

# Processamento Geométrico de Imagens

Cicero Mota

14 de outubro de 1999

À IVANILDE,  
sem cuja presença este trabalho não teria sido possível.

# Resumo

Neste trabalho estudamos o problema de reconstrução de uma imagem a partir de uma segmentação perceptual baseada na classificação geométrica de seus pontos. A classificação dos pontos é feita usando operadores não-lineares baseados nas curvaturas da superfície de Monge da imagem. É apresentada uma prova teórica de que uma imagem pode ser reconstruída a partir de seus pontos de curvatura Gaussiana não-nula. Este resultado dá fundamento teórico ao método de processamento não-linear de sinais bidimensionais proposto por C. Zetzche, E. Barth e B. Wegmann. São usados filtros de curvaturas para a detecção de arestas e vértices de uma imagem e mostramos que é possível reconstruí-la a partir desses elementos.

O problema de quantização e percepção é estudado do ponto de vista da otimização usando diferentes métricas no espaço de cor. Uma consequência dos resultados apresentados é que o processo de quantização por equalização do histograma de frequências é perceptualmente ótimo. Este fato é bem conhecido e largamente usado, mas até onde nos é conhecido, uma prova não aparece na literatura de processamento de imagens.

**Palavras-chaves:** processamento de imagens, curvatura gaussiana, visão, filtros não-lineares.

# Abstract

We study the problem related to the reconstruction of an image from a perceptual segmentation based on the geometric classification of its points. The classification of these points is accomplished by non-linear operators based on the curvatures of Monge's patch associated to the image. A theoretical proof is presented that an image can be reconstructed solely from its points with non-zero Gaussian curvatures. This result provides theoretical background to a method of non-linear two-dimensional signal processing proposed by C. Zetzche, E. Barth e B. Wegmann. Curvature operators are use to detect edges and vertices from images and we show that it is possible to reconstruct them from these elements.

We study the perceptual problem related to image quantization from an optimization point of view, using different metrics on the color space. A consequence of the results presented is that quantization using histogram equalization provides optimal perceptual results. This fact is well known and widely used but, to our knowledge, a proof has never appeared on the computer graphics literature.

**Key-words:** image processing, Gaussian curvature, vision, non-linear filters.

# Agradecimentos

Ao Professor Jonas Gomes, por sua amizade, por sua personalidade magnânima, pela dedicação a seus alunos.

A Romildo José da Silva, pela amizade e convivência no IMPA durante estes anos.

Aos colegas e funcionários do IMPA e do Projeto Visgraf, pelo excelente ambiente de trabalho.

Aos colegas do Departamento de Matemática da Universidade do Amazonas.

Ao CNPq e à CAPES, pelo suporte financeiro.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Curvatura e percepção . . . . .	2
1.2	Bugs detectors . . . . .	3
1.3	Lei de Weber . . . . .	3
1.4	Conectividade da variedade visual . . . . .	5
1.5	Esboço primário . . . . .	5
<b>2</b>	<b>Imagens</b>	<b>9</b>
2.1	Paradigma de Abstração . . . . .	9
2.2	Modelo funcional de imagem . . . . .	10
2.3	Derivadas em espaços de escala . . . . .	10
2.4	Arestas . . . . .	11
2.5	Segmentações . . . . .	13
2.5.1	Segmentações regulares . . . . .	14
2.5.2	Segmentação a partir das arestas . . . . .	14
2.6	Superfícies paramétricas . . . . .	15
2.6.1	As formas fundamentais . . . . .	15
2.6.2	Curvaturas . . . . .	16
2.6.3	A matriz da derivada da aplicação de Gauss . . . . .	17
<b>3</b>	<b>Geometria diferencial e análise de imagens</b>	<b>19</b>
3.1	Segmentação perceptual de uma imagem . . . . .	19
3.2	Segmentação geométrica de uma imagem . . . . .	20
3.3	Transformadas geométricas . . . . .	22
3.4	Reconstrução a partir dos pontos de curvatura não-nula . . . . .	24
<b>4</b>	<b>Reconstrução</b>	<b>27</b>
4.1	Cálculo da segmentação geométrica . . . . .	27
4.2	Arestas e curvatura Gaussiana . . . . .	29
4.3	Reconstrução . . . . .	31
4.4	Reconstrução a partir da curvatura Gaussiana . . . . .	33
4.4.1	Filtros quadráticos . . . . .	33
4.4.2	Inversão de filtros racionais . . . . .	34

<b>5</b>	<b>Quantização</b>	<b>37</b>
5.1	Partições . . . . .	37
5.2	Partições e quantização . . . . .	37
5.3	Partição e quantização binária . . . . .	38
5.4	Quantização ótima . . . . .	38
5.5	Função de distribuição de probabilidade . . . . .	40
5.6	Medidas de distorção . . . . .	40
5.7	Quantização ótima unidimensional . . . . .	41
5.8	Implementação . . . . .	44
5.9	Quantização e ganho de informação . . . . .	44
<b>6</b>	<b>Aplicações, extensões e trabalhos futuros</b>	<b>53</b>
	<b>Referências Bibliográficas</b>	<b>55</b>

# Lista de Figuras

1.1	O experimento de Attneave. . . . .	3
1.2	What the frog's eye tells the frog's brain: O <i>bug detector</i> presente no sistema visual dos sapos produz resposta não nula apenas para estímulos que apresentam variação em todas as direções. . . . .	4
1.3	Experimento mostrando que o sistema visual executa um processo de reconstrução da imagem vista pelo olho. . . . .	6
1.4	Filtros de passa banda no domínio da frequência. . . . .	7
1.5	Imagens ambíguas. . . . .	7
1.6	A figura em (b) é a mesma figura em (a) a menos de uma rotação de 180 graus. No entanto a imagem em (a) é percebida como um monte enquanto a imagem em (b) é percebida como uma cratera. . . . .	8
2.1	Níveis de abstração no modelo dos quatro universos. . . . .	10
2.2	Conjunto de arestas na imagem de <i>Lena</i> . . . . .	12
2.3	Segmentação a partir das arestas . . . . .	14
3.1	Segmentação perceptual: As regiões 1D são as vizinhanças das arestas e as regiões 2D são as vizinhanças dos vértices. Todos os outros pontos pertencem à região 0D. . . . .	20
3.2	Segmentação geométrica: As regiões g1D são as vizinhanças das arestas e as regiões g2D são as vizinhanças dos vértices. Todos os outros pontos pertencem à região g0D. . . . .	21
3.3	O catenóide e helicóide são duas superfícies mínimas isométricas. Portanto, possuem os mesmos coeficientes métricos $E, F, G$ , e as mesmas curvaturas Gaussiana e média. . . . .	23
3.4	Superfícies compactas distintas com as mesmas curvaturas Gaussiana e média. . . . .	23
4.1	Segmentação geométrica: os pontos em cinza representam os pontos planares, os outros pontos na cor original da imagem (preto e branco) são os pontos selecionados para cada segmento. . . . .	30
4.2	Segmentação e reconstrução da imagem de <i>Lena</i> . Os pontos em tom médio de cinza representam os pontos planares que foram descartados. Os pontos parabólicos, elípticos e hiperbólicos foram mantidos na sua cor original. . . . .	32

4.3	Imagens distintas com mesma curvatura Gaussiana. As figuras (c) e (d) ilustram o perfil das superfícies de Monge das imagens em (a) e (b). . . . .	35
5.1	Segmentação, e quantização de uma função. . . . .	38
5.2	Partição e quantização binária . . . . .	39
5.3	$d(x, y) =  x - y $ . . . . .	47
5.4	$d(x, y) =  x - y ^2$ . . . . .	48
5.5	$d(x, y) =  \log(x) - \log(y) $ . . . . .	49
5.6	$d(x, y) =  \log(x) - \log(y) ^2$ . . . . .	50
5.7	Mediana . . . . .	51

# Capítulo 1

## Introdução

*“O desenho não é uma brincadeira. É muito grave e misterioso que um traço possa representar um ser vivo. Não apenas sua imagem, mas sobretudo aquilo que ele realmente é. Que maravilha! Não seria isso mais surpreendente que todas as prestidigitações e ‘coincidências’ do mundo?”*

*Pablo Picasso*

Há mais de trinta anos a comunidade científica tem conhecimento da existência de células nos sistemas visuais dos animais especializadas na detecção de características como *bugs*, cantos, fins de linha etc, (DeValois & DeValois, 1980; DeValois *et al.*, 1982; Lettvin *et al.*, 1959), e mais recentemente em (Barth *et al.*, 1999). Tais células são conhecidas genericamente como *endstopped cells* e apresentam resposta mínima ou nenhuma resposta a estímulos que não apresentam variação ou que apresentam apenas variação ao longo de retas paralelas (variação unidimensional). Apesar disso, as técnicas mais frequentemente utilizadas para a análise e modelagem das funções do sistema visual utilizam operadores lineares.

A presença ubíqua dos operadores lineares se deve, em grande parte, à simplicidade de cálculo com essas transformações. Por exemplo, a inversa de uma transformação linear ainda é uma transformação linear e há métodos bem conhecidos para o seu cálculo. O cálculo da inversa de uma transformada, caso exista, pode ser essencial para algumas aplicações, como por exemplo na codificação e compressão de imagens. Entre as transformadas lineares mais populares encontramos, por exemplo, a transformada de Fourier, wavelets e o filtro laplaciano.

Apesar do grande sucesso dos métodos lineares, alguns métodos não-lineares de processamento de imagens têm sido desenvolvidos. Entre esses, um dos mais conhecidos é certamente a codificação fractal de imagens. Nele, cada imagem é associada ao único ponto fixo de um sistema iterado de funções (IFS). O IFS é usado para codificar a imagem, a qual pode ser recuperada pela ação do IFS sobre qualquer imagem inicial.

A despeito do uso bem sucedido de métodos lineares no processamento digital de sinais e na modelagem de algumas funções do sistema visual dos animais por

neurofisiologistas, não é difícil notar que operadores lineares não podem oferecer um modelo adequado para *endstopped cells*. Isto ocorre porque as auto-funções do operador são sinais unidimensionais, (Zetzsche & Barth, 1990; Zetzsche *et al.*, 1993). Experimentos de psicologia revelam, ainda, que os pontos que contêm a maior quantidade de informação no contorno de um desenho são os de extremo da curvatura (Attneave, 1954).

Uma vez que não pode existir um filtro linear capaz de modelar a função de uma célula insensível a estímulos de dimensão zero ou um, o estudo de transformadas não-lineares, mais do que uma opção, torna-se uma necessidade. No artigo “The Importance of Intrinsically Two-dimensional Image Features”, C. Zetzsche, E. Barth e B. Wegmann fazem um estudo detalhado das limitações de um filtro linear e propõem a curvatura gaussiana da superfície de Monge da função imagem como um bom operador não-linear para estudos envolvendo a percepção, (Zetzsche *et al.*, 1993). A curvatura gaussiana como filtro não-linear é estudada com mais detalhes em “Image Encoding, Labeling, and Reconstruction from Differential Geometry”, onde os autores mostram que é possível reconstruir parcialmente uma imagem partindo apenas do conjunto de pontos de curvatura gaussiana não-nula, (Barth *et al.*, 1993). Neste mesmo trabalho os autores propõem que aquele conjunto contém toda a informação necessária para a reconstrução de uma imagem cinza.

A razão para a escolha da curvatura gaussiana está em uma decomposição natural de uma imagem em três regiões: uma região onde a imagem não apresenta variação, uma região onde ocorre apenas variação unidimensional e uma terceira região onde ocorre variação em todas as direções. Esta última contém as características geométricas presentes em *bugs*, cantos, fins de linhas etc, e é a única região para a qual a curvatura gaussiana produz resposta não nula.

Transformadas lineares e não-lineares têm aplicações em várias áreas relacionadas à Computação Gráfica, à Fisiologia e à Psicologia como por exemplo: Visão Biológica, Visão Artificial, Robótica, Codificação e Compressão de Imagens, Segmentação e Reconstrução de Imagens, e Reconhecimento de Padrões.

A seguir apresentaremos um rápido resumo da literatura mostrando evidências de processamento não-linear de imagens por sistemas visuais.

## 1.1 Curvatura e percepção

Em 1954 Attneave, (Attneave, 1954) realizou uma série de experimentos com a percepção humana de contornos e concluiu que entre os pontos de uma linha fechada aqueles mais representativos são os de extremo da curvatura.

Attneave conduziu seu experimento do seguinte modo: submete-se a um indivíduo um contorno como o da figura 1.1(a) e pede-se que ele escolha um determinado número de pontos sobre a figura de maneira que se possa reconstruí-la. Se o número de pontos é pequeno demais a informação nele contida não será suficiente para a reconstrução. Se o número de pontos é grande demais essa informação é largamente redundante. Para um número apropriado de pontos, digamos seis pontos para a figura 1.1(a), o resultado do experimento para um grupo de indivíduos é

similar à figura 1.1(b), onde as barras indicam a frequência dos pontos escolhidos.

Attneave inferiu que entre os pontos escolhidos os mais prováveis são os de extremo da curvatura. Para ilustrar suas conclusões, calculou os pontos de extremo da curvatura para o contorno de um gato e o reconstruiu a partir desses pontos, com segmentos de reta, resultando em um contorno que ainda pode ser reconhecido como de um gato, figura 1.1(c).

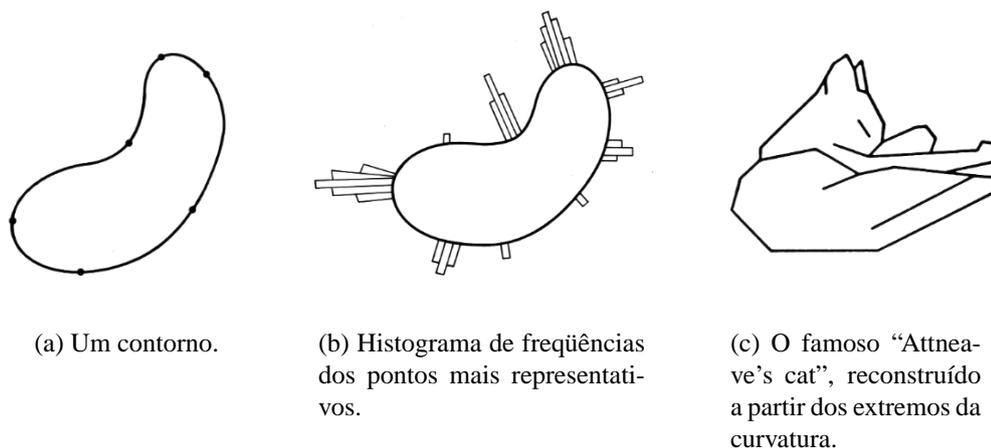


Figura 1.1: O experimento de Attneave.

## 1.2 Bug detectors

Há mais de trinta anos neurofisiologistas sabem que existem no sistema visual dos animais células especializadas na detecção de *bugs*, cantos, fins de linha etc.

Um dos experimentos pioneiros neste campo foi realizado por Lettvin e Maturana em 1959. O experimento consistiu em medir as respostas elétricas de algumas células do sistema visual dos sapos enquanto o olho era estimulado com diferentes figuras. Descobriu-se que o sistema visual desses animais apresenta células que não produzem resposta quando estimuladas com imagens que apresentam características de variação unidirecional. O experimento e o tipo de resposta para algumas imagens é ilustrado na figura 1.2.

