Teoria da Informação

INFORMASÃO

Conceito de Informação

- A informação se define em relação a uma mensagem.
- A mensagem pode ser qualquer coisa, em princípio, enviada por um emissor, com destino a um receptor
  - Por exemplo , uma imagem
- Ao se abstrair os aspectos físicos da mensagem, o que resta é a informação

Conceito de Informação

- Podemos considerar, em princípio, tanto o o emissor quanto o receptor como entidades abstratas
  - Supor que o emissor é o processo capaz de gerar uma imagem
  - Supor que o receptor é o processo capaz de ler uma imagem (carregá-la na memória, p.ex)

Conceito de Informação

- A informação é o "conteúdo" da mensagem
- Não o suporte físico da mesma
- Por exemplo, na frase:
  - A aula já começou !
    - **Seu suporte físico são as letras, as palavras, escritas e apresentadas**
    - **Seu conteúdo é aquilo que elas significam, a idéia de aula, o conceito de começar, o fato de ela já ter ou não começado e a exclamação no final.**
    - **A questão de se a aula já havia começado ou não quando a mensagem foi recebida, muda o conteúdo de informação da mensagem**

Conceito de Informação

- Portanto, <u>o estado do receptor</u> é significativo para se determinar as características da informação
  - Se a aula ainda não tivesse começado no entender dos alunos, a mensagem poderia ter o significado: “Vou começar agora”.
  - Se os alunos entendessem que a aula já havia começado, poderiam interpretar a mensagem como: “Vocês não estão prestando atenção”.

Conceito de Informação

- A informação permite alterar o estado de conhecimento do receptor sobre o emissor

Conceito de Informação

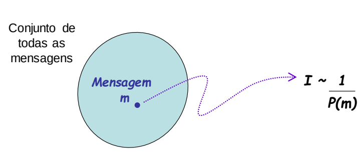
- Quanto <u>mais improvável</u> a mensagem, mais informação ela transporta

Medida da Informação

- Resumindo:
  - Quanto mais provável a mensagem (i.e., seu conteúdo), tanto menor a informação que ela traz
- Com base nesse conceito vamos “construir” uma medida para a informação
- Em seguida, vamos interpretar essa medida assim construída

Medida da informação

- A quantidade de informação em uma mensagem depende inversamente da probabilidade da mensagem
  - Em nosso caso a mensagem é a imagem



Medida da informação

- Quando se tem a conjunção de duas mensagens :
  - Do ponto de vista lógico booleano, é uma operação “e” (AND)
  - Do ponto de vista de teoria de conjuntos é uma intersecção
- Exemplo: se você envia a mesma mensagem duas vezes, deve ser entendida como uma só:

$$M \cdot M = M$$

Medida da informação

- Quando se tem a conjunção de duas mensagens :
  - Desejamos que a informação seja **aditiva** nesse caso, i.e., dadas duas mensagens concomitantes, queremos que suas informações se somem
- Porém, nesse caso suas probabilidades se **multiplicam**:

$$P(m1 \bullet m2) = P(m1).P(m2)$$

Conjunção (AND) Multiplicação

Medida da informação

- Estabelecemos que a medida da informação deve somar quando as probabilidades se multiplicam
  - A solução funcional para isso é empregar o logaritmo:

$$I \sim \frac{1}{P(m)} \;,\; \left.\begin{array}{l} I(m1 \bullet m2) = I(m1) + I(m2) \\ P(m1 \bullet m2) = P(m1).P(m2) \end{array}\right\} \quad I = \log\left(\frac{1}{P(m)}\right)$$

Conjunção (AND) Multiplicação

Medida da informação

$$I \sim \frac{1}{P(m)} \;,\; \left.\begin{array}{l} I(m1 \bullet m2) = I(m1) + I(m2) \\ P(m1 \bullet m2) = P(m1).P(m2) \end{array}\right\} \quad I = \log\left(\frac{1}{P(m)}\right)$$

