método.

A análise dos resultados

Na figura 5.1 apresentamos uma sequência de resultados.

Em (c) e (d) aparecem a detecção manual e automática das feições da cena para 45 graus. Para detecção automática obtivemos 412 feições. Na detecção manual obteve-se 397 feições, estabelecendo uma diferença percentual de 3.78 entre as detecções nesta direção. Os lineamentos ficam bem determinados visualmente tanto em (c) quanto em (d).

O problema de "ruídos", feições inexistentes ou excesso de feições foi encontrado. Isto já era esperado, uma vez que nenhum conhecimento adicional foi estabelecido, como é o caso no processo seletivo manual das feições. Desta forma, variações que não representavam reais eventos geológicos também foram "capturadas". Acreditamos que um trabalho simples, baseando-se nas outras propriedades das primitivas do EPB, e nas informações relativas ao momento de aquisição de imagem (já fornecidas na imagem de satélite da região), podem melhorar bastante esta característica ruidosa da detecção automática.

O nosso objetivo central com este experimento, foi demonstrar a potencialidade da extração automática, usando a abordagem simbólica para processar imagens de satélites. Os resultados obtidos nos entusiasma a trabalhar mais especificamente esta aplicação, objetivando uma operacionalidade para a extração automática de mapas de feições e geração das rosetas de direção.

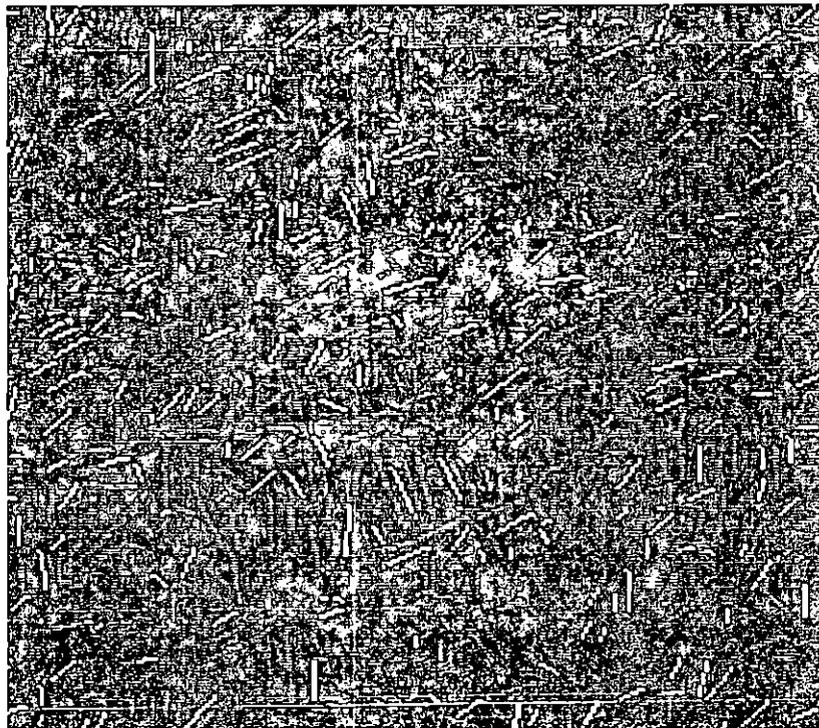
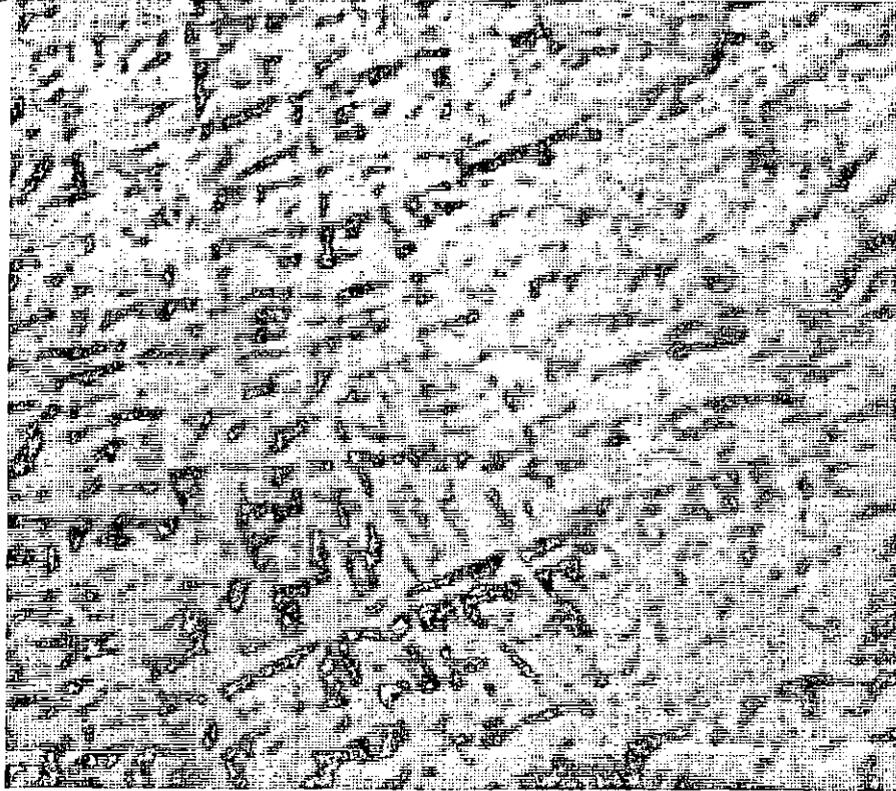


Fig. 5.1 - (a) Área TM estudada. (b) o EPB completo para a cena. As primitivas com maior intensidade (mais brancas) são as BORDAS. As de intensidade média são as BARRAS e as de menor intensidade são os SCZ's.

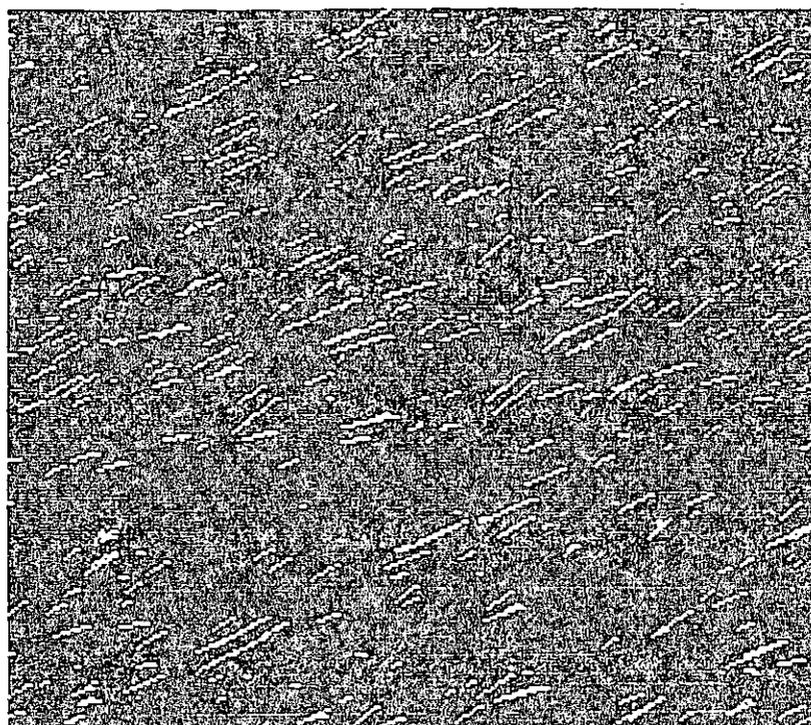
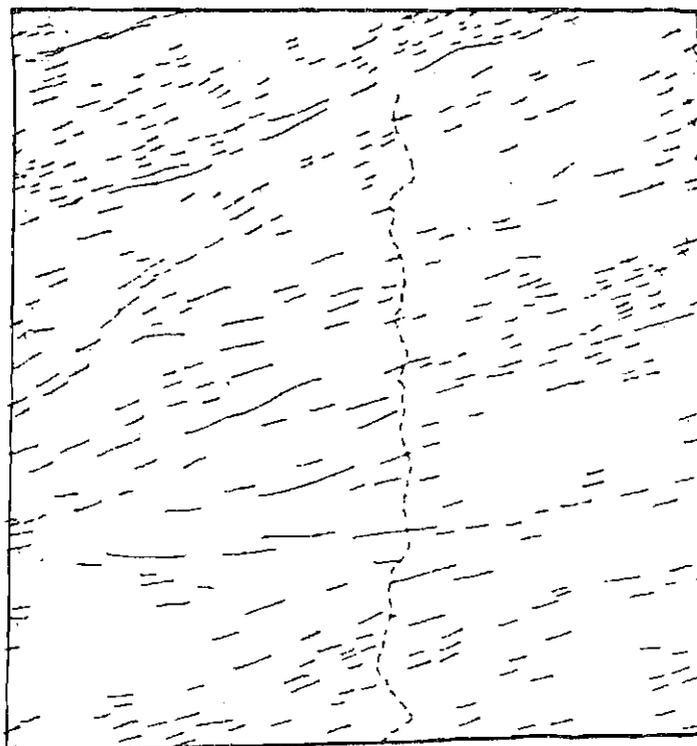