## 1.3 Lei de Weber

Um aspecto interessante do sistema visual humano é que a percepção de intensidade de luz ocorre em quantidades. A *Lei de Weber* afirma que se  $\Delta x_k = x_{k+1} - x_k$  é o menor valor tal que se pode perceber diferença entre as intensidades de iluminação  $x_k$  e  $x_{k+1}$ , então

$$\frac{\Delta x_k}{x_k} = \varepsilon = \text{constante.} \quad (1.1)$$

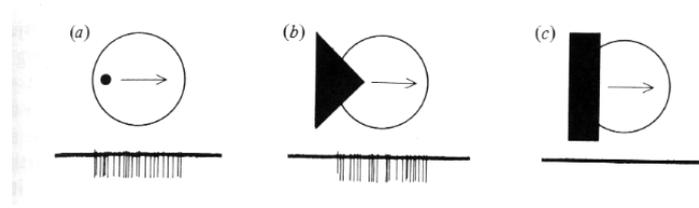


Figura 1.2: “What the frog’s eye tells the frog’s brain”: O *bug detector* presente no sistema visual dos sapos produz resposta não nula apenas para estímulos que apresentam variação em todas as direções.

Portanto, se começarmos com um valor de intensidade  $x_0$ , o  $k$ -ésimo valor de intensidade será

$$x_k = x_0(1 + \varepsilon)^k = x_0 e^{k \log(1+\varepsilon)} \quad (1.2)$$

que é equivalente a

$$\log(x_k) = \log(x_0) + k \log(1 + \varepsilon). \quad (1.3)$$

Isto mostra que o sistema visual humano faz uma quantização não uniforme do espaço de cor.

A função logaritmo transforma a escala não-linear induzida no espaço de cores pelo sistema visual em uma escala linear. Portanto, pode-se modelar a resposta psicofísica  $\phi(x)$ , da intensidade  $x$ , pela equação

$$\phi(x) = C \log(x) \quad (1.4)$$

A constante  $C$  incorpora uma possível mudança de base no logaritmo e pode ser fisicamente estimada.

A função de resposta psicofísica pode ser estudada do ponto de vista da teoria da informação, (Resnikoff, 1987). Nesse contexto, pode-se mostrar que as únicas possibilidades para a função  $\phi$  são:

$$\phi(x) = C \log(x) \quad \text{ou} \quad (1.5)$$

$$\phi(x) = Cx^r, \quad r \neq 0. \quad (1.6)$$

O modelo  $\phi(x) = C \log(x)$  foi proposto por Fechner, (Fechner, 1858). Enquanto o modelo  $\phi(x) = Cx^r$  foi proposto por Stevens, (Stevens, 1961).

Outra maneira de interpretar a Lei de Weber é imaginar que a unidade com a qual medimos distância no *espaço de cores* muda de acordo com a cor. Tal mudança é proporcional à intensidade luminosa da cor. O conceito matemático que abstrai a noção intuitiva de mudança de escala para cada ponto é o de *métrica Riemanniana*.

No contexto da Geometria Riemanniana, a Lei de Weber indica que a métrica do espaço de cor unidimensional é

$$\alpha^2 \left( \frac{dx}{x} \right)^2. \quad (1.7)$$

A generalização dessa métrica para o espaço de cores tridimensional é conhecida como *métrica de Helmholtz*, (von Helmholtz, 1891),

$$\alpha^2 \left(\frac{dx}{x}\right)^2 + \beta^2 \left(\frac{dy}{y}\right)^2 + \gamma^2 \left(\frac{dz}{z}\right)^2. \quad (1.8)$$

A distância entre as cores  $A = (x_0, y_0, z_0)$  e  $B = (x_1, y_1, z_1)$  nessa métrica é dada por

$$d^2(A, B) = \alpha^2 \log^2(x_1 - x_0) + \beta^2 \log^2(y_1 - y_0) + \gamma^2 \log^2(z_1 - z_0). \quad (1.9)$$

Se duas cores diferem apenas em brilho, isto é,  $B = rA$ , então

$$d(A, B) = \sqrt{\alpha^2 + \beta^2 + \gamma^2} |\log(r)| \quad (1.10)$$

que é a função psicofísica de Fechner, (Resnikoff, 1987).

## 1.4 Conectividade da variedade visual

Um experimento simples pode ser usado para mostrar que o cérebro humano executa um processo de reconstrução da imagem visualizada. Fixando-se o olho direito na cruz branca da figura 1.3(a) e afastando-a cerca de 30 cm do rosto, é possível encontrar um ponto onde não se vê a bola branca presente na figura. O desaparecimento da bola branca ocorre devido à existência de uma região cega no olho humano, no entanto quando olhamos para uma imagem, como por exemplo esta folha de papel, não percebemos nenhuma falha na imagem percebida.

A figura 1.3(b) mostra que o processo de reconstrução não é simplesmente um processo de interpolação a partir do contorno da região que não é vista pelo olho. Neste caso, a figura percebida apresentaria uma variação suave do branco para o preto na região obstruída pelo ponto cego do olho; no entanto, a parte escura da imagem é reconstruída com arestas em linha reta.

## 1.5 Esboço primário

Desde David Marr, a comunidade científica tem procurado caracterizar uma imagem a partir de um número mínimo de informações. A proposta de Marr era caracterizar uma imagem a partir de suas arestas calculadas em várias escalas (*primal sketch*). Neste caso, as arestas seriam calculadas como *zero crossings* do laplaciano

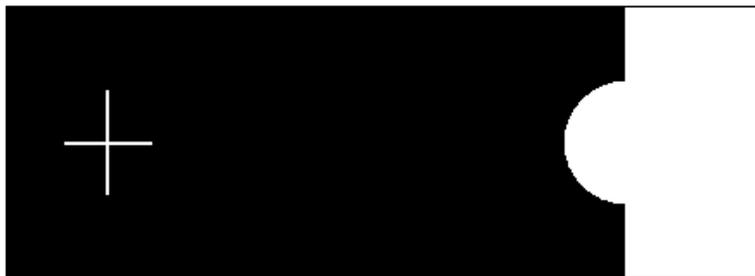
$$\Delta(h * g_\sigma) = h * \Delta g_\sigma, \quad \text{onde } g_\sigma(u, v) = \frac{1}{2\pi\sigma} e^{-\frac{u^2+v^2}{2\sigma}}. \quad (1.11)$$

O filtro  $\Delta g_\sigma$  é chamado de *LOG* a partir da expressão em inglês *Laplacian of Gaussian*. A função de transferência desse filtro é mostrada na figura 1.4(b).

A razão para o uso do filtro *LOG* reside em um teorema de amostragem de sinais o qual afirma que um *filtro de passa-oitava* pode ser reconstruído de maneira



(a)



(b)

Figura 1.3: Experimento mostrando que o sistema visual executa um processo de reconstrução da imagem vista pelo olho.

exata a partir de suas raízes, (Logan Jr., 1977). Um filtro de passa-oitava é um filtro onde a razão entre as frequências de menor e maior valor absoluto é menor que  $1/2$ , veja figura 1.4(a). A figura 1.4(b) mostra claramente que o *LOG* é um filtro de passa-banda, entretanto, não é um filtro de passa-oitava.

É claro que se conseguirmos recuperar  $\Delta(h * g_\sigma)$ , então  $h * g_\sigma$  pode ser recuperada, pois a equação

$$\Delta f = g \tag{1.12}$$

determina a função  $f$  a menos da condição de contorno. Entretanto, o próprio D. Marr chama a atenção para o fato de que a percepção humana envolve características mais complexas, como pode ser visto nas figuras 1.5 e 1.6.

A conjectura de Marr foi provada por Hummel e Moniot, supondo que se tem informações sobre o gradiente da imagem ao longo das arestas, (Hummel & Moniot, 1989).

Posteriormente, Stephane Mallat usou um esquema de *Wavelets* para calcular e reconstruir uma imagem a partir de suas arestas, (Mallat, 1991). O método de Mallat pode ser visto como um esquema de amostragem e reconstrução, em que os pontos de amostragem são as arestas em uma determinada resolução e as amostras são as derivadas da imagem suavizada. Mallat também propôs que seu esquema

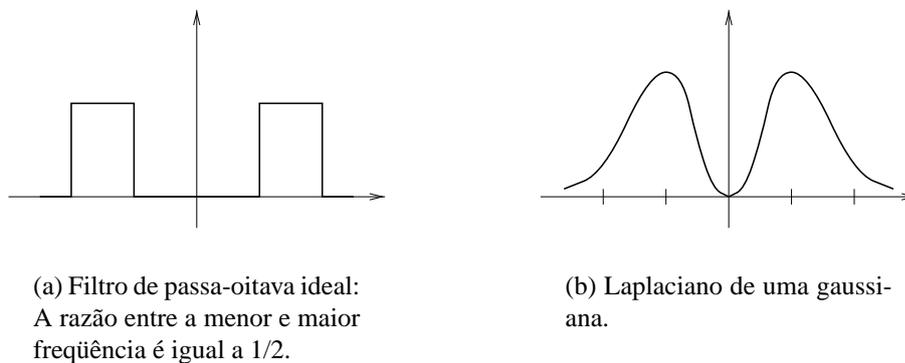


Figura 1.4: Filtros de passa banda no domínio da frequência.



(a) Cachorro.

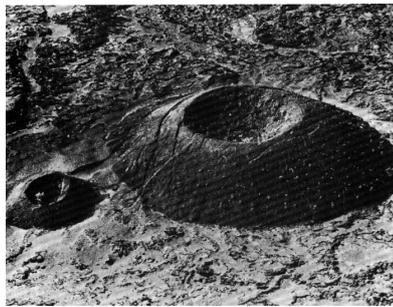


(b) Mulher jovem ou idosa?

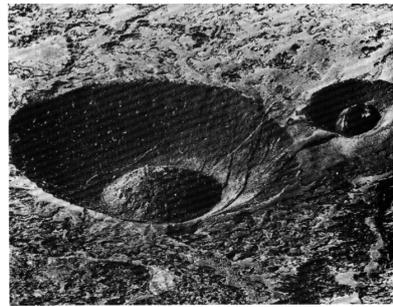
Figura 1.5: Imagens ambíguas.

de amostragem é completo e implementou um algoritmo para a reconstrução da imagem que na prática apresentou bons resultados.

Apesar dos trabalhos bem sucedidos de Mallat, Hummel e Moniot, Meyer mostrou através de exemplos analíticos que ambas as conjecturas (de Marr e Mallat) são falsas, (Meyer, 1991). O contra-exemplo de Meyer, reside na escolha específica da wavelet escolhida por Mallat para a detecção das arestas e reconstrução da imagem. Portanto, o problema de saber se outra escolha da wavelet pode levar a uma demonstração da conjectura de Mallat permanece em aberto.



(a)



(b)

Figura 1.6: A figura em (b) é a mesma figura em (a) a menos de uma rotação de 180 graus. No entanto a imagem em (a) é percebida como um monte enquanto a imagem em (b) é percebida como uma cratera.

# Capítulo 2

## Imagens

Neste capítulo escolheremos um modelo matemático para o estudo analítico das imagens e descreveremos algumas técnicas de processamento que decorrem naturalmente desta escolha. Em particular, descreveremos brevemente o espaço de imagens, derivadas em espaço de escala, arestas, segmentações e técnicas de geometria diferencial.

### 2.1 Paradigma de Abstração

Para aplicar métodos matemáticos e computacionais ao estudo das imagens é conveniente estabelecer uma hierarquia de abstrações. Uma vez escolhida tal hierarquia, podemos usar para cada nível de abstração um modelo matemático apropriado. Neste trabalho usaremos o *paradigma dos quatro universos* introduzido em (Gomes & Velho, 1995).

No modelo dos quatro universos, associamos a cada imagem quatro representações: a imagem física, uma idealização matemática da noção física, uma representação discreta e uma codificação da imagem. Os pares de operações de passagem entre os vários níveis de abstração são denominados: *abstração e realização*, *discretização e reconstrução* e *codificação e decodificação*. A figura 2.1 ilustra o paradigma dos quatro universos.

A escolha do modelo matemático em cada um dos vários níveis de abstração conduz a problemas distintos no estudo das imagens. Intrínseco a cada escolha dos modelos estão os problemas de estudo dos modelos matemáticos escolhidos e das operações de passagem entre eles. Neste trabalho, aplicaremos métodos de geometria diferencial na análise e processamento de imagens, portanto escolheremos um modelo matemático para as imagens que nos permita aplicar tais métodos. Estaremos principalmente interessados no estudo dos modelos matemático e discreto de uma imagem e nas operações de discretização e reconstrução.

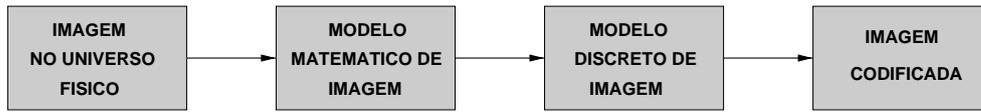


Figura 2.1: Níveis de abstração no modelo dos quatro universos.

## 2.2 Modelo funcional de imagem

No modelo funcional, uma imagem cinza é representada por uma função  $h : U \rightarrow \mathbb{R}$ , onde o *suporte da imagem*,  $U$ , é um subconjunto do plano e  $h(u, v)$  representa a intensidade de cinza no ponto  $(u, v) \in U$ . Neste modelo, um espaço de imagens é um subespaço vetorial do espaço de funções  $\mathcal{L} = \{h : U \rightarrow \mathbb{R}\}$  munido da soma e produto por escalar usuais:

$$(g + h)(x) = g(x) + h(x) \quad (2.1)$$

e

$$(\lambda h)(x) = \lambda h(x) \quad (2.2)$$

para  $g, h \in \mathcal{L}$ .

Se além disso supomos que existe  $\int_U h(x)^2 dx$ , o que significa que a energia da imagem é finita, o espaço  $\mathcal{L}$  fica naturalmente munido de um produto interno  $\langle f, g \rangle = \int_U g(x)h(x) dx$ . Este é um espaço de funções muito bem estudado na matemática e será o espaço usado com modelo para o espaço de imagens neste trabalho. Convém lembrar que se dispõe de outros modelos que também podem ser usados para representar apropriadamente uma imagem, como por exemplo as *distribuições temperadas* e os *modelos estocásticos*.

## 2.3 Derivadas em espaços de escala

Para um grande número de técnicas de processamento de imagens é necessário o cálculo das derivadas da função  $h$ . Algumas das dificuldades em se utilizar essas técnicas se devem aos seguintes problemas:

- Em geral uma imagem não é diferenciável e apresenta uma grande quantidade de ruídos;
- Pequenas perturbações de uma função diferenciável podem acarretar grandes variações na sua derivada.

Esses problemas tornam a diferenciabilidade discreta um problema mal posto, mas podem ser contornados usando técnicas de regularização em espaços de escala, (Lindeberg, 1994). O método, essencialmente, consiste em escolher uma função

diferenciável positiva  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  satisfazendo  $\int g(x) dx = 1$ . Neste caso é um fato conhecido que a família de funções definida por

$$h * g_\sigma(x) = \int h(y)g_\sigma(x - y)dy, \quad \text{onde, } g_\sigma(x) = \frac{1}{\sigma}g\left(\frac{x}{\sigma}\right)$$

tem boas propriedades de convergência e diferenciabilidade. Em particular valem as seguintes proposições:

**Proposição 1.** Se  $h$  é contínua e limitada então  $h * g_\sigma$  converge uniformemente para  $h$  quando  $\sigma \rightarrow 0$ .

**Proposição 2.** Se  $g$  é derivável então  $h * g$  é derivável e vale

$$\frac{\partial}{\partial x_j}(h * g) = h * \left(\frac{\partial g}{\partial x_j}\right). \quad (2.3)$$

Quando a função  $h$  é derivável ambas as expressões na equação são iguais a  $\frac{\partial h}{\partial x_j} * g$ . A proposição 1 junto com a equação 2.3 permitem calcular aproximações para derivadas de  $h$  a partir das derivadas da função de escala  $g_\sigma$ .