Conjunção (AND) Multiplicação

ou $I = -\log P(m)$

portanto $I(m1.m2) = -\log P(m1.m2) = -\log [\, P(m1) \,.\, P(m2) \,]$

$= -\log P(m1) - \log P(m2) = I(m1) + I(m2)$

Medida da Informação

- Interpretemos agora a medida da informação que acabamos de construir seguindo um procedimento heurístico

$$I = -\log P(m)$$

- Primeiramente, vamos nos ater a um caso limite, em que as mensagens são simples e equiprováveis

Medida da Informação

- Suponhamos que todas as mensagens sejam equiprováveis. Vamos ordená-las em ordem lexicográfica e atribuir a cada uma delas um número.

Se tivermos ao todo 10 diferentes mensagens e usarmos 10 números para representá-las, quantos dígitos serão necessários ?

Medida da Informação

a símbolos (algarismos)
d posições (dígitos)

$$\underbrace{\boxed{a_5}\boxed{a_2}\boxed{a_6}\boxed{a_7}\boxed{a_2}}$$
$$d_{d-1} \;..\; d_2 \; d_1 \; d_0$$

Se cada número de **d** posições preenchidas com algarismos do conjunto {0,1,...,**a**-1} simbolizar uma mensagem, ao todo quantas **m** mensagens teremos ?

$$m = a^d$$

Medida da Informação

- O resultado significa que com ***a*** símbolos e ***d*** dígitos pode-se representar $m = a^d$ mensagens.
- Inversamente: sabendo-se que temos ***m*** mensagens e ***a*** símbolos, quantos dígitos ***d*** são necessários para representar essas mensagens ?
  - Resposta: $d = \log_a m$

Medida da Informação

- Exemplo:
  - Se **a = 2** algarismos temos: símbolos = {0,1} → conjunto binário (apenas 2 símbolos)
  - Quantos dígitos binários (i.e., bits) serão necessários para representar ***m*** mensagens ?

$$d = \log_2 m$$

- Portanto se usarmos log na base 2, a resposta será dada em bits.

Medida da Informação

- Suponhamos que as m mensagens sejam equiprováveis. Então, qual a probabilidade de cada mensagem m_i ?

$$P(m_i) = \frac{1}{m}$$

- Logo: $m = \dfrac{1}{P(m_i)}$

Medida da Informação

Portanto, uma mensagem $\mathbf{m_i}$ de um conjunto de **m** mensagens equiprováveis pode ser representada com **a** símbolos requerendo **d** dígitos:

$$m = \frac{1}{P(m_i)}\ ,\quad d = \log_a m \rightarrow d = \log_a \frac{1}{P(m_i)}$$

Então, a medida da informação da mensagem $\mathbf{m_i}$ de probabilidade $\mathbf{P(m_i)}$ pode ser interpretada como o <u>comprimento em dígitos requerido para representar</u> a mensagem com **a** símbolos

$$I_i = -\log_a P(m_i)$$

Entropia

- Define-se a **<u>entropia</u>** de um conjunto de símbolos (mensagem) como a **<u>informação média</u>** dos símbolos presentes na mensagem.
- Por ser informação média, a entropia mede a quantidade média de dígitos por símbolo, requeridos para representar a mensagem.

Entropia

- Para cada símbolo $\mathbf{m_i}$:

$$I_i = -\log_a P(m_i)$$

- Logo, para uma mensagem com **n** símbolos:

$$H = \langle I \rangle = -\sum_{i=0}^{n-1} \log_a P(m_i)\ .\ P(m_i)$$

Entropia de uma imagem

- Uma imagem pode ser considerada como uma mensagem composta de n = L x C símbolos, os pixels, sendo L a quantia de linhas e C a de colunas.
- Sendo a imagem digital, a quantidade média de bits requeridos para representar os pixels da imagem será dada por sua entropia:

$$H = - \sum_{i,j} \log_2 p(I_{ij}) \; . \; p(I_{ij})$$

Entropia de uma imagem

- As probabilidades dos valores dos pixels I_{ij} pode ser obtida do histograma da imagem.
- O histograma de frequências relativas h_v *($v = 0,\ldots,255$)* é o estimador da densidade de probabilidade dos pixels na imagem

$$H = - \sum_{i,j} \log_2 p(I_{ij}) \; . \; p(I_{ij})$$

ou

$$H = - \sum_{v} \log_2 (h_v) \; . \; h_v$$

Entropia de uma imagem

- Advertência:
  - A forma correta de se denominar a grandeza abaixo é: entropia dos tons monocromáticos da imagem, ou intensidades (*grey-level entropy*)

$$H = - \sum_{i,j} \log_2 p(I_{ij}) \; . \; p(I_{ij})$$

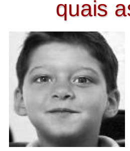
ou

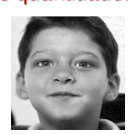
$$H = - \sum_{v} \log_2 (h_v) \; . \; h_v$$

    - **Para se definir uma entropia para as imagens em si, seria necessário conhecer mais sobre as correlações entre os pixels, sobre a estrutura espacial da imagem, sobre o processo estocástico que produz as imagens**

Exemplos

- Vamos calcular as entropias das intensidades de algumas imagens.
  - As imagens abaixo diferem na quantidade de bits usados para representá-las. Descobrir quais as quantidades de cada uma:

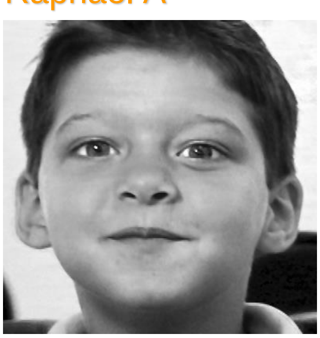

Raphael A


Raphael B


Raphael C


Raphael D

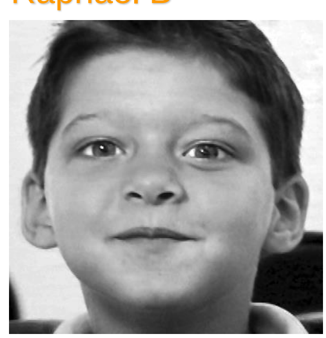
Raphael A



Entropia

6,888 bits

Raphael B



Entropia

5,835 bits

Raphael C

Entropia

7,782 bits

Raphael D

Entropia

4,839 bits

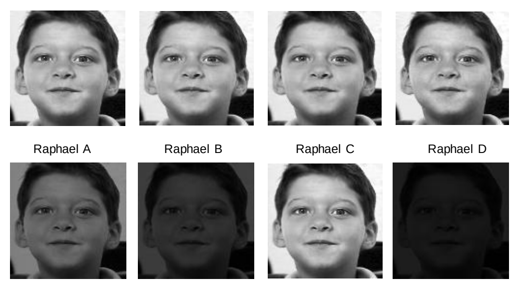
Observação

- As imagens mostradas não são de fato as que foram utilizadas nos cálculos. Elas foram equalizadas para poderem ser vistas:

Raphael A Raphael B Raphael C Raphael D

- Vejam como são as imagens reais:

Observação

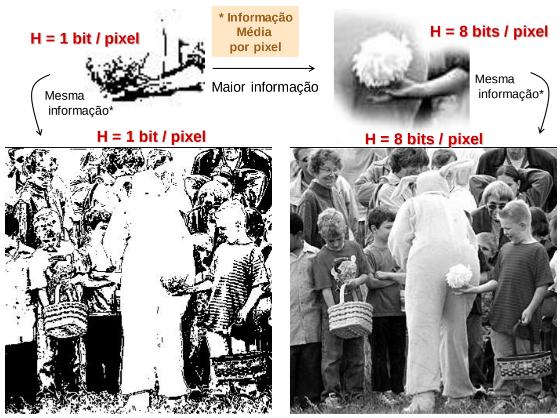


Teorema

- A função H tem valor máximo no, e unicamente no, ponto onde a probabilidade de cada evento é igual a 1/n
  - Dem: A demonstração, a ser feita na lousa, baseia-se na técnica dos multiplicadores de Lagrange.