Fig. 5.1 - (c) O mapa de feições para 45 graus obtido ma
nualmente. (d) O mapa obtido automaticamente.

Mais experimentos serão feitos, e a comparação com dados manuais será nosso parâmetro para controlar a qualidade da detecção automática, usando a abordagem aqui proposta. Além disto, operações de seleção e agrupamento das primitivas podem ser determinadas. Com isto podemos isolar primitivas do conjunto geral, que é o EPB, e auxiliar o geólogo na identificação mais precisa do que seja uma "falha" ou "fratura" ou "dobramento" nas rochas daquela região. As informações do arquivo de primitivas, podem também servir como base primária de conhecimento, para um sistema mais inteligente que, além de detectar e representar as feições, possui algum grau de compreensão da cena.

5.2 - 2ª EXPERIÊNCIA : A SEGMENTAÇÃO PRIMITIVA DE UMA CENA

No processamento de imagens de satélites, o propósito da segmentação é particionar a imagem analisada em "regiões significativas" para compreensão global da cena. As "regiões significativas" são uma função direta do problema que está sendo resolvido. Pela característica da representação para imagens em sistemas de processamento digital de imagens de satélites (PDI), os processos para encontrar estas "regiões" têm sido, basicamente, de dois tipos:

(1) métodos que se baseiam no exame ponto a ponto da imagem;

(2) métodos que se baseiam na informação extraída para uma "vizinhança" bem definida, para os pontos na imagem.

Ambos os processos, praticamente utilizam uma única variável, a informação de intensidade de cada pixel. Esta simplificação do problema, tem levado a um tratamento puramente estatístico na implementação da segmentação. Com a tecnologia dos sistemas especialistas (sistemas peritos,

expert systems), tem-se buscado melhorar os resultados dos processos segmentadores. Para isso, um conhecimento a priori é estabelecido para auxiliar a partição da cena. Mantendo-se a mesma representação inicial da cena, a IMAGEM INTENSIDADE, busca-se sofisticar os processos de controle da informação para se chegar a melhores resultados.

Nossa proposição é usar o ESBOÇO PRIMÁRIO BRUTO (EPB), como a representação inicial sobre a qual operaremos transformações. Após a extração do EPB, o que temos é um conjunto de primitivas simbólicas confusamente distribuídas, definindo diversos limiares de regiões e/ou objetos. Veja Fig. 5.2. Note que a definição do contorno do rosto é uma entidade visualmente isolada, do ponto de vista da percepção da cena global, assim como os olhos e o nariz.

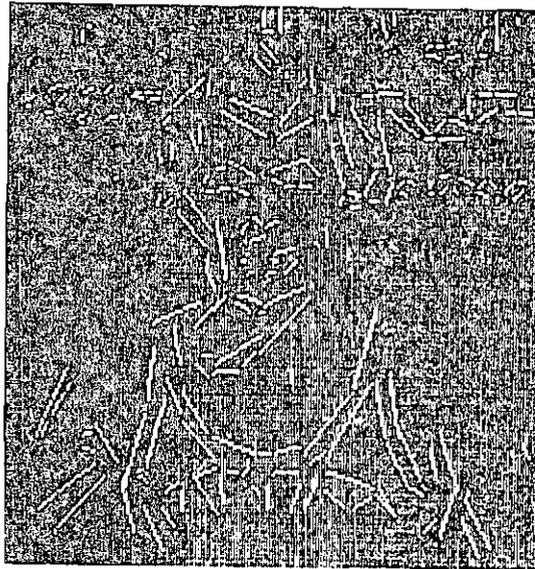


Fig. 5.2 - O EPB com $\nabla^2 G$ de $w = 6$ e 8 pixels.

Estas primitivas que definem o EPB possuem, como já vimos, um conjunto de atributos. Estes foram extraídos a partir da cena original, e buscam refletir características físicas da cena imageada. Na Fig. 5.3 temos as estru

turas dos descritores para um SEGMENTO CRUZAMENTO POR ZERO (SCZ); uma BARRA e uma BORDA.

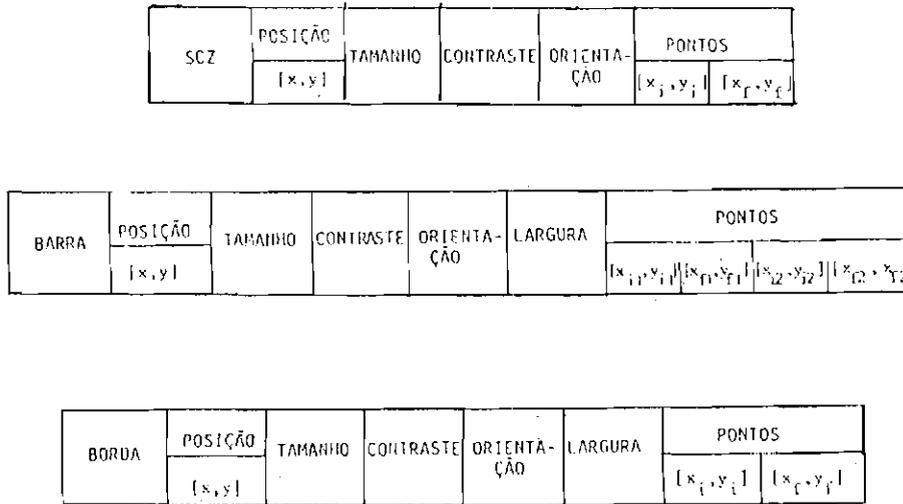


Fig. 5.3 - (a) Um SCZ. (b) Uma BARRA. (c) Uma BORDA.

Os PONTOS limitam os segmentos de reta. São os pontos de início e fim do segmento que define a primitiva. Os outros atributos foram discutidos no capítulo 4.

Observe na Fig. 5.2 que nosso sistema visual é capaz de distinguir "lotes" de primitivas, em deferentes regiões, que parecem se agrupar e definir "contornos" na cena. Conseguindo determinar dentro de nossa confusa massa de primitivas (o EPB), estruturas maiores resultado da *seleção e do agrupamento* das primitivas básicas podemos gerar diversos "quadros" isolados da cena. Cada "quadro" conteria algumas destas estruturas. Estas estruturas definiriam regiões e/ou objetos na cena. Ora, este resultado é uma forma de SEGMENTAÇÃO da cena.