Embora a função de escala possa ser escolhida de maneira bastante geral, o olho humano é menos sensível a erros que apresentam características de simetria. Por esta razão a função de escala  $g$  geralmente é escolhida com propriedades de simetria.

Uma análise mais atenta da última igualdade da equação 2.3 mostra que não é necessário conhecer explicitamente a função de escala para calcular as derivadas de  $h$ . Para isto é suficiente conhecer as derivadas  $\frac{\partial g}{\partial x_j}$ . Esta idéia foi utilizada para introduzir novos filtros de derivação em (Ma & Li, 1998).

A noção de espaço de escala apresentada ligeiramente acima está relacionada de modo muito próximo com os conceitos matemáticos de *wavelet* e *identidade aproximada*.

## 2.4 Arestas

Quando olhamos para uma imagem podemos claramente identificar regiões onde há pouca variação na intensidade de cor da imagem. Estas regiões aparecem separadas por curvas onde ocorrem grandes variações ao se mudar de uma região para outra. Isto significa que a função  $h$  que modela uma imagem não é necessariamente diferenciável, podendo até mesmo apresentar pontos de descontinuidade. Tais pontos são chamados de *arestas* da imagem. As arestas são elementos de grande importância para a percepção como foi mostrado por David Marr, (Marr, 1982).

O processo de regularização introduzido na seção 2.3 transforma a função  $h$  em uma função diferenciável e portanto sem pontos de descontinuidade. Entretanto, se a escala é suficientemente fina podemos recuperar esses pontos, como indica a proposição 3 a seguir.

Para simplificar a notação e as hipóteses, no que segue restringiremos a discussão na seção seguinte a uma função  $h : \mathbb{R} \rightarrow \mathbb{R}$  seccionalmente diferenciável e com suporte compacto.

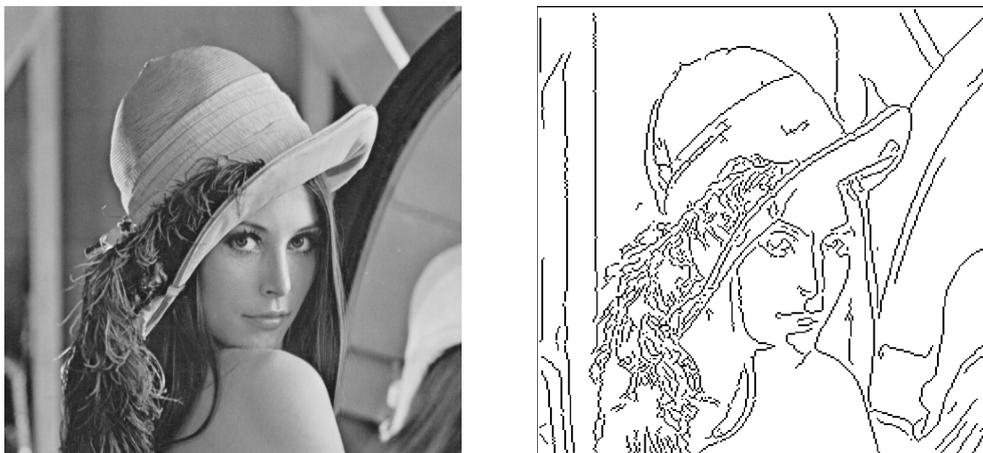


Figura 2.2: Conjunto de arestas na imagem de *Lena*.

**Proposição 3.** Sejam  $h$  uma função seccionalmente contínua e  $g$  é uma função diferenciável tal que

1.  $g(x) \geq 0$  e  $g(x) = 0$  se  $|x| > s/2$ ,
2.  $\int g(x) dx = 1$  e o único ponto de máximo local de  $g$  é  $x = 0$ .

Suponha que  $p$  é um ponto de descontinuidade de  $h$  e  $s$  é suficientemente pequeno. Então,  $|\frac{d}{dx}(h * g)|$  possui um ponto de máximo local  $p_s$ , com  $|p - p_s| < s$ .

*Demonstração.* Como  $h$  é seccionalmente contínua, se  $s$  é suficientemente pequeno podemos supor que  $p$  é o único ponto de descontinuidade de  $h$  no intervalo  $[p - s, p + s]$ . Como  $x = 0$  é o único ponto de máximo local de  $g$ , segue que  $g(x - p)$  não muda de sinal em cada um dos intervalos  $[p - s/2, p]$  e  $[p, p + s/2]$ . Podemos então usar o teorema do valor médio para integrais e obter constantes  $a_1 \in (p - s/2, p)$  e  $a_2 \in (p, p + s/2)$  tais que

$$\begin{aligned} \frac{d}{dx}(h * g)(p) &= \int_{p-s/2}^p h(x)g'(p-x) dx + \int_p^{p+s/2} h(x)g'(p-x) dx \\ &= (h(a_1) - h(a_2))g(0). \end{aligned}$$

Repetindo o argumento para  $p_1 = p - s/2$  e  $p_2 = p + s/2$ , obtemos constantes  $b_1, b_2 \in (p - s, p)$  e  $c_1, c_2 \in (p, p + s)$  tais que

$$\begin{aligned} \frac{d}{dx}(h * g)(p_1) &= (h(b_1) - h(b_2))g(0) \quad \text{e} \\ \frac{d}{dx}(h * g)(p_2) &= (h(c_1) - h(c_2))g(0). \end{aligned}$$

Como  $h$  é contínua por partes e descontínua em  $p$ , se  $s$  é pequeno temos

$$\left| \frac{d}{dx}(h * g)(p) \right| > \left| \frac{d}{dx}(h * g)(p_1) \right|, \left| \frac{d}{dx}(h * g)(p_2) \right|.$$

Portanto, como  $|\frac{d}{dx}(h * g)|$  é contínua, assume um máximo local em  $p_s \in (p - s, p + s)$ .  $\square$

Se  $p$  é um ponto de máximo de  $|\frac{d}{dx}(h * g)|$  então  $\frac{d^2}{dx^2}(h * g)(p) = 0$ . Portanto, a proposição acima diz que os pontos de aresta de  $h$  são aproximados pelos pontos de máximo da derivada primeira (*Canny edges*) ou pelos zeros da derivada segunda de  $h * g$  (*zero crossings*).

## 2.5 Segmentações

Uma técnica freqüentemente utilizada em processamento de imagens consiste em particionar o suporte da imagem em subconjuntos cujos pixels apresentem características de similaridade. Este processo é conhecido como *segmentação da imagem*. Em processos automáticos de segmentação de uma imagem pode acontecer que as regiões resultantes apresentem algumas características não desejadas como linhas conectadas a regiões bem definidas, pontos isolados etc. Quando isto ocorre, dizemos que as regiões não são regulares. A proposição 4 mostra que sempre é possível obter uma “boa segmentação” a partir de uma segmentação qualquer.

Seja  $X$  um subconjunto do plano, indicaremos por  $\text{int}(X)$ ,  $\overline{X}$  e  $\partial X$  o conjunto dos pontos interiores, o fecho topológico e a fronteira de  $X$ , respectivamente.

A operação que consiste em tomar o fecho topológico do conjunto dos pontos interiores de  $X$  é chamada *regularização* de  $X$  e é indicada por  $\mathcal{R}(X) = \overline{\text{int}(X)}$ . O conjunto  $X$  é *regular* se a operação de regularização não altera  $X$ .

**Definição 1.** Um subconjunto regular  $U \subset \mathbb{R}^2$  é uma *região* se sua fronteira é uma curva diferenciável por partes.

**Definição 2.** Seja  $U$  uma região. Uma *segmentação* de  $U$  é uma coleção  $\mathcal{S} = \{U_\lambda\}$  de subconjuntos de  $U$ , tais que:

1.  $U = \bigcup U_\lambda$ ;
2. Se  $\alpha \neq \beta$  então  $\overline{U}_\alpha \cap \overline{U}_\beta$  tem interior vazio em  $U$ .

Cada subconjunto  $U_\lambda$  é chamado um *segmento* de  $U$ . Uma segmentação é *finita* se possui um número finito de subconjuntos. Uma segmentação é *localmente finita* se cada ponto de  $U$  possui uma vizinhança que intercepta apenas um número finito de subconjuntos da segmentação. Neste trabalho, todas as segmentações serão consideradas finitas a menos que seja explicitamente enunciado.

**Observação 1.** Se  $U$  é uma região e  $\mathcal{S} = \{U_\lambda\}$  é uma segmentação de  $U$ , então  $\overline{\mathcal{S}} = \{\overline{U}_\lambda\}$  é uma segmentação de  $U$  por conjuntos fechados.

### 2.5.1 Segmentações regulares

Diremos que uma segmentação é regular se  $U_\lambda$  é regular para todo  $\lambda$ . Em particular, se a segmentação é regular os conjuntos  $U_\lambda$  são fechados. A proposição a seguir mostra que toda partição finita pode ser regularizada.

**Proposição 4.** Se  $U$  é uma região, toda segmentação finita  $\mathcal{S} = \{U_\lambda\}$  de  $U$  pode ser regularizada através da substituição de  $U_\lambda$  pela sua regularização  $R_\lambda = \mathcal{R}(U_\lambda)$ .

*Demonstração.* Pela observação 1, podemos supor que os conjuntos  $U_\lambda$  são fechados. Neste caso,  $R_\lambda \subset U_\lambda$  e portanto  $R_\alpha \cap R_\beta$  tem interior vazio se  $\alpha \neq \beta$ .

Resta mostrar que  $R = \bigcup R_\lambda$  coincide com  $U$ . Seja  $C_\lambda = \bigcup_{\beta \neq \lambda} U_\beta$ ,  $C_\lambda$  é essencialmente o complementar de  $U_\lambda$  em  $U$ . Como cada  $U_\lambda$  é fechado e a família é uma segmentação finita, segue que  $R$  e  $C_\lambda$  são fechados e  $U_\lambda \cap C_\lambda$  tem interior vazio.

Seja  $p \in U$  e suponha, por absurdo, que  $p \notin R$ , então  $p \in \bigcup \partial U_\lambda$ . Por outro lado, como  $U_\lambda$  e  $C_\lambda$  são fechados e  $U = U_\lambda \cup C_\lambda$ , temos  $\partial U_\lambda = U_\lambda \cap C_\lambda$ . Portanto, se  $p \notin R$  então  $p \in \bigcup \partial U_\lambda \subset \bigcup (U_\lambda \cap C_\lambda) = \bigcup_{\alpha \neq \beta} (U_\alpha \cap U_\beta)$ . Como a segmentação é finita, segue que o complementar de  $R$  tem interior vazio e como  $R$  é fechado, o complementar de  $R$  é aberto. Concluimos que o complementar de  $R$  é aberto e tem interior vazio, logo é vazio, o que contradiz  $p \notin R$ . Portanto  $U = R = \bigcup R_\lambda$ .  $\square$

### 2.5.2 Segmentação a partir das arestas

Uma técnica largamente utilizada em processamento de imagens consiste em obter uma segmentação  $U = \bigcup U_\lambda$  de modo que não ocorra grandes variações de intensidade em cada componente  $U_\lambda$ , figura 2.3. Como as grandes variações de intensidade da imagem ocorrem nas vizinhanças das arestas, a segmentação deve ser obtida de modo que as arestas façam parte das fronteiras da segmentação.

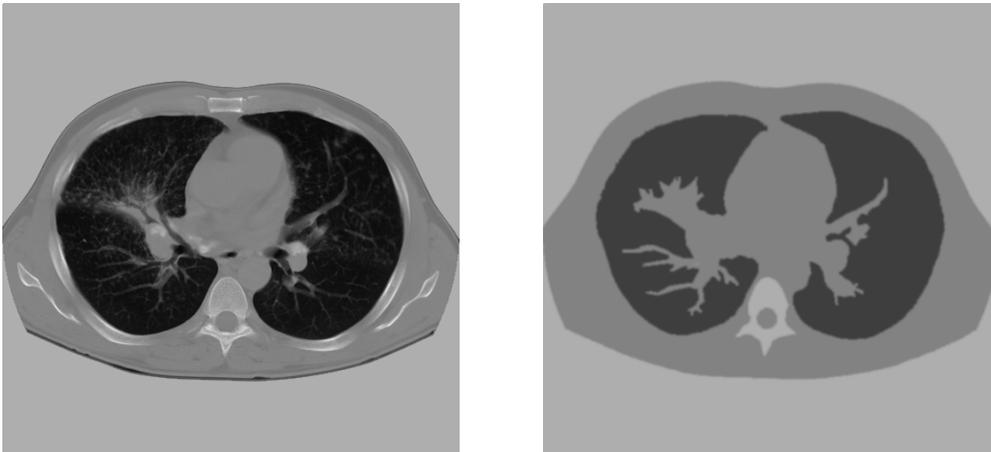


Figura 2.3: Segmentação a partir das arestas

Métodos de segmentação a partir das arestas de uma imagem têm recebido especial atenção desde que David Marr propôs que o conjunto de arestas de uma imagem forneceriam uma representação completa desta. Veja secção 1.5.

## 2.6 Superfícies paramétricas

Nos últimos cinco anos, tem havido crescente interesse na utilização de técnicas de geometria diferencial em processamento de imagens. Nesta secção apresentaremos de maneira breve alguns conceitos básicos de geometria diferencial que serão utilizados nos capítulos seguintes.

Seja  $U$  um subconjunto aberto do  $\mathbb{R}^2$  com coordenadas  $(u, v)$ . Uma *superfície paramétrica*,  $S$ , é uma aplicação diferenciável  $f : U \rightarrow \mathbb{R}^3$  tal que  $f_u$  e  $f_v$  são linearmente independentes ao longo de todo o domínio de  $f$ . Dado  $p \in U$ , os vetores  $f_u(p)$  e  $f_v(p)$  geram um subespaço vetorial do  $\mathbb{R}^3$  chamado *plano tangente* de  $S$  em  $p$ . Indica-se esse plano por  $T_p S$ .

No modelo funcional, uma imagem  $h$  está naturalmente associada a uma superfície através da parametrização

$$(u, v) \mapsto f(u, v, h(u, v)).$$

Chamamos essa superfície de *superfície de Monge* da imagem.

### 2.6.1 As formas fundamentais

A forma geométrica de uma superfície fica determinada a partir do conhecimento de duas formas quadráticas que introduziremos a seguir.

Se  $\alpha' = u'f_u + v'f_v$  é um vetor no plano tangente de  $S$ , a forma quadrática definida por

$$I_p(\alpha') = \langle \alpha', \alpha' \rangle = E(u')^2 + 2Fu'v' + G(v')^2,$$

onde,

$$E = \langle f_u, f_u \rangle, \quad F = \langle f_u, f_v \rangle, \quad G = \langle f_v, f_v \rangle,$$

é chamada *primeira forma fundamental* de  $S$ .

Para cada  $p \in U$ , existe um único vetor unitário  $N$  tal que  $\{f_u, f_v, N\}$  é uma base positiva do  $\mathbb{R}^3$ . A aplicação

$$N : U \rightarrow \mathbb{R}^3 \tag{2.4}$$

é a *aplicação normal de Gauss* de  $S$ . Como

$$\langle N, N \rangle = 1,$$

segue que

$$\langle N, N_u \rangle = \langle N, N_v \rangle = 0$$

e portanto a derivada de  $N$  é uma transformação linear de  $T_p S$  em  $T_p S$ .  $dN$  é uma transformação auto-adjunta, veja (do Carmo, 1976) para uma demonstração.