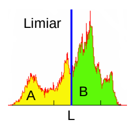
Exercício

- Explicar porquê as imagens ficaram mais escuras ao se diminuir a quantidade de bits usada para representá-las
- Construir uma transformação de LUT para produzir a partir de uma imagem de 8 bits as imagens com tons de cinza que correspondam a 7,6,5,4,3,2 e 1 bit.
- Obter os histogramas dessas imagens.
- Calcular as entropias dessas imagens.



Limiarização
(Thresholding)

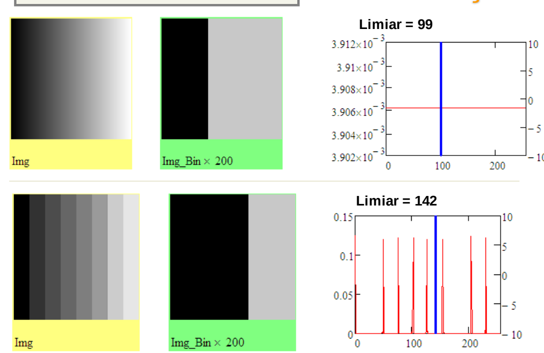
Limiarização



- Limiar (threshold) → valor limite L para tomar uma decisão
  - X ≤ L → decisão A
  - X > L → decisão B
- No caso de limiarização da intensidade, ou limiarização de níveis de cinza (*grey-level thresholding*):
  - X = intensidade ou nível de cinza
  - Decisão → classificar o nível X como pertencente à classe A ou B
  - A limiarização particiona o histograma em 2 distribuições que correspondem a duas *classes*
- A limiarização é um caso particular de classificação de regiões
  - Classificação binária com critério global baseado no histograma

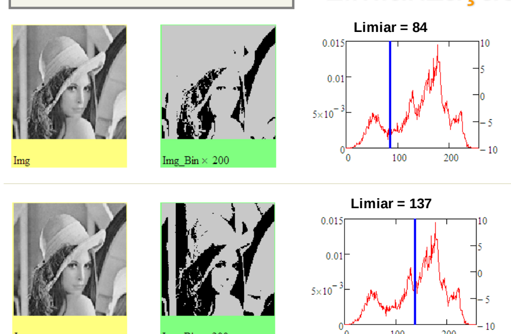
$$Img\_Bin_{i,j} := \text{if}\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização



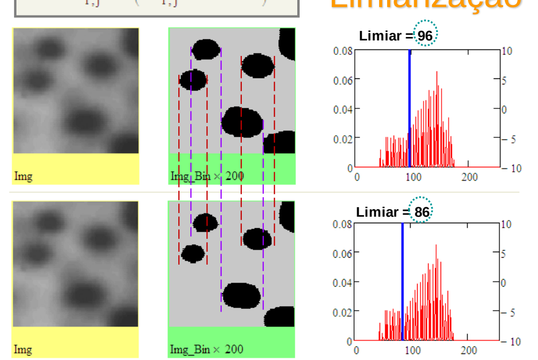
$$Img\_Bin_{i,j} := \text{if}\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização



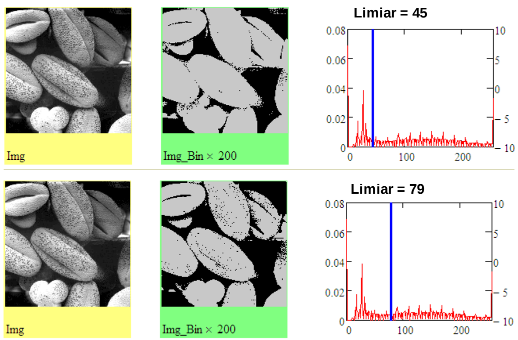
$$Img\_Bin_{i,j} := \text{if}\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização



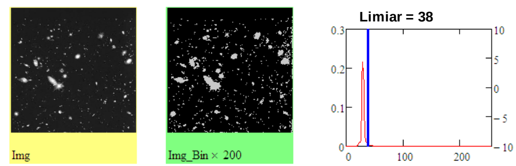
$$Img\_Bin_{i,j} := if\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização



$$Img\_Bin_{i,j} := if\left(Img_{i,j} > Limiar, 1, 0\right)$$

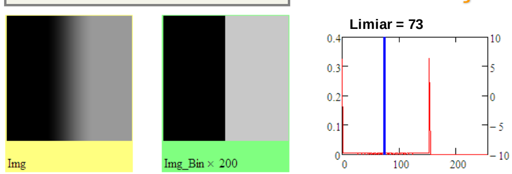
Limiarização



Empregando-se escala logaritmica para o histograma →

$$Img\_Bin_{i,j} := if\left(Img_{i,j} > Limiar, 1, 0\right)$$

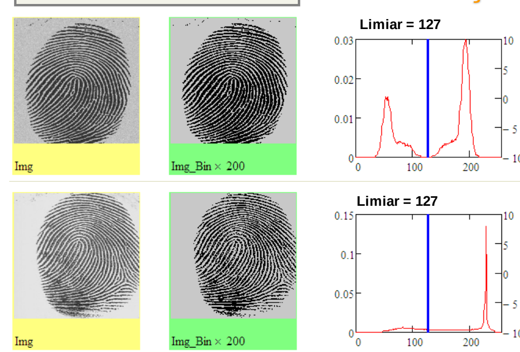
Limiarização



Empregando-se escala logaritmica para o histograma →

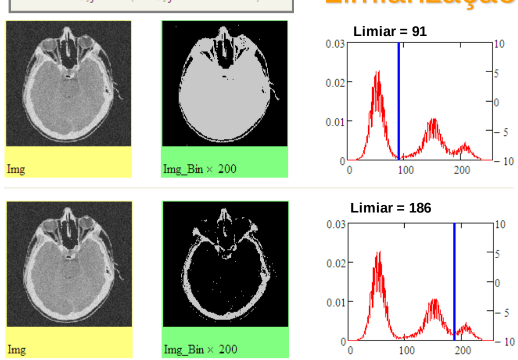
$$Img\_Bin_{i,j} := if\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização



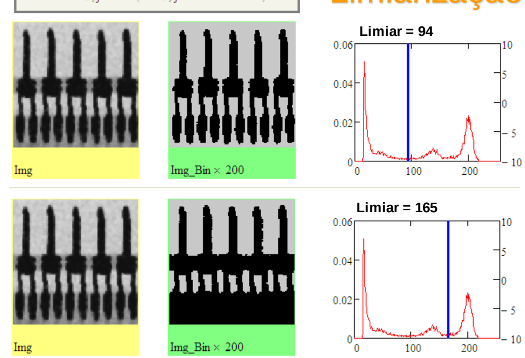
$$Img\_Bin_{i,j} := if\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização



$$Img\_Bin_{i,j} := if\left(Img_{i,j} > Limiar, 1, 0\right)$$

Limiarização

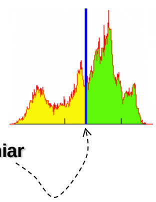


Limiarização automática

- Métodos para escolha de um limiar
  - – Baseados no histograma (momentos)
  - – Baseados na entropia
  - – outros

Estabelecer um **critério** para

Seleção automática do limiar



Limiarização – método de Otsu (1978)

- Também chamado de limiarização por maximização da variância inter-classes
- Baseado no histograma
- Supõe a mistura de duas distribuições formando o histograma total
- O objetivo é encontrar o limiar que separa essas distribuições
- A técnica usada é análise discriminante linear

Limiarização – método de Otsu (1978)