Com base nestas idéias implementamos nosso processo de segmentação. Não adotamos conhecimento de contexto

ou do especialista. A partição da cena em "regiões significativas", se dá baseando-se apenas na informação possível de extração a partir da IMAGEM INTENSIDADE e dos processos de SELEÇÃO e AGRUPAMENTO definidos sobre o EPB. Considerando que o EPB gerado descreve, simbólica e estruturalmente, a cena original.

Os processos de agrupamento e seleção

A geometria euclidiana estabelecia que o todo seria sempre igual a soma de suas partes. Nos anos 20 e 30, um grupo de psicólogos alemães, estudando experimentalmente os processos visuais de percepção das cenas, determinava um novo conceito: "O todo perceptual é sempre diferente que a simples soma de suas partes". Eles notaram que a percepção das cenas, sempre indicava a realização de processos ativos de seleção e agrupamento entre os elementos que a compunham. Estes processos determinavam, que não se podia tratar as partes como entidades perceptivas isoladas e disjuntas (Frisby, 1978). Este grupo estabeleceu uma escola, a GESTALTISTA, vinda da palavra alemã "gestalt", que pode ser traduzida por "forma" ou "configuração". Foram sobre os mecanismos que nos levam a percepção das "gestalts", que estes psicólogos trabalharam.

O interesse nestes estudos, é que os princípios de agrupamento descobertos pela escola gestaltista, são de uso para os processos de segmentação sobre o EPB. Os processos de agrupamento vão agir sobre a massa de primitivas, que é o EPB. Para facilitar o trabalho, temos uma escolha anterior ocorrendo no conjunto de primitivas. Esta escolha estabelece as primitivas mais prováveis para os posteriores processos de agrupamento. O conjunto SELEÇÃO + AGRUPAMENTO, agindo sobre o EPB, vai definir a SEGMENTAÇÃO da cena, para a aplicação de interesse.

O problema

Para realizar um experimento, testando a via utilização destas idéias, escolhemos o problema de isolar em uma imagem de satélite, de uma determinada região, um rio e seus possíveis afluentes. Utilizamos uma imagem do LANDSAT TM, da região de Assis, no estado de São Paulo, Brasil. Da de às restrições do hardware, no momento, nossa imagem teve dimensões reduzidas.

A solução

Primeiro definimos os canais V^2G a serem usa dos na geração do EPB. Chegamos a $W = 6$ e $W = 10$ pixels. O canal $W = 4$ mostrou-se bastante ruidoso, e $W = 8$ continua muita repetição da informação em $W = 6$. Após gerado o EPB, passamos a trabalhar sobre o conjunto de primitivas. A cena deixou de ser a IMAGEM INTENSIDADE de entrada, e passou a ser representada pela lista de descritores, que representam simbolicamente as estruturas físicas da cena.

Utilizando critérios de seleção, buscamos eli minar as primitivas do EPB não capacitadas como possíveis ele mentos constituintes dos contornos do rio. Estes critérios se basearam no princípio da SIMILARIDADE. Ele fica estabele cido com as propriedades de TAMANHO e CONTRASTE de cada pri mitiva.

Uma informação do TAMANHO MÉDIO e CONTRASTE MÉDIO é obtida no canal maior que forma o par gerador do EPB (No caso V^2G com $W = 10$ pixels). Neste canal, baseado na suposição de organização hierárquica (descrita no capítulo 3) , temos maiores estruturas refletindo a organização espacial da função de reflectância das superfícies na cena. Certamen te, os grandes contornos que definem o rio devem estar pre sentes neste canal. Com base na "similaridade" entre os atri

butos das primitivas do EPB e as médias obtidas, *seleccionamos* os possíveis candidatos aos contornos do rio.

Depois de seleccionadas as primitivas, passamos a nos preocupar em agrupá-las. Para esta tarefa usamos o conceito da CONTINUIDADE.

A CONTINUIDADE neste caso, está ligada a PROXIMIDADE ESPACIAL e ao atributo ORIENTAÇÃO das primitivas do EPB. Veja Fig. 5.4.

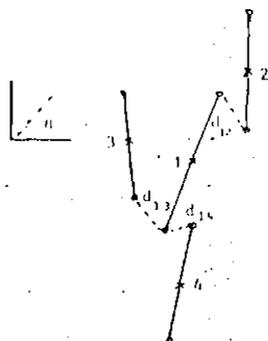


Fig. 5.4 - Algumas primitivas do EPB orientadas no espaço.

Temos (1) como a primitiva sendo verificada, e (2), (3) e (4) possíveis candidatos para a sua "continuação". As distâncias entre (1) e seus vizinhos são obtidas. O elemento que possuir a menor distância (d) e a menor diferença entre as suas ORIENTAÇÕES será o escolhido. No caso da Fig. 5.4, o mais provável é que (4) seja o escolhido, e a partir dele repete-se o procedimento.

Utilizando este conceito, vamos montando uma lista encadeada com todas as primitivas que são validadas. A lista gerada será o produto final e deve ser a imagem de "bordas" do rio e seus afluentes. Na Fig. 5.5. estão os resultados obtidos.