A segunda forma fundamental de  $S$  é a forma quadrática definida por

$$II_p(\alpha') = -\langle dN(\alpha'), \alpha' \rangle = e(u')^2 + 2fu'v' + g(v')^2,$$

onde,

$$e = -\langle N_u, f_u \rangle, \quad f = -\langle N_u, f_v \rangle = -\langle N_v, f_u \rangle, \quad g = -\langle N_v, f_v \rangle,$$

Os coeficientes  $E, F, G, e, f, g$  determinam completamente uma superfície paramétrica. Melhor ainda, o *Teorema Fundamental da Teoria Local das Superfícies* afirma que se esses coeficientes satisfazem um conjunto de equações de compatibilidade, conhecidas como equações de Gauss e Mainardi-Codazzi, então existe uma única superfície paramétrica com esses coeficientes. Um enunciado mais preciso bem como a demonstração desse teorema pode ser encontrada em (do Carmo, 1976).

## 2.6.2 Curvaturas

Chama-se curvatura gaussiana ( $K$ ) e curvatura média ( $H$ ) ao determinante e ao hemi-traço de  $-dN$ , respectivamente. Isto é

$$K = \det(dN) \tag{2.5}$$

$$H = -1/2 \text{traço}(dN). \tag{2.6}$$

Os autovalores de  $-dN$  chamam-se curvaturas principais da superfície e podem ser facilmente calculados a partir de  $K, H$  e da relação

$$k^2 - 2Hk + K = 0.$$

A segunda forma fundamental tem grande influência na forma que a superfície assume em torno de um ponto. Se  $(u', v', z)$  representa as coordenadas do  $\mathbb{R}^3$  em relação à base  $\{f_u, f_v, N\}$  transladada para o ponto  $p \in S$ , então

$$z = e(u')^2 + 2fu'v' + g(v')^2$$

é a equação do parabolóide osculador à superfície em  $p$ . Os pontos de  $S$  são classificados com relação à forma desse parabolóide em planares, parabólicos, elípticos e hiperbólicos. De modo mais preciso, dizemos que  $p \in S$  é:

1. Planar, se  $K = H = 0$ ;
2. Parabólico, se  $K = 0$  e  $H \neq 0$ ;
3. Elíptico, se  $K > 0$ ;
4. Hiperbólico se  $K < 0$ .

### 2.6.3 A matriz da derivada da aplicação de Gauss

A matriz de  $dN$  em relação à base ordenada  $\{f_u, f_v\}$  pode ser calculada a partir dos coeficientes  $E, F, G, e, f, g$  da primeira e segunda formas fundamentais.

Como  $N_u$  e  $N_v$  pertencem ao plano tangente de  $S$ , podemos escrever

$$\begin{aligned} N_u &= a_{11}f_u + a_{21}f_v \\ N_v &= a_{12}f_u + a_{22}f_v. \end{aligned}$$

Tomando-se produtos internos com  $f_u$  e  $f_v$  tem-se

$$\begin{aligned} -e &= a_{11}E + a_{21}F \\ -f &= a_{11}F + a_{21}G \\ -f &= a_{12}E + a_{22}F \\ -g &= a_{12}F + a_{22}G \end{aligned}$$

de onde se obtém,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = - \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} e & f \\ f & g \end{pmatrix} \quad (2.7)$$

A equação 2.7 permite calcular as curvaturas gaussiana e média.



## Capítulo 3

# Geometria diferencial e análise de imagens

A toda imagem  $h : U \rightarrow \mathbb{R}$  corresponde naturalmente uma superfície de Monge através da identificação entre a função  $h$  e o seu gráfico  $S = \{(u, h(u)), u \in U\}$ . A associação entre imagens e superfícies nos permite aplicar métodos da geometria diferencial à análise e síntese de imagens.

### 3.1 Segmentação perceptual de uma imagem

A equivalência entre imagens e funções dada pela nossa escolha do modelo matemático para as imagens, a cada imagem corresponde uma função e vice-versa, faz parecer que o estudo das imagens reduz-se simplesmente ao estudo das funções. Por que, então, as imagens são especiais?

A resposta é que imagens naturais apresentam grande quantidade de redundância. Para quase todos os pontos de uma imagem natural, a diferença entre o valor de um pixel e seus vizinhos é muito pequena ou simplesmente nenhuma. É essa redundância que os métodos de codificação procuram detectar e eliminar.

Com o objetivo de melhor explorar os diferentes tipos de redundância que podem ocorrer em uma imagem, podemos segmentar o suporte da imagem,  $h : U \rightarrow \mathbb{R}^n$ , em três regiões de acordo com o nível de redundância perceptual na vizinhança de cada ponto. Indicaremos essas regiões por 0D, 1D e 2D.

**Definição 3 (Segmentação perceptual de uma imagem).** Sejam  $h : U \rightarrow \mathbb{R}^n$  uma imagem e  $U \subset \mathbb{R}^2$ , diremos que:

1.  $u \in 0D$ , se  $h$  é constante em uma vizinhança de  $u$ ;
2.  $u \in 1D$ , se  $u \notin 0D$  e existe uma decomposição de uma vizinhança de  $u$  como a união de segmentos de retas paralelas, e  $h$  é constante ao longo de cada um desses segmentos;
3.  $u \in 2D$ , se nenhuma das duas condições anteriores se verifica, isto é,  $h$  apresenta variações ao longo de todas as direções em uma vizinhança de  $u$ .

Chamaremos de segmentação perceptual da imagem  $h$ , à segmentação regular de  $U$  obtida a partir dos conjuntos 0D, 1D e 2D.

Por um abuso da notação indicaremos os conjuntos 0D, 1D e 2D e suas versões regularizadas pelo mesmo símbolo. De acordo com a terminologia introduzida em (Zetsche *et al.*, 1993), diremos que as componentes 0D, 1D e 2D da segmentação perceptual da imagem têm *dimensão intrínseca zero, um e dois*, respectivamente. Do ponto de vista perceptual, a componente 0D corresponde às regiões onde há pouca ou nenhuma variação na intensidade da imagem, a componente 1D corresponde às vizinhanças de arestas, enquanto a componente 2D corresponde às regiões de grande variação da imagem. Imagens naturais são predominantemente 0D. A figura 3.1 ilustra a segmentação perceptual da imagem suavizada de um quadrado.

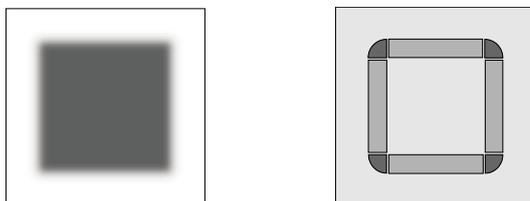


Figura 3.1: Segmentação perceptual: As regiões 1D são as vizinhanças das arestas e as regiões 2D são as vizinhanças dos vértices. Todos os outros pontos pertencem à região 0D.

A segmentação de uma imagem, nas suas regiões de dimensionalidade intrínseca, feita acima está estreitamente relacionada à classificação dos pontos de uma superfície em *planares, parabólicos, elípticos e hiperbólicos* dada pela geometria diferencial clássica. Neste caso, a classificação dos pontos da superfície de Monge da imagem.

## 3.2 Segmentação geométrica de uma imagem

Podemos usar a classificação local da superfície de Monge associada à imagem  $h : U \rightarrow \mathbb{R}$  para obter uma segmentação similar àquela dada pela dimensão intrínseca da imagem. Indicaremos as regiões desta segmentação por  $g0D$ ,  $g1D$  e  $g2D$ .

**Definição 4 (Segmentação Geométrica de uma Imagem).** Se  $h : U \rightarrow \mathbb{R}$  é uma imagem e  $U \subset \mathbb{R}^2$ , diremos que:

1.  $u \in g0D$ , se  $(u, h(u))$  é planar;
2.  $u \in g1D$ , se  $(u, h(u))$  é parabólico;
3.  $u \in g2D$ , se  $(u, h(u))$  é elíptico ou hiperbólico.

Chamaremos de segmentação geométrica da imagem  $h$ , à segmentação regularizada de  $U$  obtida a partir dos conjuntos  $g0D$ ,  $g1D$  e  $g2D$ .

**Observação 2.** A segmentação geométrica da imagem possui algumas propriedades bem interessantes. Pode-se mostrar, por exemplo, que a fronteira comum entre as regiões  $g0D$  e  $g1D$  é um segmento de reta, e a fronteira comum às regiões  $g1D$  e  $g2D$  é uma curva diferenciável, veja figura 3.2.

Novamente, abusaremos da notação e indicaremos os conjuntos  $g0D$ ,  $g1D$  e  $g2D$  e suas versões regularizadas pelo mesmo símbolo. A figura 3.2 ilustra a segmentação geométrica da imagem suavizada do quadrado em 3.2. Observe que, neste caso, as segmentações perceptual e geométrica coincidem, figuras 3.1 e 3.2, respectivamente.

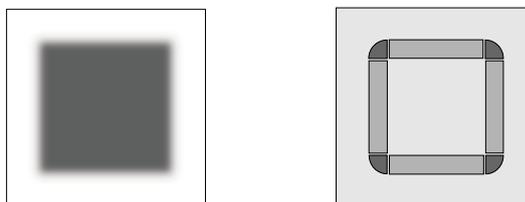


Figura 3.2: Segmentação geométrica: As regiões  $g1D$  são as vizinhanças das arestas e as regiões  $g2D$  são as vizinhanças dos vértices. Todos os outros pontos pertencem à região  $g0D$ .

A segmentação da imagem dada pela classificação dos pontos da superfície de Monge é similar à segmentação dada pela dimensionalidade intrínseca. A região  $g0D$  é basicamente a união de pedaços de plano, a região  $g1D$  é basicamente a união de pedaços de cones e cilindros, enquanto a região  $g2D$  tem comportamento não-linear em todas as direções. Apesar da similaridade, as regiões das duas segmentações podem não coincidir exatamente, como se pode concluir da possibilidade de pedaços de cones aparecerem na região  $2D$ . Entre outras, temos as seguintes inclusões que podem ser facilmente verificadas:

$$0D \subset g0D, \quad 1D \subset g0D \cup g1D, \quad 2D \subset g1D \cup g2D, \quad g2D \subset 2D$$

A similaridade entre as regiões dadas pela classificação local das superfícies e aquelas dadas pela dimensão intrínseca de uma imagem indicam que as medidas de curvatura da geometria diferencial são um bom ponto de partida para a detecção e eliminação de redundâncias em imagens.

Além disso, ambas as segmentações gozam de propriedades quase sempre desejadas em computação gráfica e processamento de imagens, como por exemplo, a invariância por translações, rotações. A segmentação perceptual também é invariante por alterações no brilho e contraste, enquanto que a segmentação perceptual pode sofrer mudar sob essas transformações.

### 3.3 Transformadas geométricas

No processo de análise de uma imagem, freqüentemente é necessário mudar de um modelo de imagem para outro. A operação de mudança entre dois modelos de imagem é chamada de *transformada*. Na seção anterior, vimos que a geometria diferencial pode oferecer um bom ponto de partida para a eliminação de redundâncias em uma imagem. Podemos então usar medidas de curvatura para construir uma transformada do espaço de imagens. Neste ponto aparecem dois problemas:

1. A seleção das medidas de métrica e curvatura a serem usadas na transformada;
2. A reconstrução da imagem original a partir da imagem transformada.

**Seleção da transformada.** A geometria diferencial é muito rica em funções que podem ser usadas como transformadas não-lineares, como por exemplo: as *curvaturas gaussiana e média*; os *auto-valores da aplicação normal de Gauss*; os *coeficientes das primeira e segunda formas fundamentais*; etc. Entretanto, algumas dessas funções podem não dar resposta nula para imagens que são perceptualmente 0D ou 1D. Isto pode ocorrer, por exemplo, se forem usadas, diretamente, os auto-valores da aplicação normal de Gauss ou a curvatura média.

**Reconstrução da imagem original.** Para um grande número de aplicações a reconstrução da imagem original pode ser de importância fundamental, como por exemplo na compressão de imagens. É interessante notar que para transformações lineares, o cálculo da transformada inversa é imediato: A inversa de uma transformação linear é outra transformação linear. Para transformações não-lineares, este pode ser um problema de solução extremamente complicada. Porém, para outras aplicações a recuperação exata da imagem original pode não ser necessária.

As restrições que as medidas de métrica ou curvatura impõem sobre a forma das superfícies é um dos principais pontos de pesquisa em geometria diferencial. O conhecimento nessa direção se encontra largamente desenvolvido. O *Teorema Fundamental da Teoria Local das Superfícies* afirma que os coeficientes  $E, F, G, e, f, g$ , da primeira e segunda forma fundamental, são suficientes para determinar a superfície a menos de um movimento rígido. Portanto, pelo menos do ponto de vista teórico, algum tipo de transformada geométrica invertível é possível. Todavia, existe um alto grau de correlação entre estes coeficientes. Podemos ver esta redundância através da necessidade de equações extras entre os coeficientes para assegurar a compatibilidade entre eles. Tais equações são conhecidas como *Equações de Gauss e Mainardi-Codazzi*.

Escolher uma transformada geométrica invertível e que elimine a correlação existente entre os coeficientes da primeira e segunda forma fundamental pode ser um problema difícil. Por exemplo, o catenóide e o helicóide, figura 3.3, são duas superfícies isométricas e portanto possuem os mesmos coeficientes métricos  $E, F, G$  e a mesma curvatura gaussiana. Além disso, como ambas são superfícies mínimas,

possuem a mesma curvatura média (igual a zero). Mesmo no caso compacto podemos ter esse tipo de comportamento como ilustra a figura 3.3.

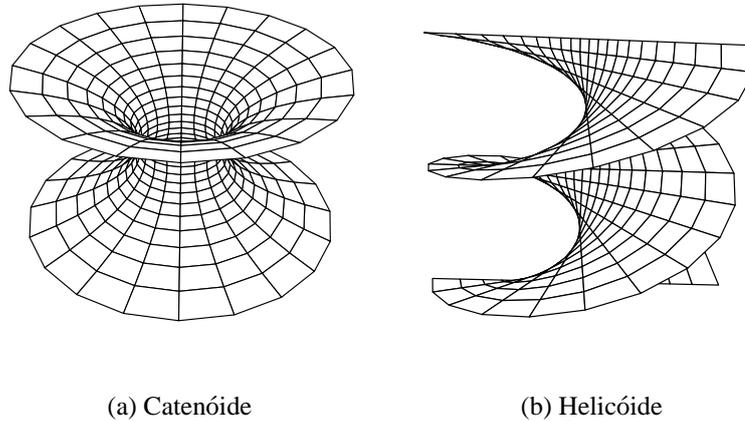


Figura 3.3: O catenóide e helicóide são duas superfícies mínimas isométricas. Portanto, possuem os mesmos coeficientes métricos  $E, F, G$ , e as mesmas curvaturas gaussianas e média.

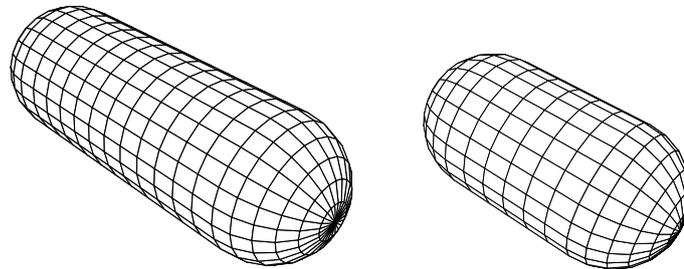


Figura 3.4: Superfícies compactas distintas com as mesmas curvaturas gaussianas e média.

Na seção anterior, vimos ainda que a segmentação de uma imagem a partir da classificação local das superfícies é muito próxima daquela obtida por meio de características perceptuais. Tal classificação é feita essencialmente a partir do sinal das curvaturas gaussianas e média da superfície. Se  $K$  e  $H$  indicam as curvaturas gaussianas e média da imagem, temos essencialmente que:

1.  $u \in g0D$  se  $K(u) = H(u) = 0$ ;
2.  $u \in g1D$  se  $K(u) = 0$  e  $H(u) \neq 0$ ;
3.  $u \in g2D$  se  $K(u) \neq 0$ .

Em (Zetsche *et al.*, 1993) os autores observaram que, como podemos ver acima, a curvatura gaussiana dá resposta nula para pontos da imagem que possuem dimensão intrínseca zero ou um. Baseados nesta informação, propuseram esta medida geométrica como um bom filtro não-linear para estudos envolvendo a percepção humana. Vale a pena observar que a curvatura gaussiana também pode dar resposta nula para alguns pontos que possuem dimensão intrínseca dois, por exemplo, se a imagem é parte de um cone.

Uma vez que apenas a curvatura gaussiana não é suficiente para determinar uma superfície, e portanto uma imagem, surge uma questão fundamental:

**Questão 1.** Quais as restrições extras necessárias para garantir a reconstrução da imagem?

A questão 1 será o principal tópico de interesse na seção e capítulos seguintes.

### 3.4 Reconstrução a partir dos pontos de curvatura não-nula

A proposição a seguir mostra que podemos recuperar uma imagem partindo apenas dos seus pontos que apresentam características 2D, como foi proposto em (Zetsche *et al.*, 1993) e (Barth *et al.*, 1993).

**Teorema 1.** Uma superfície compacta fica completamente determinada pelos seus pontos de curvatura gaussiana não-nula.

Se  $S$  é uma superfície com curvatura gaussiana  $K = 0$ . Indicaremos por  $F$  o conjunto dos pontos planares de  $S$  e por  $P = S - F$  o conjunto dos pontos parabólicos.

Observe que se  $p$  é um ponto da superfície e  $k_1, k_2$  são os auto-valores da aplicação normal de Gauss, então  $k_1(p) = k_2(p) = 0$  é uma propriedade fechada. Portanto, os conjuntos  $F$  e  $P$  são, respectivamente, fechado e aberto em  $S$ .

Para demonstrar o teorema 1 precisamos das duas proposições a seguir. Elas dão informações qualitativas sobre as linhas assintóticas e curvatura média de uma superfície de curvatura gaussiana nula. As demonstrações podem ser encontradas em (do Carmo, 1976).

**Proposição 5.** A única linha assintótica que passa por um ponto parabólico  $p \in P \subset S$  de uma superfície  $S$  de curvatura gaussiana  $K \equiv 0$  é um segmento (aberto) de reta contido em  $S$ .

**Proposição 6.** Se  $s$  é o comprimento de arco da linha assintótica  $r$  passando por um ponto parabólico  $p \in P \subset S$  de uma superfície  $S$  de curvatura gaussiana  $K \equiv 0$  e  $H(s)$  é a curvatura média ao longo de  $r$ , então  $H(s) = \frac{1}{as+b}$  para algum par  $a, b \in \mathbb{R}$ .

Estes dois resultados são necessários para a demonstração do teorema 1 que apresentaremos a seguir.

**Demonstração do teorema 2.** Seja  $S$  uma superfície compacta e  $R$  o fecho topológico do conjunto de pontos de curvatura gaussiana não-nula.  $R \neq \emptyset$  visto que  $S$  é compacta. Indique por  $S'$  o complementar de  $R$  em  $S$ , isto é,  $S' = S - R$ .  $S'$  é uma superfície de curvatura gaussiana  $K \equiv 0$ .

A seguir mostraremos que a direção dos auto-vetores da aplicação normal de Gauss na fronteira de  $R$  determinam a superfície fora de  $R$ .

Se  $p \in P \subset S'$  é um ponto parabólico e  $r$  é a linha assintótica maximal de  $S'$  passando por  $p$ , então pela proposição 5,  $r$  é um segmento de reta. Como  $S$  é compacta,  $r$  é limitado. Sejam  $p_1$  e  $p_2$  os pontos extremos de  $r$ . Podemos usar a proposição 6 para concluir que a curvatura média de  $S$  satisfaz,

$$H(s) = \frac{1}{as + b}$$

ao longo de  $r$ . Pela continuidade de  $H$ , temos  $H(p_1) \neq 0$  (o mesmo argumento se aplica a  $H(p_2)$ ). Logo,  $p_1$  não é planar e como  $K = 0$  é uma propriedade fechada, concluimos que  $p_1$  é parabólico.

Pela maximalidade de  $r$ , temos que  $p_1$  pertence à fronteira  $\partial S' = \partial R$ . Uma vez que  $p_1$  não é umbílico, pois nesse caso seria planar, a direção do auto-vetor, da derivada da aplicação normal de Gauss, correspondente ao auto-valor zero, está unicamente determinada e coincide com a direção de  $r$ . Isto determina  $r$  unicamente para cada ponto parabólico de  $\partial S'$ .

Por outro lado, como a derivada da aplicação normal de Gauss é identicamente nula em  $F$ , as componentes conexas de  $F$  são pedaços de plano e portanto ficam determinadas por um ponto em  $\partial R$  mais o valor do vetor normal nesse ponto. Isto conclui a prova do teorema.  $\square$

É interessante notar que a prova acima é construtiva, isto é, mostra-se como se pode reconstruir a superfície a partir dos pontos de curvatura gaussiana não-nula. Para isto, basta seguir as linhas de curvatura principal nula que são segmentos de reta começando no bordo do conjunto de pontos de curvatura gaussiana não-nula.

A demonstração acima vale apenas para superfícies compactas sem bordo. Em geral, o teorema é falso para superfícies com bordo, como mostra o exemplo de um tronco de cilindro (a porção de um cilindro entre dois cones paralelos). Nesse caso, a curvatura gaussiana é identicamente nula e a superfície não pode ser reconstruída a partir de seus pontos de curvatura gaussiana não-nula. Mesmo quando incluímos os pontos de bordo da superfície ainda temos duas possibilidades para esta: o tronco de cone ou duas elipses em planos paralelos. Uma extensão do resultado para gráficos compactos com bordo seria de interesse prático. Além disso, certamente é falso para superfícies completas. Para verificar esta última afirmação, basta notar que tomando variação em uma única direção podemos construir uma infinidade de superfícies de curvatura gaussiana nula.

Para aplicar o teorema acima para imagens, que em nosso modelo são superfícies compactas com bordo, devemos estender essas superfícies de modo a ter uma superfície sem bordo. Para isto, imagine a imagem com a tampa de uma caixa. Suavizando os cantos e arestas da caixa, podemos então substituir a imagem por uma

superfície compacta. Durante esse processo introduzimos pontos de curvatura não-nula ao longo da fronteira da imagem. Este processo é, portanto, equivalente a reconstruir a imagem a partir dos pontos de curvatura gaussiana mais os pontos de bordo da imagem e as direções normais nesses pontos.

# Capítulo 4

## Reconstrução

No capítulo anterior, mostramos que o conjunto dos pontos onde a curvatura gaussiana não se anula caracteriza a superfície. Neste capítulo, estudaremos as peculiaridades inerentes a um algoritmo de reconstrução de uma imagem a partir de seus pontos  $2D$  e apresentamos um algoritmo simples que reconstrói a imagem a partir de seus pontos  $1D$  e  $2D$ .

A possibilidade de reconstrução de uma imagem através desses pontos foi estudada originalmente em (Barth *et al.*, 1993). Embora os autores mostrem substancial evidência da possibilidade de reconstrução da imagem por meio desse conjunto de dados, os algoritmos apresentados reconstróem a imagem apenas de forma parcial.

Uma diferença fundamental entre filtros lineares e não-lineares ocorre no que diz respeito à reconstrução da imagem. Para transformações lineares o problema de reconstrução se reduz a encontrar soluções de sistemas de equações lineares e equivale, portanto, à inversão de uma matriz ou ao cálculo da pseudo-inversa no caso em que a transformação não é invertível. A inversão de filtros não-lineares envolve a solução de sistemas de equações não-lineares e por conseguinte é um problema muito mais difícil.

### 4.1 Cálculo da segmentação geométrica

Como vimos na seção 2.6, uma imagem  $h : U \rightarrow \mathbb{R}$  está naturalmente identificada com uma superfície de Monge  $S$  através da parametrização

$$(u, v) \mapsto f(u, v) = (u, v, h(u, v)).$$

Para esta superfície temos,

$$f_u = (1, 0, h_u) \quad \text{e} \quad f_v = (0, 1, h_v),$$

de onde se pode calcular os coeficientes da primeira forma fundamental

$$\begin{aligned} E &= \langle f_u, f_u \rangle = 1 + h_u^2 \\ F &= \langle f_u, f_v \rangle = h_u h_v \\ G &= \langle f_v, f_v \rangle = 1 + h_v^2. \end{aligned}$$

Usando-se o produto vetorial para calcular  $N$  a partir de  $f_u$  e  $f_v$ , obtém-se

$$N = \frac{1}{\sqrt{1+h_u^2+h_v^2}}(-h_u, -h_v, 1).$$

De onde se pode calcular os coeficientes da segunda forma fundamental

$$\begin{aligned} e &= -\langle N_u, f_u \rangle = \langle N, f_{uu} \rangle = \frac{h_{uu}}{\sqrt{1+h_u^2+h_v^2}} \\ f &= -\langle N_u, f_v \rangle = \langle N, f_{uv} \rangle = \frac{h_{uv}}{\sqrt{1+h_u^2+h_v^2}} \\ g &= -\langle N_v, f_v \rangle = \langle N, f_{vv} \rangle = \frac{h_{vv}}{\sqrt{1+h_u^2+h_v^2}}. \end{aligned}$$

Em particular, para superfícies de Monge temos:

$$A = \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} 1+h_u^2 & h_u h_v \\ h_u h_v & 1+h_v^2 \end{pmatrix} \quad (4.1)$$

e

$$B = \begin{pmatrix} e & f \\ f & g \end{pmatrix} = \frac{1}{\sqrt{1+h_u^2+h_v^2}} \begin{pmatrix} h_{uu} & h_{uv} \\ h_{uv} & h_{vv} \end{pmatrix}. \quad (4.2)$$

Podemos, portanto, calcular facilmente as curvaturas gaussiana e média de  $S$  a partir da relação

$$-dN = A^{-1}B, \quad (4.3)$$

obtendo-se

$$K = \frac{h_{uu}h_{vv} - h_{uv}^2}{(1+h_u^2+h_v^2)^2} \quad (4.4)$$

e

$$2H = \frac{(1+h_u^2)h_{vv} - 2h_u h_v h_{uv} + (1+h_v^2)h_{uu}}{(1+h_u^2+h_v^2)^{3/2}}. \quad (4.5)$$

Uma análise mais detalhada da equação 4.3 permite obter fórmulas mais simples para o cálculo da segmentação geométrica. Neste caso, a matriz  $-dN$  é o produto de uma matriz invertível de determinante positivo pela matriz

$$\text{Hess } h = \begin{pmatrix} h_{uu} & h_{uv} \\ h_{uv} & h_{vv} \end{pmatrix}$$

chamada *Hessiana* da função  $h$ . Como  $-dN$  e  $\text{Hess } h$  diferem pelo produto de uma função invertível temos a seguinte proposição.

**Proposição 7.** Seja  $h : U \rightarrow \mathbb{R}$  uma função diferenciável e  $p = (u, v, h(u, v))$  um ponto na superfície de Monge de  $h$ . Então:

1.  $p$  é elíptico se e apenas se  $\det \text{Hess } h > 0$ ;
2.  $p$  é hiperbólico se e apenas se  $\det \text{Hess } h < 0$ ;
3.  $p$  é parabólico se e apenas se  $\det \text{Hess } h = 0$  e  $\Delta h \neq 0$ ;
4.  $p$  é planar se e apenas se  $\det \text{Hess } h = \Delta h = 0$ .

Onde,  $\det \text{Hess } h = h_{uu}h_{vv} - h_{uv}^2$  é o determinante da Hessiana de  $h$  e  $\Delta h = h_{uu} + h_{vv}$  é o Laplaciano.

A equação 4.4 mostra que o determinante da Hessiana difere da curvatura de Gauss por um fator que depende da quarta potência da norma do gradiente  $\nabla h = (h_u, h_v)$  da imagem. Portanto, embora ambas tenham o mesmo sinal e se anulem nos mesmos pontos, o determinante da Hessiana produz repostas fortes nas vizinhas das arestas enquanto a curvatura gaussiana não. Se no processo de cálculo da segmentação geométrica estiver envolvido algum tipo de limiar podem ocorrer diferenças substanciais entre os resultados obtidos.

A figura 4.1 mostra a segmentação geométrica de uma imagem. Os pontos em tom médio de cinza representam os pontos descartados. Os pontos em branco e preto são os pontos da imagem, em sua cor original, selecionados para o respectivo segmento.

## 4.2 Arestas e curvatura gaussiana

Nesta seção vamos relacionar a noção de aresta de uma imagem com os zeros da curvatura gaussiana da superfície de monge associada a esta imagem. Com isto, poderemos esclarecer algumas das peculiaridades intrínsecas a um algoritmo que reconstrua a imagem partindo apenas de seus pontos  $2D$ .

Existem várias definições de aresta na literatura. Um modelo comumente aceito consiste em descrever as arestas como os pontos de máximo local de

$$\frac{dh}{dt} = |\nabla h|^2,$$

onde  $h(t)$  indica a restrição de  $h$  às linhas integrais do gradiente  $\nabla h = (h_u, h_v)$  da função  $h$ . Portanto, se  $p \in U$  é um ponto de aresta da imagem e indicarmos por  $\nu = \nabla h(p)$ , temos

$$\frac{\partial}{\partial \nu} \langle \nabla h, \nabla h \rangle = 0$$

ou equivalentemente

$$\langle \text{Hess } h \nabla h, \nabla h \rangle = 0 \tag{4.6}$$

Portanto,  $\nabla h$  é uma direção assintótica de  $\text{Hess } h$ . Isto só pode ocorrer em duas possibilidades:

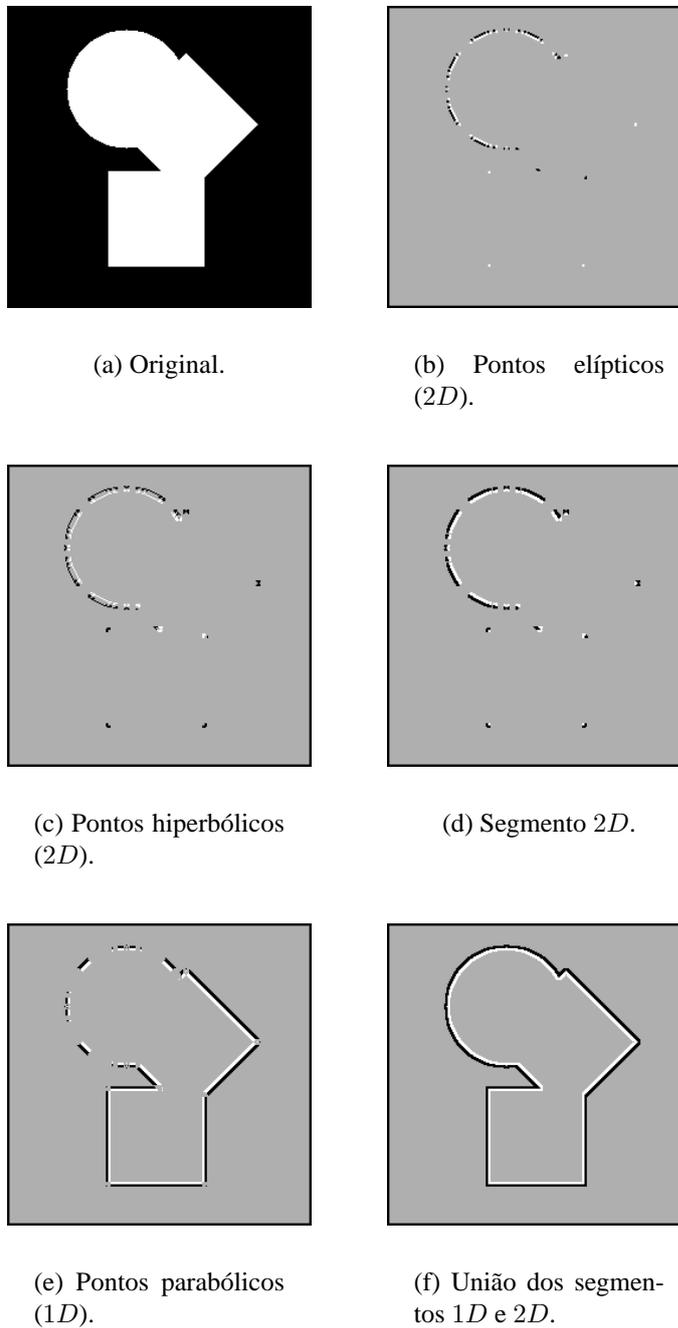


Figura 4.1: Segmentação geométrica: os pontos em cinza representam os pontos planares, os outros pontos na cor original da imagem (preto e branco) são os pontos selecionados para cada segmento.