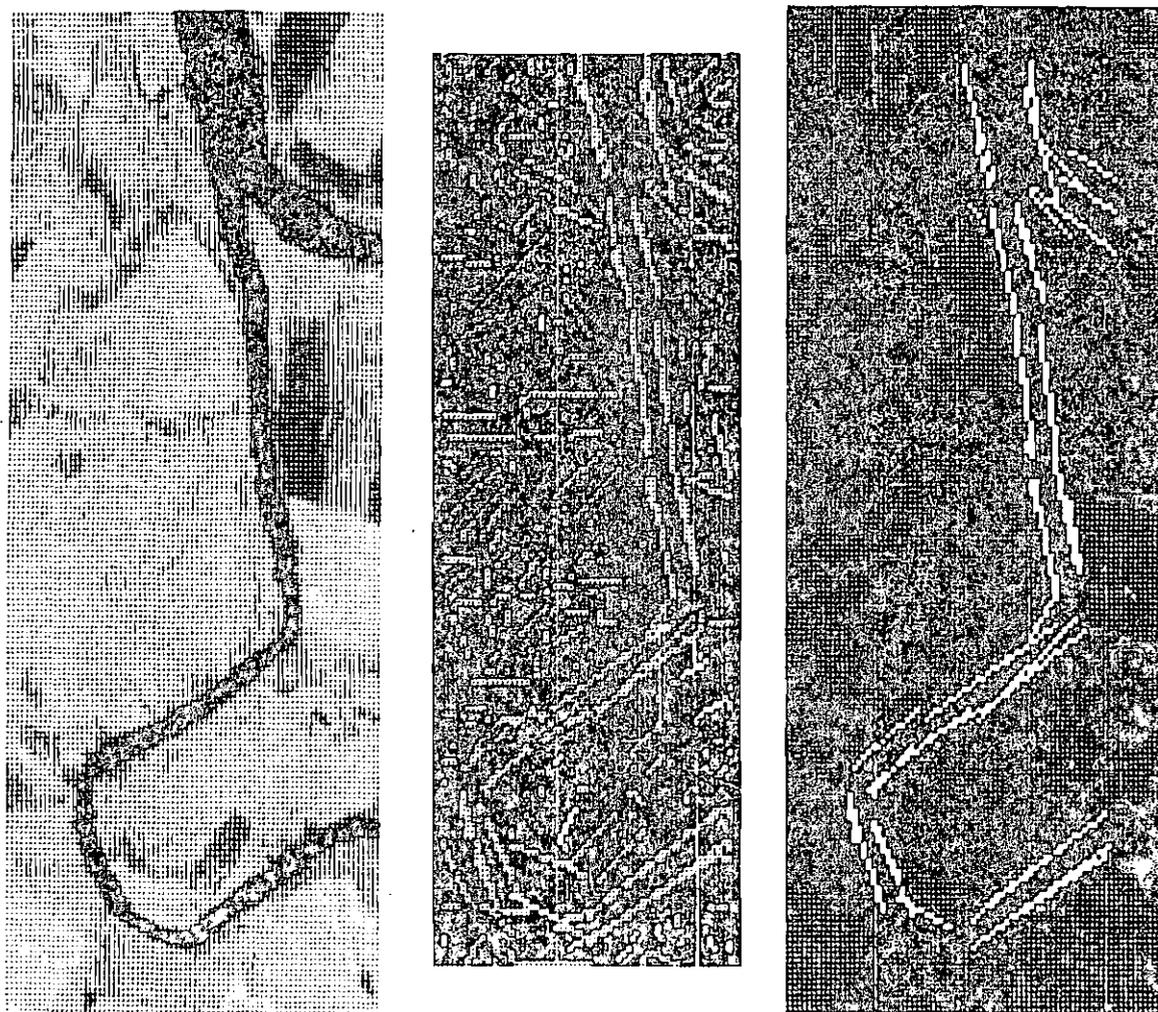


Fig. 5.5 - (a) Imagem TM original. (b) EPB com ∇^2G de $W = 6$ e 10 pixels. (c) Imagem segmentada, sobreposta a original, utilizando critérios de seleção e agrupamento.

Análise dos resultados

Mostrou-se possível a realização da tarefa de segmentar, para uma aplicação definida, utilizando o EPB e portanto algum tipo de processamento simbólico. Nenhum co

nhecimento adicional, seja do contexto ou de especialista foi necessário. As restrições se basearam em suposições gerais sobre o mundo visível. O tempo de processamento ainda é um problema. No entanto, é bom ressaltar que nossa implementação é serial e sem otimizações. Todos os processos são essencialmente locais e se adequariam a implementação em hardware paralelo, o que melhoraria sensivelmente o tempo de respostas. O experimento demonstra a potencialidade do modelo para a tarefa de segmentação, indicando uma nova possibilidade a seguir quando existe dificuldades em segmentar usando critérios puramente estatísticos.

5.3 - 3ª EXPERIÊNCIA : A DISCRIMINAÇÃO DE TEXTURAS

A análise de textura tem-se mostrado fundamental para a percepção de regiões e superfícies em imagens. Por exemplo, nós facilmente distinguimos entre um tecido de lã e um de algodão, na mesma coloração. Propriedades de textura são sempre usadas pelos especialistas, direta ou indiretamente, quando classificando as diferenciadas regiões em uma imagem. Este fato levou a muitos estudos e algoritmos propondo reconhecimento e descrição de texturas por computador. No entanto, como poucas definições de textura têm sido dadas, e nenhuma precisamente define o problema, as técnicas de discriminação da textura são grandemente "ad hoc" (Brady, 1982).

Duas são as abordagens principais para a discriminação de textura: a estatística e a estrutural. As propriedades para a caracterização da textura, têm sido derivadas no domínio estatístico, usando principalmente matrizes de concorrência. (Em Haralick et al. (1973) e Connors et al. (1984) este tipo de análise se encontra bem descrito).

Processos mais recentes, sugerem que a extração de características locais de textura, e as relações en

tre as regiões que ficam determinadas por estas características, levariam a definição de elementos de textura, ou primitivas de textura. Estas primitivas teriam papel central na percepção de regiões e/ou superfícies de texturas diferenciadas.

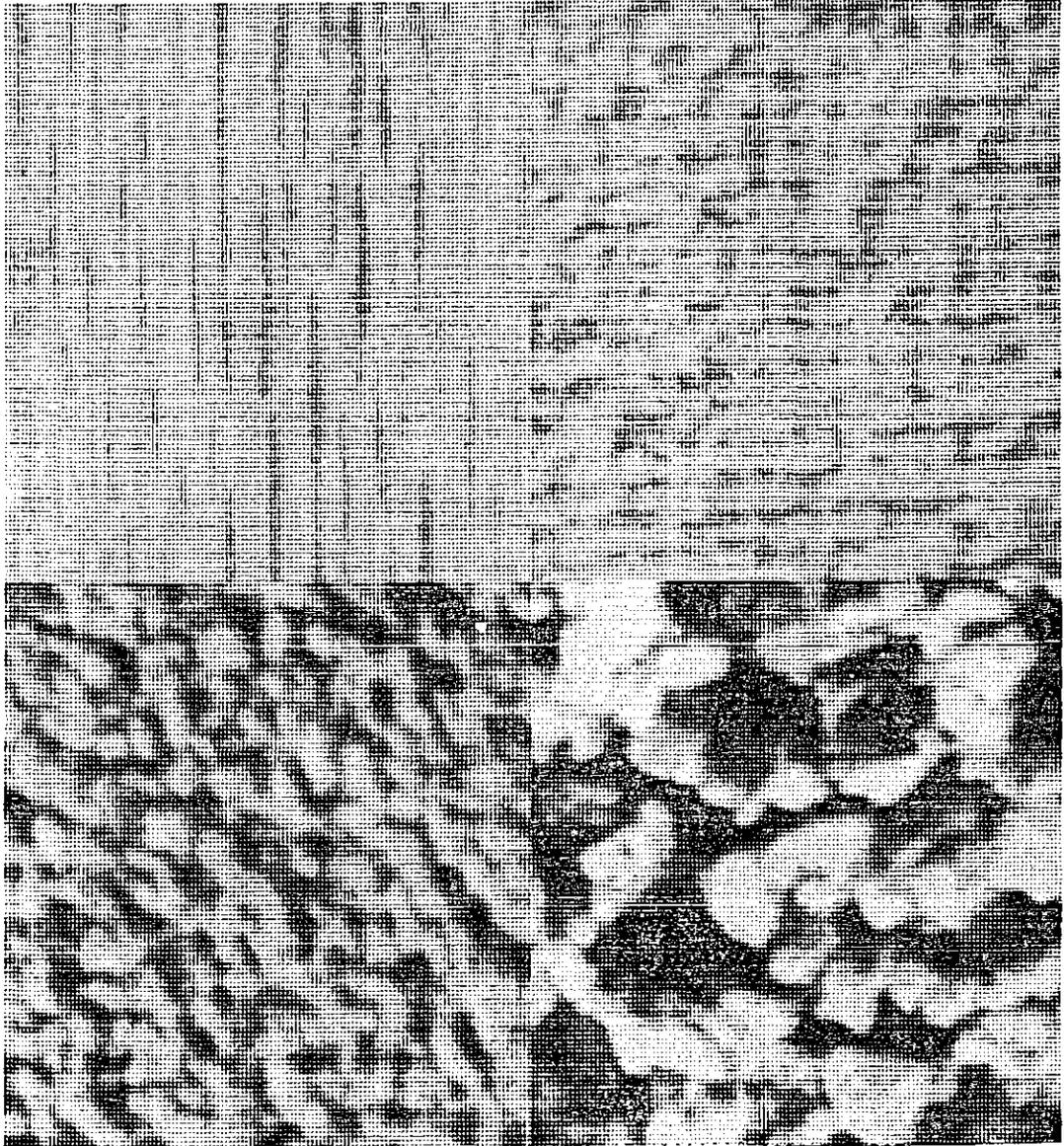
A idéia de elementos primitivos de textura, casa-se bem com a proposta do ESBOÇO PRIMÁRIO. Nesta implementação, nenhuma primitiva específica para a textura foi estudada. No entanto, a textura se caracteriza por algum tipo de "regularidade" nos elementos primitivos que compõem a cena. A repetição ou quase-repetição dos elementos fundamentais, podem caracterizar uma região e/ou superfície. As características de organização destes elementos também podem fazê-lo.