1.  $p$  é um ponto parabólico de  $S$  e  $\nabla h$  é um autovetor de Hess  $h$  associado ao autovalor zero;
2.  $p$  é um ponto hiperbólico de  $S$  e  $\nabla h$  uma das direções assintóticas de Hess  $h$ .

De onde se conclui que as arestas estão contidas na união das regiões hiperbólicas e parabólicas da superfície de Monge de  $h$ . Portanto, a exclusão dessas necessariamente exclui as arestas da image. A sensibilidade do sistema visual humano a estas curvas é bem conhecida. É a exclusão das arestas a provável causa das distorções obtidas ao se tentar reconstruir uma imagem sem levar em consideração as regiões hiperbólicas e parabólicas. Veja por exemplo as reconstruções em (Barth *et al.*, 1993; Zetsche *et al.*, 1993).

A figura 4.1 ilustra as regiões elíptica, hiperbólica (2D), parabólica e, por último, a união das três regiões curvas.

### 4.3 Reconstrução

O teorema 1 da secção 3.4 afirma que uma imagem pode ser reconstruída a partir do conhecimento parcial de seus valores desde que tenham sido preservados os pontos onde a *curvatura gaussiana* é não-nula. Além disso, o teorema indica um processo para a reconstrução da superfície. Na prática, entretanto, esse processo de reconstrução é inviável devido aos erros numéricos resultantes de arredondamentos.

A imagem pode ser reconstruída facilmente se, além de conhecermos seus valores nas regiões de curvatura gaussiana não-nula (g2D), conhecermos também a imagem nas regiões parabólicas (g1D). Neste caso, os pontos perdidos pertencem a pedaços de planos e portanto, podemos usar interpolação linear ao longo das direções  $u$  e  $v$  para obter

$$h_{ij} = 1/2(h_{i-1j} + h_{i+1j})$$

e

$$h_{ij} = 1/2(h_{ij-1} + h_{ij+1})$$

de onde obtemos,

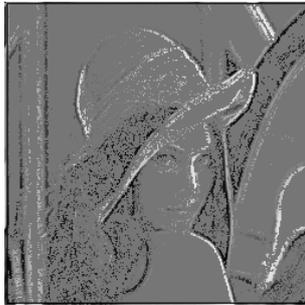
$$h_{i-1j} + h_{i+1j} + h_{ij-1} + h_{ij+1} - 4h_{ij} = 0 \quad (4.7)$$

O sistema 4.7 é linear e sua solução reconstrói completamente a imagem. A figura 4.2(e) mostra a imagem de *Lena* reconstruída usando o método descrito acima.

No cálculo da segmentação usa-se um limiar  $\varepsilon$  para decidir se um ponto é planar ou não. Em conseqüência, as regiões classificadas como planares não o são de forma exata. Essas regiões são reconstruídas como pedaços de superfícies hiperbólicas.



(a) Original.



(b) Pontos parabólicos (1D).



(c) Pontos elípticos e hiperbólicos (2D).



(d) União da regiões 1D e 2D.



(e) Reconstrução a partir dos pontos em 4.2(d)

Figura 4.2: Segmentação e reconstrução da imagem de *Lena*. Os pontos em tom médio de cinza representam os pontos planares que foram descartados. Os pontos parabólicos, elípticos e hiperbólicos foram mantidos na sua cor original.

## 4.4 Reconstrução a partir da curvatura gaussiana

O problema de reconstruir uma superfície a partir da curvatura gaussiana é equivalente a encontrar as soluções da equação diferencial

$$h_{uu}h_{vv} - h_{uv}^2 = K(u, v)(1 + h_u^2 + h_v^2)^2. \quad (4.8)$$

Como resolver a equação 4.8 é um problema difícil, e como estamos interessados principalmente em aplicações ao processamento de imagens, estudaremos a viabilidade de resolver a equação 4.8 do ponto de vista discreto.

### 4.4.1 Filtros quadráticos

Dado um operador bilinear  $\mathcal{B} : \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{S}$ , onde  $\mathcal{S}$  indica o espaço de imagens, chamaremos de *filtro quadrático* ao operador  $B : \mathcal{S} \rightarrow \mathcal{S}$  dado por  $B(h) = \mathcal{B}(h, h)$ .

Uma maneira simples de construir tais filtros é escolher dois filtros lineares  $S$  e  $T$  e definir  $\mathcal{B}(h, g) = ShTg$ . É imediato que  $\mathcal{B}$  é um operador bilinear e o filtro associado é  $B(h) = ShTh$ . Se os filtros  $S$  e  $T$  são contínuos e espacialmente invariantes, então

$$B(h * \delta) = S(h * \delta)T(h * \delta) = (h * S\delta)(h * T\delta), \quad (4.9)$$

o que mostra que essa classe de filtros quadráticos fica determinada pelas funções de resposta de impulso,  $s = S\delta$  e  $t = T\delta$  dos respectivos filtros. A discretização de  $B$  também é imediata, se a seqüência  $x$  ( $x_i = h(i)$ ) representa uma discretização de  $h$  e

$$(Sh)_n = (h * s)_n = \sum s_{n-i} x_i \quad (4.10)$$

e

$$(Th)_n = (h * t)_n = \sum t_{n-j} x_j \quad (4.11)$$

então

$$B(h)_n = \sum_{i,j} b_{ij}^n x_i x_j, \quad (4.12)$$

onde

$$b_{ij}^n = 1/2\{s_{n-i}t_{n-j} + s_{n-j}t_{n-i}\} \quad (4.13)$$

A classe dos filtros quadráticos é bastante rica. Dois exemplos importantes que podem ser colocados nesse contexto são o *determinante Hessiano* e o quadrado do *gradiente* da imagem.

Podemos escrever o determinante Hessiano como filtro quadrático a partir dos operadores lineares

$$D_{uu}(h) = h_{uu}, \quad D_{vv}(h) = h_{vv} \quad \text{e} \quad D_{uv}(h) = h_{uv}, \quad (4.14)$$

ou seja

$$D(h) = h_{uu}h_{vv} - h_{uv}^2 = D_{uu}(h)D_{vv}(h) - D_{uv}(h)D_{uv}(h). \quad (4.15)$$

O quadrado do gradiente é obtido a partir dos operadores lineares

$$D_u(h) = h_u \quad \text{e} \quad D_v(h) = h_v \quad (4.16)$$

isto é

$$|\nabla h|^2 = h_u^2 + h_v^2 = D_u(h)D_u(h) + D_v(h)D_v(h), \quad (4.17)$$

Portanto, do ponto de vista discreto, a curvatura gaussiana pode ser vista como o quociente entre duas séries

$$K_{pq} = \frac{D_{pq}(x)}{(1 + |\nabla_{pq} x|^2)^2}, \quad (4.18)$$

onde  $x = (x_{ij})$  é a discretização da imagem  $h$ .

Um filtro que é o quociente de duas séries, como a discretização da curvatura gaussiana feita acima, é chamado de *filtro racional*, (Zayed, 1993).

#### 4.4.2 Inversão de filtros racionais

Do ponto de vista computacional, podemos usar apenas um subconjunto finito do conjunto de amostras da imagem e a série da equação 4.12 torna-se um polinômio. Em particular a fórmula 4.18 para a curvatura discreta torna-se o quociente de dois polinômios. Desse ponto de vista, recuperar a imagem a partir das curvaturas discretas é equivalente a resolver o sistema de  $mn$  equações polinomiais

$$D_{pq}(x) = K_{pq}(1 + |\nabla_{pq} x|^2)^2 \quad (4.19)$$

a  $mn$  incógnitas, onde  $m$  e  $n$  são as dimensões da imagem e  $D_{pq}(x)$ ,  $|\nabla_{pq} x|^2$  são polinômios homogêneos de grau dois na variável  $x = (x_{ij})_{m \times n}$ .

O espaço adequado para o estudo do sistema de equações 4.19 é o *espaço projetivo complexo* de dimensão  $mn$ ,  $\mathbb{C}\mathbb{P}^{mn}$ . Para estudar o sistema de equações 4.19 no espaço projetivo complexo, faz-se a mudança de variáveis

$$z_{ij} = \frac{x_{ij}}{y}, \quad (4.20)$$

obtendo-se o sistema de equações homogêneas de grau quatro

$$y^2 D_{pq}(z) = K_{pq}(y^2 + |\nabla_{pq} z|^2)^2 \quad (4.21)$$

O *Teorema de Bézout* permite concluir que dada uma seqüência  $(K_{pq})$ , o sistema de equações 4.21 sempre possui uma solução. De fato, o teorema de Bézout afirma que o sistema 4.21 possui  $4^{mn}$  soluções contadas com multiplicidade ou então possui

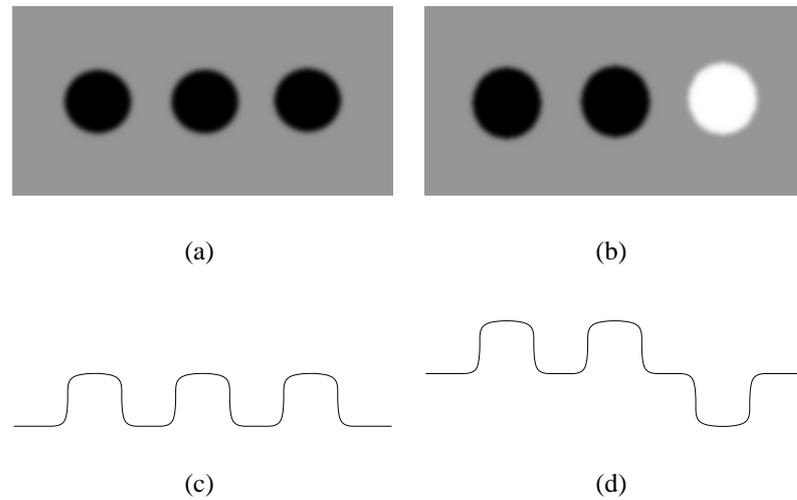


Figura 4.3: Imagens distintas com mesma curvatura gaussiana. As figuras (c) e (d) ilustram o perfil das superfícies de Monge das imagens em (a) e (b).

um *continuum* de soluções. O Teorema de Bézout garante que o erro cometido no cálculo das curvaturas não leva a um sistema de equações sem solução. Entretanto, deve-se manter alguma cautela, pois o teorema aceita soluções complexas, bem como soluções no infinito, para o sistema de equações.

Como o sistema de equações 4.21 tipicamente apresenta um grande número de soluções, precisamos impor equações extras de modo que a imagem seja a única solução do sistema a menos de multiplicidades. Uma possibilidade, para esse conjunto de equações extras, é supor que a imagem é conhecida na fronteira do conjunto suporte da mesma, (Barth *et al.*, 1993). Entretanto essa condição de contorno não é suficiente para garantir a unicidade da solução, como mostra a figura 4.4.2.

Supor a imagem conhecida na fronteira de seu conjunto suporte também não é uma condição plausível do ponto de vista perceptual. De fato, a acuidade do sistema visual diminui à proporção que se aproxima do bordo da imagem percebida pelo olho humano, uma vez que a densidade de células fotossensíveis diminui à medida que se afasta da fóvea (região do olho responsável pela visão central).

Portanto, descobrir quais condições de contorno são suficientes para reconstruir a imagem diretamente da sua curvatura gaussiana permanece um problema importante ainda em aberto.



# Capítulo 5

## Quantização

Como vimos na introdução, o processamento de uma imagem pelo sistema visual humano apresenta características altamente não-lineares. Em particular, a distância perceptual entre as cores apresenta características logarítmicas. Estas características têm grande influência em processos de discretização do espaço de cor de uma imagem, comumente chamado de *quantização* da imagem.

Neste capítulo apresentamos um estudo detalhado de um processo de quantização conhecido como quantização ótima. Também será feita uma análise do *algoritmo do corte mediano* do ponto de vista da teoria da informação. Começaremos discutindo rapidamente uma generalização natural do conceito de segmentação de uma imagem.

### 5.1 Partições

Se na definição de segmentação exigirmos apenas que  $U = \bigcup U_\lambda$  e  $U_\alpha \cap U_\beta = \emptyset$  sempre que  $\alpha \neq \beta$ , então a família é chamada uma *partição*. Toda partição define uma relação de equivalência através da relação:

$$u \sim v \quad \text{se } u \text{ e } v \text{ pertencem ao mesmo segmento.}$$

Reciprocamente, toda relação de equivalência induz uma partição de  $U$ , onde cada segmento é uma classe de equivalência.

### 5.2 Partições e quantização

Se  $f : U \rightarrow \mathbb{R}^n$  é uma função e  $\mathcal{V} = \{V_\lambda\}$  é uma partição do  $\mathbb{R}^n$  então  $f$  induz uma partição  $\mathcal{U} = \{U_\lambda\}$  de  $U$ , onde  $U_\lambda = f^{-1}(V_\lambda)$ . De modo recíproco, se  $U = \bigcup U_\lambda$  é uma partição, podemos definir uma função  $f_q(x) = q_\lambda$  assumindo valor constante  $q_\lambda$  em cada segmento  $U_\lambda$ , portanto toda partição induz uma função de classificação em  $U$ . Chamaremos a função  $f_q$  de *função indicadora* da partição.

Se a partição de  $U$  é induzida por  $f$ , e  $q_\lambda \in V_\lambda$  então  $f_q$  é uma aproximação constante por partes de  $f$ , figura 5.1. O método de segmentar o suporte de uma imagem e aproximá-la por uma função constante em cada segmento é essencialmente

o processo conhecido como *quantização*.  $U_\lambda$  é chamado de *célula de quantização* e  $q_\lambda$  é chamado de *valor de quantização*. A função  $q : \mathbb{R}^n \rightarrow \{q_\lambda\}_\lambda$ , definida por  $q(y) = q_\lambda$  se  $y \in V_\lambda$ , é chamada de *função de quantização*. É imediato que  $f_q(x) = q(f(x))$ .

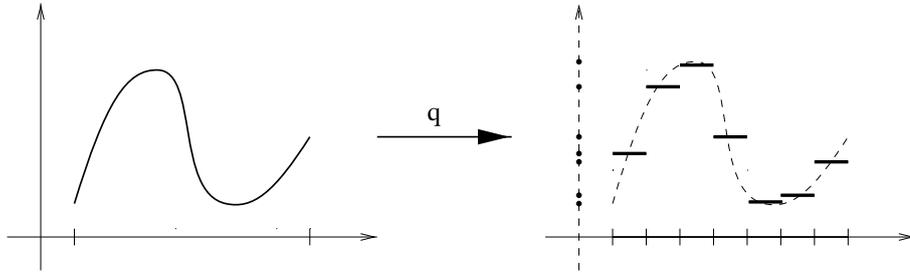


Figura 5.1: Segmentação, e quantização de uma função.

### 5.3 Partição e quantização binária

Um exemplo simples é a partição binária definida por uma função real. Se  $f : U \rightarrow \mathbb{R}$  é uma função e  $c \in \mathbb{R}$ , definimos uma partição de  $U$  por:

$$\begin{aligned} U_1 &= \{u \in U; \quad f(u) < c\} \\ U_2 &= \{u \in U; \quad f(u) \geq c\}. \end{aligned}$$

A figura 5.2 mostra a quantização binária de uma imagem cinza.

A despeito da simplicidade das partições binárias elas surgem em um grande número de aplicações como por exemplo no problema de *dithering*. De um modo geral, dada uma imagem cinza  $h : U \rightarrow \mathbb{R}$  o problema de *dithering* consiste em obter uma partição de  $U$  em dois subconjuntos de tal modo que a função indicadora desta partição satisfaz:

$$\int_V h(u) \, du = \int_V h_q(u) \, du \quad (5.1)$$

para toda vizinhança  $V \subset U$ .

### 5.4 Quantização ótima

Quando quantizamos uma imagem cometemos um erro de aproximação entre uma cor  $x$  e seu valor de quantização  $q(x)$ . O erro pontual pode ser avaliado através da escolha de uma *medida de distorção*  $d(x, y)$ , no espaço de cor da imagem. Se indicarmos esse erro por

$$\epsilon(x) = d(x, q(x)), \quad (5.2)$$



(a) Imagem cinza

(b) Imagem binária

Figura 5.2: Partição e quantização binária

podemos medir o impacto do erro cometido no processo de quantização, calculando o *valor esperado* para  $\epsilon(x)$ . Isto é,

$$E(d, q) = \int d(x, q(x))p(x) dx. \quad (5.3)$$

Onde,  $p(x)$  indica a *função de densidade de probabilidade* das cores da imagem.

Se  $E(d, q)$  é mínimo,  $q$  é um *quantizador ótimo* para a imagem  $h$  (com respeito a  $d$ ). Para se obter uma função de quantização que minimize a equação 5.3, é comum escolher o quadrado da distância euclidiana:

$$d(x, y) = |x - y|^2. \quad (5.4)$$

No caso unidimensional, se  $q_k$  indica valor de quantização para a célula  $(x_{k-1}, x_k]$ , temos

$$E(x_0, \dots, x_n, q_1, \dots, q_n) = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} |x - q_k|^2 p(x) dx. \quad (5.5)$$

De onde, derivando, se obtém que

$$x_k = \frac{q_k + q_{k+1}}{2} \quad \text{e} \quad q_k = \frac{\int_{x_{k-1}}^{x_k} xp(x) dx}{\int_{x_{k-1}}^{x_k} p(x) dx}. \quad (5.6)$$

Na discussão acima, fizemos duas hipóteses que simplificaram a solução do problema de quantização ótima unidimensional:

1.  $h$  possui uma função de densidade de probabilidade;

$$2. d(x, y) = |x - y|^2.$$

Entretanto, as hipóteses acima são muito restritivas. A saber, se a imagem é constante em uma região, a hipótese 1 não é satisfeita. A presença de regiões onde a imagem é constante é muito comum. Portanto, seria interessante não se fazer uso da hipótese 1. Por outro lado, também é interessante admitir distâncias mais gerais. Admitir tais distâncias permite levar em conta, no processo de quantização, a maneira como as pessoas percebem as cores presentes em uma imagem. A seguir discutiremos o problema de quantização ótima unidimensional de um ponto de vista bastante geral.

## 5.5 Função de distribuição de probabilidade

Se  $h : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  é uma imagem. Chama-se *função de distribuição de  $h$* , a função  $F : \mathbb{R} \rightarrow [0, 1]$ , definida por

$$F(x) = \text{área de } h^{-1}(-\infty, x]. \quad (5.7)$$

Pode-se mostrar facilmente que:

1.  $F(x) \geq 0$ ,  $\lim_{x \rightarrow -\infty} F(x) = 0$  e  $\lim_{x \rightarrow +\infty} F(x) = 1$ ;
2. Se  $x \geq y$  então  $F(x) \geq F(y)$ ;
3.  $F$  é contínua à direita, isto é  $\lim_{x \rightarrow a^+} F(x) = F(a)$ .

Portanto, a imagem  $h$  induz, através de  $F$ , uma *medida* (de Lebesgue-Stieltjes) na reta. A medida de um intervalo fechado é

$$m_h[a, b] = F(b) - F(a-), \quad (5.8)$$

onde  $F(a-) = \lim_{x \rightarrow a^-} F(x)$ . A medida dos outros intervalos têm fórmulas similares. Como  $m_h(\mathbb{R}) = 1$ ,  $m_h$  é uma *medida de probabilidade*, (James, 1981).

## 5.6 Medidas de distorção

De maneira geral, podemos escolher, como *medida de distorção* no  $\mathbb{R}^n$ , qualquer função  $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow R$  satisfazendo:

1.  $d(x, y) > 0$  se  $x \neq y$  e  $d(x, x) = 0$ ;
  2.  $d(x, y) = d(y, x)$ .
- Se, além disso,
3.  $d(x, z) \leq d(x, y) + d(y, z)$ ,

$d$  é uma *métrica*.

A propriedade 3 é importante, pois permite estimar a distorção, em um processo de quantização a vários passos, a partir da distorção em cada uma das etapas.

Se  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  é uma função injetiva e  $\bar{d}$  é uma medida de distorção em  $\mathbb{R}^n$ , então podemos definir uma nova medida de distorção pondo

$$d(x, y) = \bar{d}(\phi(x), \phi(y)).$$

No caso em que em  $\bar{d}$  é o quadrado da métrica euclidiana, temos

$$d(x, y) = |\phi(x) - \phi(y)|^2, \quad (5.9)$$

ou ainda de modo mais geral,

$$d(x, y) = |\phi(x) - \phi(y)|^r, \quad r \neq 0, \quad (5.10)$$

Se além disso,  $n = 1$  e  $\phi(x/y) = -\phi(y/x)$ , podemos definir

$$d(x, y) = |\phi(x/y)|, \quad (5.11)$$

supondo que a cor nula não aparece na imagem.

## 5.7 Quantização ótima unidimensional

A discussão da Lei de Weber, feita na secção 1.3, indica que a maneira como o sistema visual humano percebe as cores tem características não-lineares. Esse facto mostra que é importante estudar o problema de quantização ótima de um ponto de vista mais geral. Nessa secção abordaremos esse problema para imagens monocromáticas.

Dada uma imagem cinza  $h : [a, b] \rightarrow \mathbb{R}$ , trata-se de encontrar uma função de quantização  $q : [a, b] \rightarrow \{q_1, \dots, q_n\}$  tal que

$$E(d, q) = \int d(x, q(x)) dF(x) \quad (5.12)$$

seja mínima, onde  $F$  é a medida de probabilidade introduzida em 5.5.

**Proposição 8.** Se  $q : \mathbb{R} \rightarrow \{q_1, \dots, q_n\}$  é um quantizador ótimo para a imagem  $h$ , então  $x_k$  pode ser escolhido a meia distância de  $q_k$  e  $q_{k+1}$ , isto é,  $d(x_k, q_k) = d(x_k, q_{k+1})$ .

*Demonstração.* Podemos escrever a equação 5.12 como

$$D(x_0, \dots, x_n, q_1, \dots, q_n) = \int_{x_0}^{q_1} d(x, q_1) dF(x) + \sum_{k=1}^{n-1} \left( \int_{q_k}^{x_k} d(x, q_k) dF(x) + \int_{x_k}^{q_{k+1}} d(x, q_{k+1}) dF(x) \right) + \int_{q_n}^{x_n} d(x, q_n) dF(x).$$

Seja

$$I_k(y) = \int_{q_k}^y d(x, q_k) dF(x) + \int_y^{q_{k+1}} d(x, q_{k+1}) dF(x)$$

Reescrevendo obtemos

$$I_k(y) = C + \int_y^{q_{k+1}} g(x) dF(x)$$

Onde,

$$C = C(q_k, q_{k+1}) = \int_{q_k}^{q_{k+1}} d(x, q_k) dF(x) \quad \text{e} \quad g(x) = d(x, q_{k+1}) - d(x, q_k).$$

Se  $g(x_k) > 0$ , como  $g(q_{k+1}) = -d(q_k, q_{k+1}) \leq 0$ , podemos usar a continuidade de  $g$  para obter  $y_k$ , com  $x_k < y_k \leq q_{k+1}$ ,  $g(y_k) = 0$  e  $g(x) > 0$  no intervalo  $[x_k, y_k)$ , donde

$$I(x_k) = \int_{x_k}^{y_k} g(x) dF(x) + I(y_k) \geq I(y_k). \quad (5.13)$$

Se  $g(x_k) < 0$ , procedemos de maneira similar. Portanto, trocando  $x_k$  por  $y_k$  se necessário, temos  $d(x_k, q_k) = d(x_k, q_{k+1})$   $\square$

A equação 5.13 mostra que se não existirem intervalos de probabilidade zero, exceto os triviais, então obrigatoriamente  $d(x_k, q_k) = d(x_k, q_{k+1})$ , o que é um pouco mais do que foi enunciado.

**Corolário 1.** Seja  $q : \mathbb{R} \rightarrow \{q_1, \dots, q_n\}$  um quantizador ótimo para a imagem  $h$ , então:

1. Se  $d(x, y) = |\phi(x) - \phi(y)|^r$ , temos  $\phi(x_k) = \frac{\phi(q_k) + \phi(q_{k+1})}{2}$  ;
2. Se  $d(x, y) = |\phi(x/y)|$ , temos  $x_k^2 = q_k q_{k+1}$ .

*Demonstração.* Podemos supor que  $q_k \neq q_{k+1}$ , caso contrário, juntamos as células  $(x_{k-1}, x_k]$  e  $(x_k, x_{k+1}]$  em uma única célula. Portanto, usando a injetividade de  $\phi$ , se  $d(x, y) = |\phi(x) - \phi(y)|^r$ , segue que

$$\phi(x_k) - \phi(q_k) = \phi(q_{k+1}) - \phi(x_k),$$

o que mostra o item 1.

De modo análogo, se  $d(x, y) = |\phi(x/y)|$ , temos  $\phi(x_k/q_k) = -\phi(x_k/q_{k+1}) = \phi(q_{k+1}/x_k)$  e portanto

$$\frac{x_k}{q_k} = \frac{q_{k+1}}{x_k},$$

o que mostra o item 2.  $\square$

Por outro lado, se  $(x_{k-1}, x_k]$  é a célula de quantização de  $q_k$ , podemos escrever a equação 5.12 como

$$E(x_0, \dots, x_n, q_1, \dots, q_n) = \sum \int_{x_{k-1}}^{x_k} d(x, q_k) dF(x). \quad (5.14)$$

Como cada parcela é independente,  $q_k$  minimiza

$$E(d(X, c) \mid x_{k-1} < X \leq x_k) = \frac{\int_{x_{k-1}}^{x_k} d(x, c) dF(x)}{\int_{x_{k-1}}^{x_k} dF(x)}. \quad (5.15)$$

Um fato largamente conhecido em teoria da probabilidade é que

$$\min E(X - c)^2 = E(X - \mu)^2 \quad \text{e} \quad (5.16)$$

$$\min E|X - c| = E|X - m|, \quad (5.17)$$

onde  $\mu, m$  são, respectivamente, a *esperança* e uma *mediana* de  $X$ , (James, 1981). Portanto temos o seguinte:

**Teorema 2.** Se  $q : \mathbb{R} \rightarrow \{q_1, \dots, q_n\}$  é um quantizador ótimo para a imagem  $h$ , com células de quantização  $(x_{k-1}, x_k]$ , então:

1. Se  $d(x, y) = |\phi(x) - \phi(y)|$ ,  $q_k$  é a mediana de  $h$  no intervalo  $(x_{k-1}, x_k]$ , isto é

$$m_h(x_{k-1}, q_k] \geq 1/2 m_h(x_{k-1}, x_k] \quad \text{e} \quad m_h[q_k, x_k] \geq 1/2 m_h(x_{k-1}, x_k]; \quad (5.18)$$

2. Se  $d(x, y) = |\phi(x) - \phi(y)|^2$ ,  $\phi(q_k)$  é a média de  $\phi(h)$  no intervalo  $(x_{k-1}, x_k]$ , isto é

$$\phi(q_k) = \frac{\int_{x_{k-1}}^{x_k} \phi(x) dF(x)}{\int_{x_{k-1}}^{x_k} dF(x)}. \quad (5.19)$$

Em ambos os casos, temos

$$\phi(x_k) = \frac{\phi(q_k) + \phi(q_{k+1})}{2}. \quad (5.20)$$

*Demonstração.* Evidente a partir do corolário 1 e das equações 5.16 e 5.17.  $\square$

Um fato notável, mostrado no teorema 2, é que se  $d(x, y) = |\phi(x) - \phi(y)|$ , o valor de quantização é a mediana da célula de quantização e portanto não depende da função  $\phi$ .

Abaixo apresentamos a versão discreta das fórmulas do teorema 2 para  $\phi(x) = Cx$  ou  $\phi(x) = C \log(x)$ . No que se segue,  $z_j$  indica os valores da imagem,  $p_j$  o número de ocorrências da cor  $z_j$ , e  $n_k$  o total de ocorrências na célula  $(x_{k-1}, x_k]$ .

$$1. d(x, y) = |x - y|.$$

$$x_k = \frac{q_k + q_{k+1}}{2}, \quad q_k = \text{mediana de } \{z_j\}_{x_{k-1} < z_j \leq x_k}. \quad (5.21)$$

$$2. d(x, y) = |x - y|^2.$$

$$x_k = \frac{q_k + q_{k+1}}{2}, \quad q_k = \frac{1}{n_k} \sum_{x_{k-1} < z_j \leq x_k} p_j z_j. \quad (5.22)$$

$$3. d(x, y) = |\log(x) - \log(y)|.$$

$$x_k^2 = q_k q_{k+1}, \quad q_k = \text{mediana de } \{z_j\}_{x_{k-1} < z_j \leq x_k}. \quad (5.23)$$

$$4. d(x, y) = |\log(x) - \log(y)|^2.$$

$$x_k^2 = q_k q_{k+1}, \quad q_k^{n_k} = \prod_{x_{k-1} < z_j \leq x_k} z_j^{p_j}. \quad (5.24)$$

## 5.8 Implementação

Uma característica peculiar das fórmulas anteriores é que elas estão definidas de maneira recursiva. Uma vez conhecido os valores de  $x_1, \dots, x_n$  podemos calcular  $q_1, \dots, q_n$  e vice-versa. Pode-se mostrar que o algoritmo obtido pela atribuição de valores iniciais arbitrários para  $x_1, \dots, x_n$ , seguido do cálculo recursivo de  $q_1, \dots, q_n$  e  $x_1, \dots, x_n$ , é convergente, (Levine, 1985). No final deste capítulo apresentamos o resultado do processo de quantização da imagem da Lenna para 16, 8, 4 e 2 cores. Essas imagens foram geradas usando a implementação feita por Romildo José da Silva dos algoritmos acima descritos.