Pensando textura desta forma, podemos considerar os elementos primitivos do ESBOÇO PRIMÁRIO BRUTO (EPB) como nossos elementos de textura. Usando, então, estatísticas simples (de primeira ordem) sobre o EPB, podemos discriminar texturas, baseando nossa análise nas relações espaciais e nas propriedades de cada primitiva básica do EPB, ou de um determinado canal V^2G .

Para testar estas considerações, utilizamos imagens de texturas obtidas com uma câmara de TV, a partir do Brodatz (1968), com tamanhos de [128 x 128] pixels. Fizemos dez (10) amostras de cada textura. Na fig. 5.6 estão apresentadas uma das amostras de cada uma das imagens trabalhadas.

(a)

(b)



(c)

(d)

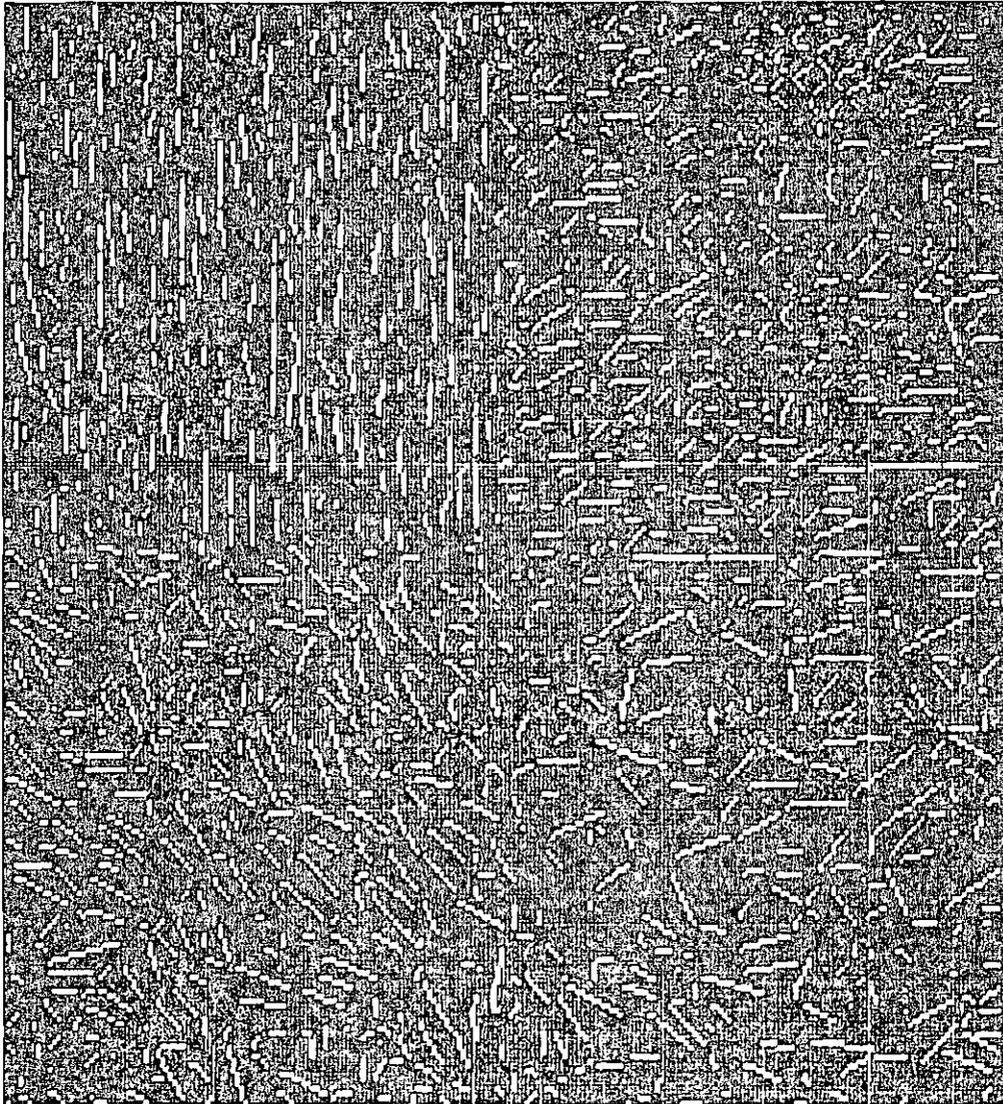
Fig. 5.6 - As imagens originais, retiradas de Brodatz (1986) de (a) veio (b) rãfia, (c) cortiça e em (d) mica.

Para cada tipo de textura, uma descrição simbólica isolada para cada canal ∇^2G com $W = 4, 6, 8$ e 10 pixels foi obtida. Escolhemos para trabalhar os canais com $W=6$ e 8 pixels, visto que em $W = 4$ pixels, houve uma detecção bastante ruidosa e em $W = 10$ havia pouca informação. Sobre estas descrições, verificamos as distribuições de TAMANHO e ORIENTAÇÃO para as primitivas encontradas. A discriminação das texturas, baseia-se na análise destas distribuições.

No canal com $W = 6$ pixels, o veio e a rãfia são diretamente separáveis observando a distribuição do TAMANHO. O veio apresenta um histograma com pico em (18) em uma dada orientação preferencial (no caso ≈ 90 graus), enquanto a rãfia apresenta picos em (18) e (4) com duas orientações preferenciais defasadas de aproximadamente 180 graus. Estas configurações das primitivas, definido estruturas diferenciadas para a rãfia e para o veio, discriminam visualmente as texturas e podem ser quantizadas. Veja na Fig. 5.7 as descrições (conjunto de primitivas geradas) para o canal $W = 6$ pixels e na Fig. 5.8 as distribuições para o veio e a rãfia.

(a)

(b)



(c)

(d)

Fig. 5.7 - Primitivas para (a), o veio, (b), a rãfia, (c) a cortiça e, em (d), a mica.

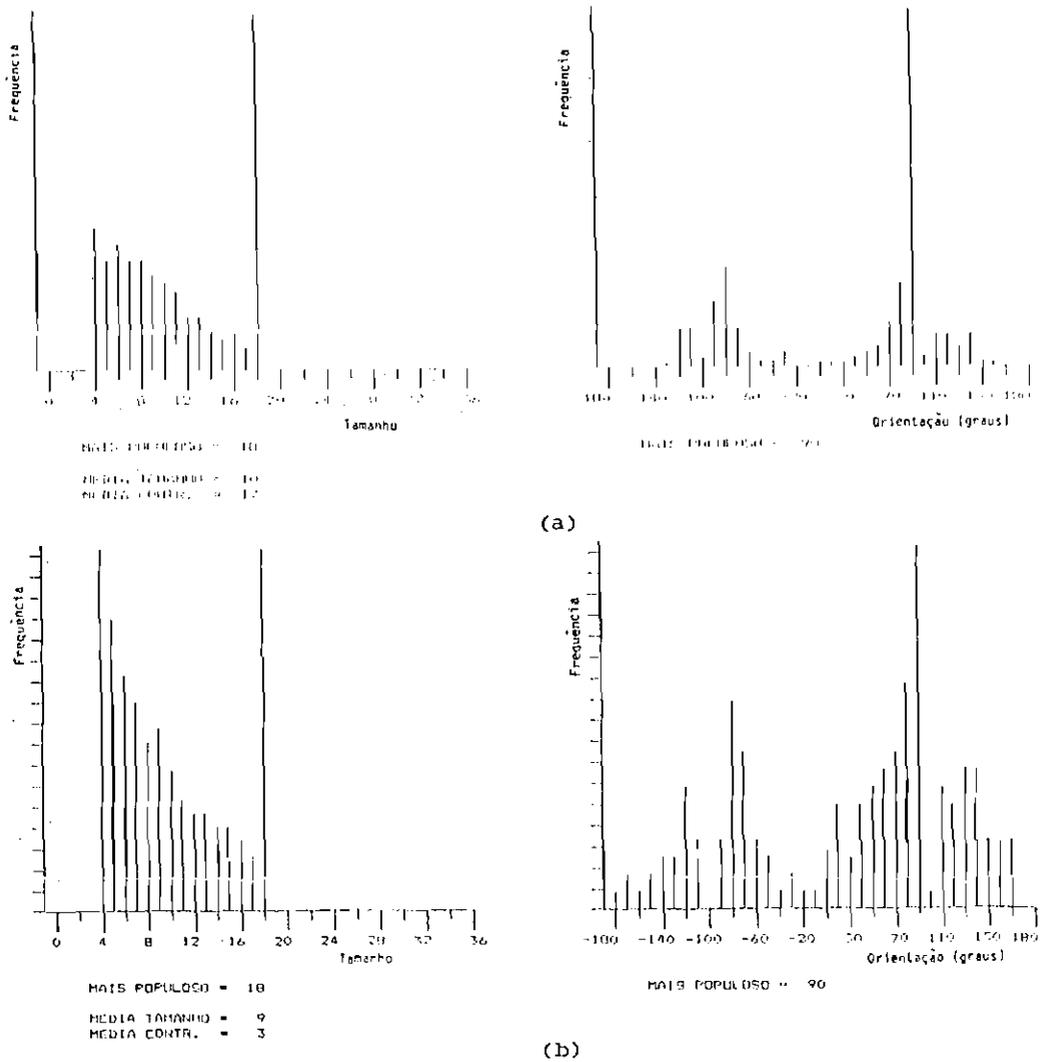


Fig. 5.8 - Distribuições de TAMANHO e ORIENTAÇÃO para (a), o veio, e (b), a rãfia.

O veio e a rãfia se diferenciam diretamente da cortiça, também observando a distribuição de TAMANHO. A cortiça tem concentração de primitivas com tamanho (4), sem nenhuma orientação preferencial. As primitivas distribuem-se em todas as direções, não se podendo determinar alinhamentos direcionais. Esta organização caracteriza a textura, diferenciando-a do veio e da rãfia. Na Fig. 5.9 apresentamos as distribuições de TAMANHO e ORIENTAÇÃO da cortiça para o canal ∇^2G com $W = 6$ pixels.

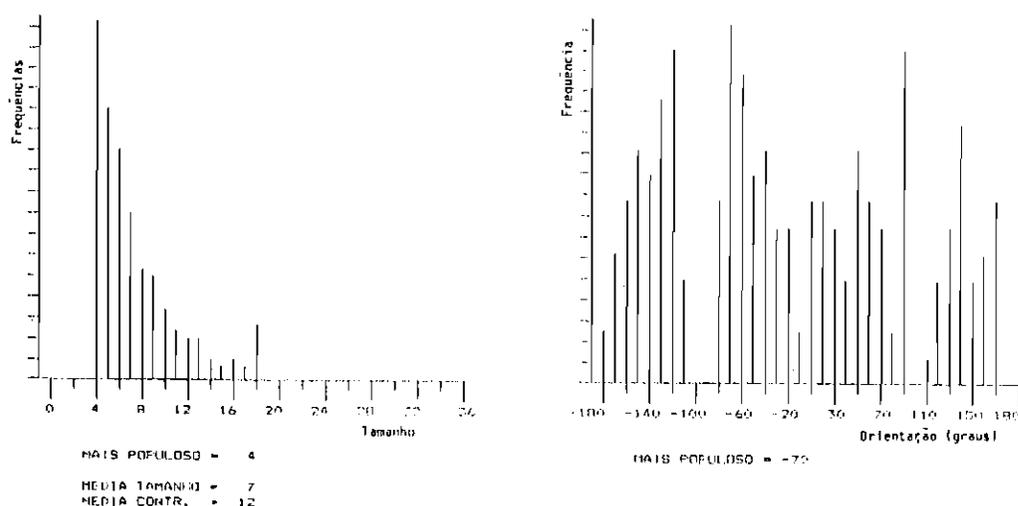


Fig. 5.9 - Distribuição em TAMANHO e ORIENTAÇÃO.

A mica diferencia-se do veio e da rãfia pelas mesmas características apresentadas em relação a cortiça. Na Fig. 5.10 tem-se as distribuições da mica para o canal ∇^2G com $W = 6$ pixels.

Aparece, então, um problema. Como diferenciar a mica e a cortiça. Observando as imagens originais de ambas na Fig. 5.1, é fácil verificar que a mica é semelhante a cor

tiça, sô que em uma escala maior. Visto isto, tentamos estabelecer um novo canal ∇^2G e verificar os resultados. Utilizando ∇^2G e verificar os resultados. Utilizando ∇^2G com $W = 8$ pixels obtivemos os resultados apresentados na Fig. 5.11.

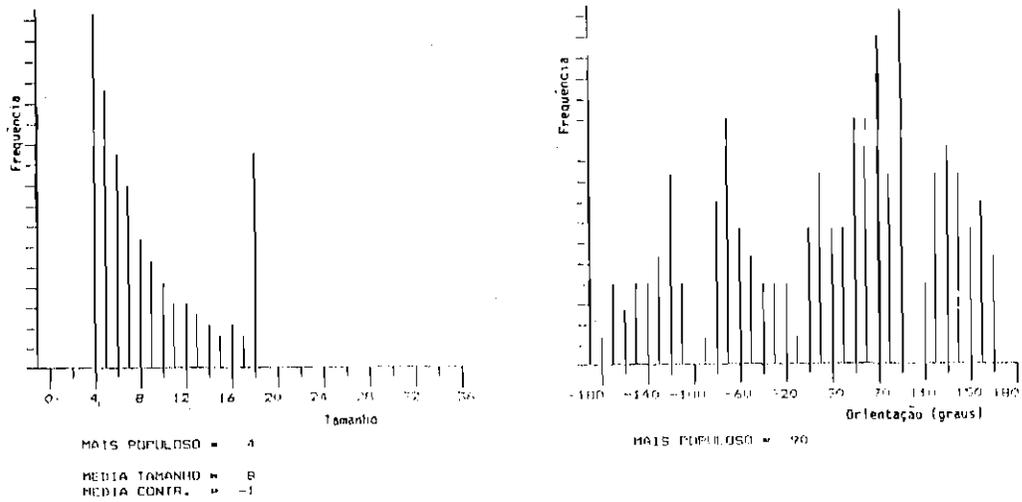


Fig. 5.10 - Distribuição em TAMANHO e ORIENTAÇÃO.