## 5.9 Quantização e ganho de informação

Dados dois intervalos  $J_2 \subset J_1$ , o ganho de informação de  $J_2$  relativo a  $J_1$  é definido como

$$G = \log\left(\frac{m(J_1)}{m(J_2)}\right), \quad (5.25)$$

onde  $m(a, b)$  é uma medida do intervalo  $(a, b)$ , (Resnikoff, 1987).

Em particular, para a medida  $m = m_h$  introduzida na secção 5.5, temos  $m_h(\mathbb{R}) = 1$  e portanto o ganho de informação de um intervalo  $J$  com relação a  $\mathbb{R}$  é

$$G = \log\left(\frac{1}{m_h(J)}\right) = -\log(m_h(J)). \quad (5.26)$$

Considere uma partição da reta por intervalos, obtida a partir de uma seqüência de números  $-\infty = x_0 < x_1 < x_2 < \dots < x_n = +\infty$  e sejam  $p_k = m_h(x_{k-1}, x_k]$ , ( $m_h(x_{n-1}, +\infty)$  se  $k = n$ ). Como  $m_h(\mathbb{R}) = 1$ , temos  $\sum_{k=0}^n p_k = 1$ , portanto  $p_k$  é uma probabilidade discreta e o *ganho esperado de informação* ou *entropia*, relativo a esta subdivisão da reta, é

$$I = I(x_1, \dots, x_{n-1}) = - \sum_{k=1}^n p_k \log(p_k). \quad (5.27)$$

$I$  mede a quantidade de informação de uma quantização da imagem  $h$ , com células de quantização  $(x_{k-1}, x_k]$ . Note que se  $x_{k-1} < y_k < x_k$ , então  $p_k = p'_{k-1} + p'_k$ , onde  $p'_{k-1} = m_h(x_{k-1}, y_k]$  e  $p'_k = m_h(y_k, x_k]$ , portanto

$$-p_k \log(p_k) = -(p'_{k-1} + p'_k) \log(p'_{k-1} + p'_k) \leq -p'_{k-1} \log(p'_{k-1}) - p'_k \log(p'_k), \quad (5.28)$$

donde,

$$I(x_1, \dots, x_{n-1}) \leq I(x_1, \dots, x_{k-1}, y_k, x_k, \dots, x_{n-1}), \quad (5.29)$$

o que corresponde à noção intuitiva de que teremos mais informação sobre a imagem se usarmos um número maior de células de quantização.

Por outro lado, cada parcela de  $I$  é uma função da forma  $f(x) = -x \log(x)$ , com  $x > 0$  e portanto satisfaz:

$$\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 1} f(x) = 0, \quad (5.30)$$

o que mostra que intervalos de probabilidade muito grande ou muito pequena contribuem pouco para o ganho médio de informação, os primeiros porque têm pouca informação e os últimos porque têm baixa probabilidade. Portanto, é natural perguntar: Em qual posição devemos colocar os pontos da seqüência  $\{x_k\}$  de modo que o ganho esperado de informação seja máximo?

**Proposição 9.** Sejam  $h : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  uma imagem e  $p_k = m_h(x_{k-1}, x_k]$ , onde  $-\infty = x_0 < x_1 < \dots < x_n = \infty$ . Então o ganho médio de informação é máximo para  $p_k = 1/n$ .

*Demonstração.* Temos que maximizar a função  $I = - \sum p_k \log(p_k)$  sujeito à restrição  $\sum p_k = 1$ .

Podemos aplicar o método dos multiplicadores de Lagrange e procurar um ponto crítico de

$$L = - \sum p_k \log(p_k) + \lambda \sum p_k.$$

Derivando, obtemos

$$0 = \frac{\partial L}{\partial p_k} = - \log(p_k) - 1 + \lambda,$$

portanto  $p_k = e^{\lambda-1}$ , e usando a restrição, temos

$$p_k = \frac{1}{n}.$$

Como  $I$  é uma função côncava, sua restrição ao subespaço

$$\{(p_1, \dots, p_n) \mid \sum p_k = 1\}$$

também é côncava, portanto o ponto encontrado é o máximo global de  $I$ .

□

A proposição acima diz que se o processo de quantização maximiza o ganho médio de informação, os intervalos de quantização devem ser escolhidos de modo que se tenha uma *equalização* do histograma de freqüências da imagem. Se  $F(x) = m_h(-\infty, x]$  é a função de distribuição de  $h$ , então devemos tomar  $x_k = F^{-1}(k/n)$ . É fato bem conhecido da literatura de computação gráfica que um bom método de quantização, pela eficiência computacional e qualidade perceptual, é o algoritmo do corte mediano introduzido por Paul Heckbert em 1982. Essencialmente o que esse algoritmo faz é equalizar o histograma da imagem no processo de quantização. Pela proposição 9, o método maximiza a entropia da imagem quantizada o que, de certo modo, explica a boa qualidade perceptual obtida.



(a) 16 cores.



(b) 8 cores.



(c) 4 cores.



(d) 2 cores.

Figura 5.3:  $d(x, y) = |x - y|$



(a) 16 cores.



(b) 8 cores.



(c) 4 cores.



(d) 2 cores.

Figura 5.4:  $d(x, y) = |x - y|^2$



(a) 16 cores.



(b) 8 cores.



(c) 4 cores.



(d) 2 cores.

Figura 5.5:  $d(x, y) = |\log(x) - \log(y)|$



(a) 16 cores.



(b) 8 cores.



(c) 4 cores.



(d) 2 cores.

Figura 5.6:  $d(x, y) = |\log(x) - \log(y)|^2$



(a) 16 cores.



(b) 8 cores.



(c) 4 cores.



(d) 2 cores.

Figura 5.7: Mediana



## Capítulo 6

# Aplicações, extensões e trabalhos futuros

Neste trabalho, usamos conceitos de geometria diferencial para segmentar e reconstruir uma imagem a partir de suas características intrinsecamente bidimensionais. Apresentamos uma prova teórica de que é possível reconstruir uma superfície a partir de seus pontos de curvatura gaussiana não nula. Este resultado responde de maneira afirmativa à pergunta colocada em (Barth *et al.* , 1993; Zetsche *et al.* , 1993). Apresentamos, também, um algoritmo que reconstrói a imagem de maneira satisfatória a partir de seus pontos  $g_{1D}$  e  $g_{2D}$ .

Analisamos as arestas de uma imagem do ponto de vista da geometria diferencial e mostramos que tipicamente os pontos de arestas pertencem a regiões parabólicas das imagens. Isto mostra que podemos introduzir erros de reconstrução nas vizinhanças das arestas se não levarmos em conta as regiões  $g_{1D}$ . Tais erros são perceptualmente graves devido à sensibilidade do sistema visual a estas características das imagens.

A teoria apresentada neste trabalho tem grande potencial para aplicações. No momento, estamos investigando várias aplicações do método apresentado, bem como prosseguindo no desenvolvimento teórico dos seus desdobramentos. A seguir apresentamos uma lista com alguns tópicos de nosso interesse que estão sendo investigados.

**Algoritmo de reconstrução.** Para a maioria das aplicações é necessário um algoritmo de reconstrução eficiente. Portanto, apresentar algoritmos que reconstruam a imagem com perdas mínimas e rapidamente é essencial para o sucesso do método. Planejamos, também investigar melhor a importância dos pontos parabólicos para a reconstrução da imagem.

**Codificação e compressão de imagens.** Codificação e compressão de imagens aparece como aplicação natural dos métodos desenvolvidos ao longo da tese. Em particular pode-se usar filtros de curvatura para detectar as arestas da imagem e posterior codificação a partir destas arestas.

**Subdivisão adaptativa de imagens.** Pode-se explorar um filtro de curvaturas para produzir subdivisões adaptativas de uma imagem. Tal esquema para subdivisões traz consequências importantes como, por exemplo, a direção em que uma aresta cruza as células da subdivisão.

**Filtragem de ruídos em imagens.** A introdução de ruídos em uma imagem leva ao aparecimento de pontos de curvatura muito grande. Um filtro de curvaturas, por exemplo a gaussiana, pode ser usada para detectar e remover esses ruídos.

**Localização de imagens em bancos de dados.** Os pontos de curvatura máxima são determinantes no reconhecimento de uma imagem. Pode-se, então, usar estes pontos como uma maneira rápida de comparar imagens. Este método pode ser bastante eficiente visto que as curvaturas podem ser calculadas em tempo linear.

**Identificação de objetos em cena.** Filtros de curvatura podem ser usados para identificar objetos que são essencialmente geométricos, em uma cena. Neste caso, o objeto seria previamente selecionado, seus vértices calculados e o padrão obtido seria pesquisado nas imagens subseqüentes.

**Amostragem baseada na curvatura média de superfícies paramétricas e implícitas.** Métodos adaptativos de amostragem de uma curva ou função através da sua curvatura têm levado a excelentes resultados. As técnicas aqui apresentadas levam naturalmente a métodos de decomposição adaptativa de superfícies implícitas ou paramétricas. Além disso a geometria dos polígonos da decomposição variam de acordo com o tipo da componente da superfície (planar, parabólica, elíptica ou hiperbólica).

**Generalizações.** A teoria descrita na tese se estende naturalmente para tratar superfícies arbitrárias e não apenas superfícies de Monge. De um modo geral, uma imagem é uma função  $h : S \rightarrow \mathbb{R}^n$ , onde  $S$  é uma variedade. Casos particulares de grande interesse são:

- $\dim S = 3$ . Imagens 3D;
- $\dim S = 2$  e  $n \geq 3$ . Imagens em cores. Neste caso, podemos tratar a imagem como uma superfície de  $S \times \mathbb{R}^3$  ( $= \mathbb{R}^5$ , se  $S \subset \mathbb{R}^2$ ) ou como uma imersão  $h : S \rightarrow \mathbb{R}^3$ , neste último caso  $h$  pode ter muitos pontos singulares.
- $S$  é uma esfera. Este é um caso bastante natural, tendo em vista a distribuição das células fotossensíveis no globo ocular.

# Referências Bibliográficas

- Attneave, F. 1954. Some informational Aspects of Visual Perception. *Psychological Review*, **61**, 183–193.
- Barth, E., Caelli, T., & Zetsche, C. 1993. Image Encoding, Labeling, and Reconstruction from Differential Geometry. *CVGIP: Graphical Models and Image Processing*, **55**(6), 428–446.
- Barth, E., Beard, B. L., & Jr., A. L. Ahumada. 1999. Nonlinear features in vernier acuity. *Human Vision and Eletronic Imaging IV, SPIE Proceedings*, **8**, 3644.
- DeValois, R. L., & DeValois, K. K. 1980. Spatial Vision. *Annual Reviews in Psychology*, **31**, 309–314.
- DeValois, R. L., Albrecht, L. G., & Thorell, L. G. 1982. Spatial Frequency Selectivity of cells in Macaque Visual Cortex. *Vision Research*, **22**, 545–559.
- do Carmo, M. P. 1976. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Inc. ISBN 0-13-212589-7.
- Fechner, G. T. 1858. Über ein wichtiges psychophysiches Grundgesetz und dessen Beziehung zur Schätzung der Sterngrössen. *Abk. k. Ges. Wissensch., Math.-Phys. K1*, **4**.
- Fernandez, P. J. 1976. *Medida e integração*. Instituto de Matemática Pura e Aplicada, CNPq. Projeto Euclides.
- Froment, J., & Mallat, S. 1992. *Second Generation Compact Image Coding with Wavelets*. ISBN 0-12-174590-2.
- Gomes, J., & Velho, L. 1994. *Computação Gráfica: Imagem*. Instituto de Matemática Pura e Aplicada/SBM. Série de Computação e Matemática.
- Gomes, J., & Velho, L. 1995. Abstraction Paradigms for Computer Graphics. *The Visual Computer*, 227–239.
- Heckbert, P. 1982. Color image quantization for frame buffer display. *Computer Graphics*, **16**, 297–307.
- Hoffman, R., & Jain, A. 1987. Segmentation and Classification of Range Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-9**(5), 608–620.

- Hummel, R., & Moniot, R. 1989. Reconstructions from Zero Crossings in Scale Space. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **12**(37), 2111–1130.
- James, B. R. 1981. *Probabilidade: Um Curso em Nível Intermediário*. Instituto de Matemática Pura e Aplicada. Projeto Euclides.
- Keith, K. 1977. *Elementary Algebraic Geometry*. Springer-Verlag, New York Inc. ISBN 0-387-90199-X.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. 1959. What the Frog's Eye Tells the Frog's Brain. *Proceedings of the IRE*, **47**, 1940–1951.
- Levine, M. D. 1985. *Vision in Man and Machine*. McGraw-Hill Publishing Company. ISBN 0-07-037446-5.
- Lima, E. L. 1981. *Curso de Análise Volume 2*. Instituto de Matemática Pura e Aplicada, CNPq. Projeto Euclides.
- Lindeberg, T. 1994. *Space-Scale Theory in Computer Vision*. Kluwer Academic Publishers. ISBN 0-7923-9418-6.
- Logan Jr., B. F. 1977. Information in the Zero-crossings of Bandpass Signals. *Bell Syst. J.*
- Ma, S. D., & Li, B. 1998. Derivative computation by multiscale filters. *Image and Vision Computing*, **16**, 43–53.
- Mallat, S. 1991. Zero-crossings of a Wavelet Transform. *IEEE Trans. Inform. Theory*, July, 1019–1033.
- Mallat, S., & Zhong, S. 1992. Characterization of Signals from Multiscale Edges. *IEEE Trans. on Pattern Anal. Mach. Intell.*, **14**(7), 710–732.
- Marr, D. 1982. *Vision*. W. H. Freeman and Company. ISBN 0-7167-1567-8.
- Marr, D., & Hildreth, E. 1980. Theory of Edge Detection. *Proc. R. Soc. London*, **202**, 181–207.
- Meyer, Y. 1991. Un contre-exemple a la conjecture de Marr et a celle de S. Mallat. *Preprint*.
- Osserman, R. 1968. *Two-Dimensional Calculus*. Harcourt, Brace & World, Inc.
- Pitas, I., & Venetsanopoulos, A. N. 1990. *Nonlinear Digital Filters Principles and Applications*. Kluwer Academic Publishers. ISBN 0-7923-9049-0.
- Resnikoff, H. L. 1987. *The Illusion of Reality*. Springer-Verlag. ISBN 0-387-96398-7.

- Stevens, S. S. 1961. To honor Fechner and repeal his law. *Science*, **133**, 80–86.
- Stokely, E. M., & Wu, S. Y. 1992. Surface Parametrization and Curvature Measurement of Arbitrary 3-D Objects: Five Practical Methods. *IEEE Trans. on Pattern Anal. Mach. Intell.*, **14**(8), 833–840.
- von Helmholtz, H. 1891. Versuch einer erweiterten Anwendung des Fechnerschen Gesetzes im Fabensystem. *Z. Psychol. Physiol. Sinnesorg.*, **2**, 1–30.
- Walker, J. S. 1988. *Fourier Analysis*. Oxford University Press, Inc. ISBN 0-19-504300-6.
- Zayed, A. I. 1993. *Advances in Shannon's Sampling Theory*. CRC Press, Inc. ISBN 0-8493-4293-7.
- Zetsche, C., & Barth, E. 1990. Fundamental Limits of Linear Filters in the Visual Processing of Two-dimensional Signals. *Visual Res.*, **30**(7), 1111–1117.
- Zetsche, C., Barth, E., & Wegmann, B. 1993. The Importance of Intrinsically Two-dimensional Features in Biological Vision and Picture Coding. *Visual Factors in Eletronic Image Communication*, 109–138.