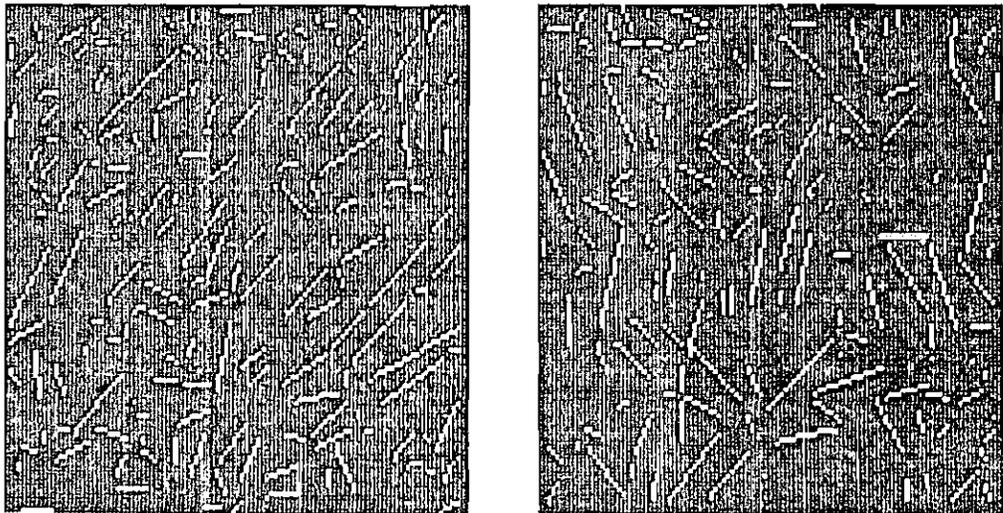


Fig. 5.11 - Descrição com $W = 8$ pixels para ∇^2G em (a) cortiça e (b) mica.

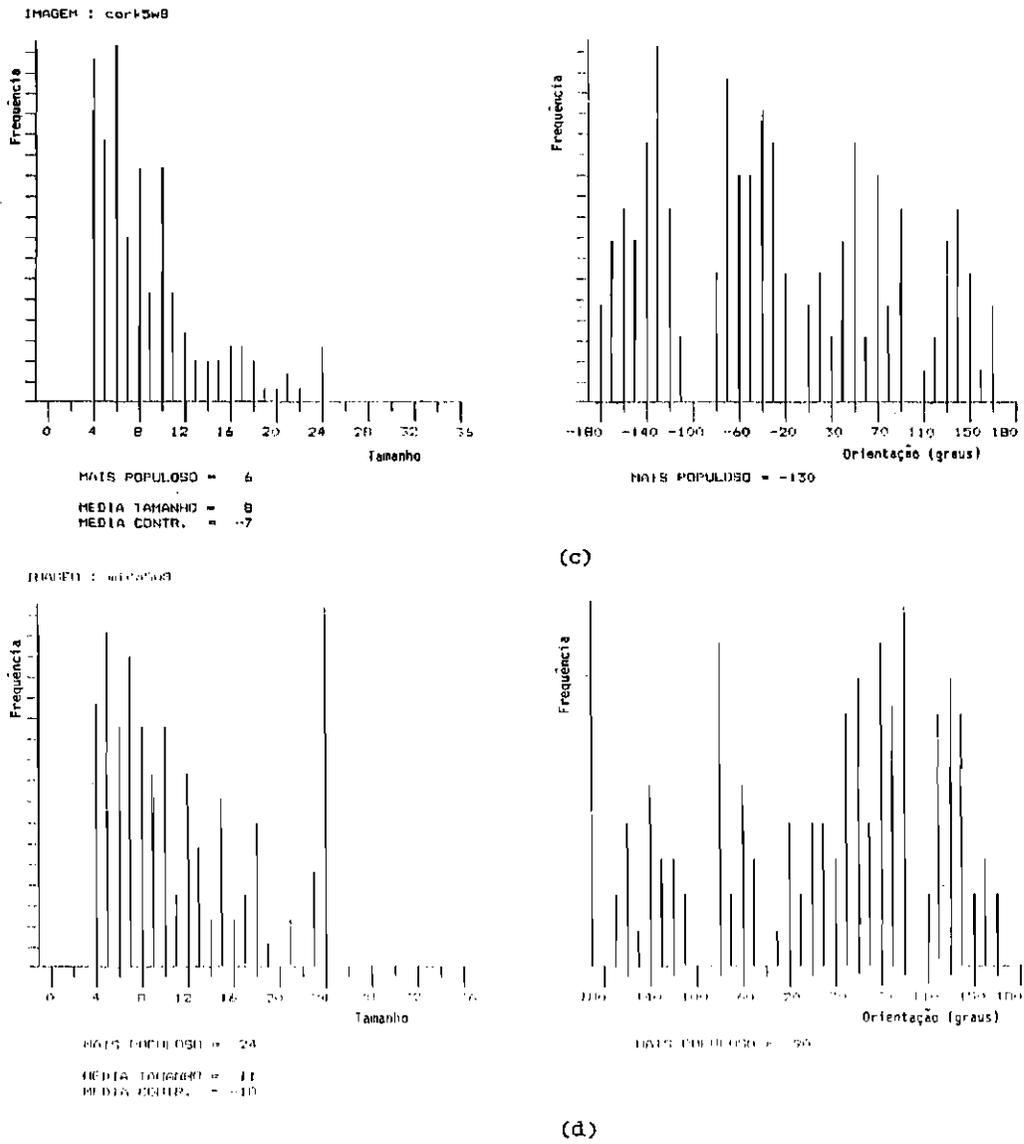


Fig. 5.11 - Distribuições de TAMANHO e ORIENTAÇÃO em (c) cortiça e (d) mica, para $W = 8$ pixels.

Para este canal, conseguimos determinar uma nova caracterização para as texturas. A distribuição de TAMANHO para a mica, passou a definir uma organização com primitivas de tamanho (24), formando a base de percepção da cena. A cortiça, continuou definindo primitivas de tamanhos entre (4) e (6) como sua base perceptiva. Ambas sem preferências direcionais, portanto não definindo alinhamentos. Esta diferenciação, possibilitada pela nova escala do operador $\nabla^2 G$, torna possível discriminar a mica da cortiça.

A análise dos resultados

No escopo deste trabalho, objetivamos com estes experimentos, estabelecer empiricamente, a possibilidade de uso do modelo proposto para caracterizar texturas. Fazendo isto, trabalhando somente estatísticas simples sobre a representação simbólica da cena. Nosso objetivo posterior é tentar a segmentação da imagem, baseando nosso algoritmo na extração de características de textura da cena, utilizando esta abordagem aqui apresentada. Um EPB local na imagem, forneceria medidas baseadas nas distribuições, de maneira a caracterizar quantitativamente a textura na região. Trabalhando então com grafos, onde os nós representariam regiões com mesmas características de textura, buscaríamos definir regiões ou limiares entre regiões.

O experimento realizado indica a viabilidade em tentar esta abordagem para segmentação por textura e, certamente será uma das continuações deste trabalho. Com isto, estaremos utilizando uma representação simbólica da cena para executar tarefas tradicionais em processamento digital de imagens de satélites e demonstrar esta possibilidade era o que pretendíamos com estes experimentos.

CAPÍTULO 6

CONCLUSÃO

A orientação principal deste trabalho está ligada ao processo evolutivo dos sistemas de tratamento de imagens de satélites. A perspectiva atual aponta para a geração de sistemas com algum grau de compreensão das cenas analisadas. Busca-se automatizar o trabalho de interpretação das cenas ou, pelo menos, auxiliar o especialista da área a identificar e resolver mais facilmente seus problemas.

A abordagem escolhida, tratou principalmente do problema relacionado à representação inicial das cenas para análise. Apontou a existência de um processamento inicial sobre a informação visual, independente dos processos de aprendizado e experiência. O reconhecimento primitivo das formas na cena, o nível mais primário dos processamentos visuais, seria estritamente dirigido pelas informações extraídas a partir da cena original. O processamento de dados visuais, seria já no seu nível mais baixo, um processamento simbólico, baseado então, em uma representação simbólica e estrutural para a informação a ser tratada.

Desta forma, experimentos foram definidos, para que se verificasse a viabilidade de realizar tarefas de processamento digital de imagens de satélites, usando como entrada, uma representação diferente da IMAGEM INTENSIDADE, tradicionalmente utilizada.

Os três experimentos realizados, discutidos no capítulo 5, demonstram claramente a potencialidade da abordagem proposta. O tempo de processamento é ainda um problema real, quando se pensa na viabilidade operacional do modelo. No entanto, dada as características locais do proces

samento que envolve a geração do ESBOÇO PRIMÁRIO BRUTO (EPB), um hardware não convencional, com capacidade de processar in formações concorrentemente, pode melhorar em muito a resposta em tempo do sistema. Mesmo considerado este fato, uma das possíveis aplicações apresentadas, e detecção e representação automática de lineamentos geológicos pode, já neste estágio, somente com uma melhora na capacidade de armazenamento, apresentar resultados semi-operacionais.

Os experimentos abriram diversos caminhos, que devem ser seguidos, dando continuidade à orientação que este trabalho apresenta. A segmentação, usando uma caracterização estrutural para a textura das imagens, e a segmentação baseada em critérios de seleção e agrupamento das primitivas simbólicas do EPB, para áreas de difícil solução usando critérios puramente estatísticos, são alguns caminhos a seguir.

É certo que esta um sistema compreender cenas, não temos como única componente a cognição, dos níveis mais altos de processamento. Existem operações sendo efetuadas a nível da imagem de entrada, que geram informação global para a descrição final da cena. Os níveis intermediários e altos dos processamentos de dados visuais, não são discutidos aqui. Fica claro, porém, que a relação destes níveis com os processos visuais primários, torna-se facilitada se a representação inicial, já é uma representação mais adequada aos processamentos simbólicos posteriores. A segmentação do rio, no capítulo 5, demonstra este fato.

A abordagem utilizando modelos de percepção , largamente motivados por estudos do sistema visual humano, é uma opção excitante e bastante interessante. A compreensão das teorias computacionais envolvidas nos processos visuais primários é fundamental para criação de sistemas automatizados de visão, com eficiência semelhante a sistemas visuais

biológicos. No entanto, os argumentos apresentados neste trabalho, mesmo que motivados pelos estudos neurofisiológicos, psicofísicos e psicológicos, são sustentados do ponto de vista puramente matemática, físico e computacional, como apresentam os capítulos 3 e 4.

Os resultados experimentais, demonstram a adequação do modelo para a realização de tarefas básicas de processamento de imagens de satélites. Dada as características próprias a este tipo de imagens, modificações devem ser introduzidas nas primitivas da representação. Uma possibilidade, é introduzir o atributo de "COR", como mais uma propriedade da descrição simbólica. Este atributo contribuiria para melhorar os processos de segmentação, e pode ser obtido a partir de um tratamento relacionando as bandas espectrais das imagens de satélites.

Finalizando, o que temos é uma ferramenta que auxilia o trabalho com este tipo de imagens, quando se torna muito complicado ou mesmo impossível, o tratamento puramente estatístico para a solução de algum problema. Além disto, temos uma representação inicial, que certamente facilitará os processos posteriores de compreensão global das cenas, dada a quantidade de informação sobre a estrutura física das mesmas, e a representação simbólica explícita destas propriedades

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAMS, M.; BLUSSON, A.; CARRERE, V.; NGUYEN, T.; RABU, Y.;
Image processing applications for geologic mapping. *IBM
Journal of Research and Development*, 29 (2):177-187, mar.1985.
- BARROW, H.G.; TENENBAUM, J.M. *Recovering intrinsic scene charac-
teristics from images*. In: HANSON, A.R.; RISEMAN, E.M.; ed.
Computer vision systems. New York, Academic, 1978, p. 3-25.
- BARROW, H.G.; TENENBAUM, J.M. Computational Vision. *Proc. of
the IEEE*, 69(5):572-595, May, 1981.
- BRADY, M. Computational approaches to image understanding.
Computing Surveys, 14(1):3-71, Mar, 1982.
- BROW, C.M. Computer vision and natural constraints. *Science*,
224(4655):1299-1304, June, 1984.
- CONNERS, R.W.; TRIVELDI, M.M.; HARLOW, C.A. Segmentation of
a high-resolution urban scene using texture operators.
Computer Vision, Graphics and Image Process, 25:273-310, 1984.
- CORNSWEET, T.N. *Visual perception*. New York, Academic, 1970.
- FISCHLER, M.A. *On the representation of natural scenes*. In: HANSON,
A.R.; RISEMAN, E.M., ed. *Computer vision systems*. New York
Academic, 1978, p. 41-51.
- FORGUS R.H. *Percepção. O processo básico do desenvolvimento cognitivo*
São Paulo, Editora Herder, 1971.
- FRISBY, J.P. *Seeing. Ilusion, brain and mind*. Oxford, University
Press, 1979.
- GONZALEZ, R.C.; WINTZ, P. *Digital Image Processing*. New York,
Addison-Wesley, 1977.
- HARALICK, R.M.; SHANMUGAN, K.; DINSTEN, I. Textural feature
for image classification. *IEEE Trans. Sys. Man. Cyb.*, 3(s/f):
610-621, 1973.

- HILDRETH, E.C. The detection of intensity changes by computer and biological vision systems. *Computer Vision, Graphics, and Image Processing*, 22(s/f):1-27, 1983.
- HILDRETH, E.C. *Edge detection*. Local, MIT, AI Lab, 1985 (AI LAB. Tech. Rep. 858).
- HILDRETH, E.C. *Implementantion of a theory of edge detection*. Local, MIT, AI Lab., 1980. (AI Lab. Tech. Rep. 579).
- LINDSAY, P.H.; NORMAN, D.A. *Human information processing*. New York, Academic Press, 1977.
- MARR, D. Artificial intelligence - A personal view. *Artificial Intelligence*, 9(s/f):37-48, 1977.
- MARR, D. Early processing of visual information. *Phil. Trans. R. Soc. London*, B275(57):483-519, Oct. 1976.
- MARR, D. *Representing a visual information - a computacional approach*
In: HANSON, A.R.; RISEMAN, E.M., ed. *Computer vision systems*. New York, Academic, 1978.
- MARR, D. *Vision*. San Francisco, W.H. Freeman and company, 1982.
- MARR, D. Visual information processing: the structure and creation of visual representations. *Phil. Trans. R. Soc. London*, B290(s/f):199-218, 1980.
- MARR, D.; HILDRETH, E.C. *Theory of edge detection*. Proc. Royal Soc. London, B207, p. 187-217, 1980.
- MARR, D.; ULLMAN, S.; POGGIO, T. Bandpass channels, zero-crossings, and early visual information processing. *J. Opt. Soc. Am.*, 69(6):914-916, June, 1976.
- MASCARENHAS, N.D.A.; VELASCO, F.R.D. *Processamento digital de imagens*. São Paulo, Editora da Universidade de São Paulo, 1984.
- POGGIO, T. Early vision: from computational structure to algorithms and parallel hardware. *Computer Vision Graphics, and Image Processing*, 31(s/f): 139-155, 1985.

- POGGIO, T. Vision by man and machine. *Scientific American*, 250 (s/f):105-116, Apr. 1984.
- POGGIO, T.; TORRE, V.; KOCH, C. Computational vision and regularization theory. *Nature*, 317(s/f):314-319, Sept. 1985.
- RICH, E. *Artificial intelligence*. New York, McGraw Hill, 1984.
- TENENBAUM, J.M.; FISCHLER, M.A.; BARROW, H.G. Scene modelling: a structural basis for image description. *Computer Vision, Graphics and Image Processing*, 12(s/f):407-425, 1980.
- ULLMAN, S. Analysis of visual motion. *Computer*, 14(8):57-69, Aug. 1981.
- VICKERS, D. *Decision processes in visual perception*. New York, Academic Press, 1979.