$\omega_A(L) := \sum_{v=0}^{L} h_v$ probabilidade da classe A

$\omega_B(L) := 1 - \omega_A(L)$ probabilidade da classe B

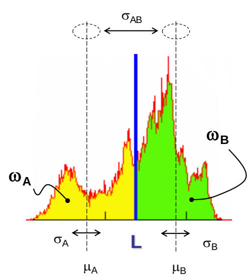
$\mu_L(L) := \sum_{v=0}^{L} (v \cdot h_v)$ média até o limiar L

$\mu_T := \frac{1}{N}\left[\sum_{v=0}^{255} (v \cdot H_v)\right]$ média da imagem

$\mu_A(L) := \frac{\mu_T}{\omega_A(L)}$ média da classe A

$\mu_B(L) := \frac{\mu_T - \mu_L(L)}{\omega_B(L)}$ média da classe B

$\sigma_{AB}^2 = \omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2$ variância inter-classes



Limiarização – método de Otsu (1978)

variância inter-classes

$$\sigma_{AB}^2 = \omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2$$

variância total

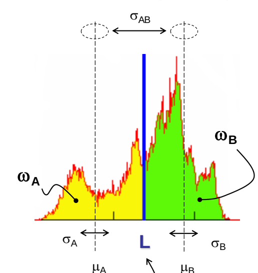
$$\sigma_T^2 = \sum_{v=0}^{255} \left[(v - \mu_T)^2 \cdot h_v\right]$$

função critério a ser maximizada

$$D = \frac{\sigma_{AB}^2}{\sigma_T^2}$$

limiar ótimo

$$L_{opt} = \underset{0 < L < 255}{\mathrm{ArgMax}}(D)$$



Exemplo

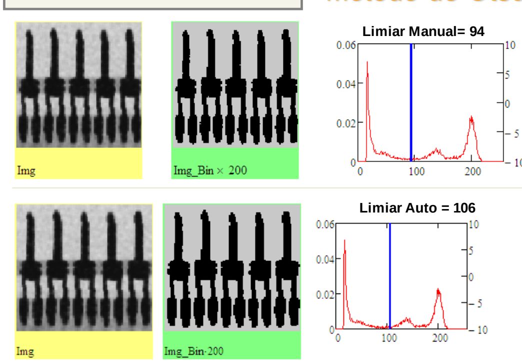
- Suponha o histograma
  - – [8,7,2,6,9,4]

0	0	1	4	4	5
0	1	3	4	3	4
1	3	4	2	1	3
4	4	3	1	0	0
5	4	2	1	0	0
5	5	4	3	1	0

http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html

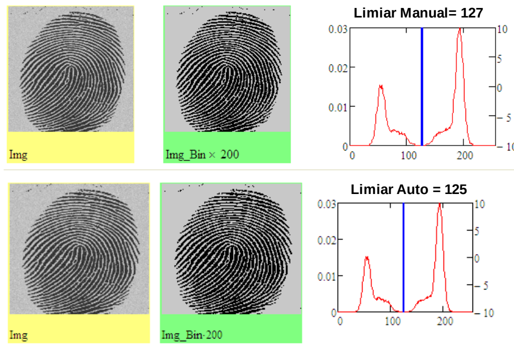
Método de Otsu

$$L_{opt} = \mathrm{ArgMax}\left[\omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2\right]$$



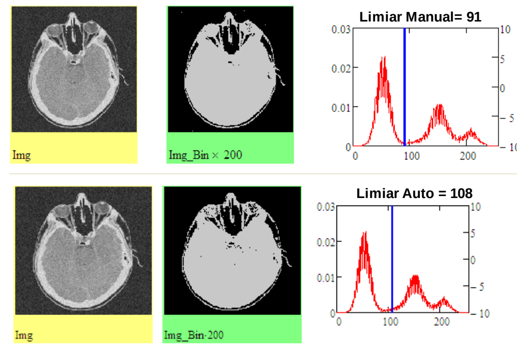
$$L_{opt} \equiv \text{ArgMax}\left[\omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2\right]$$

Método de Otsu



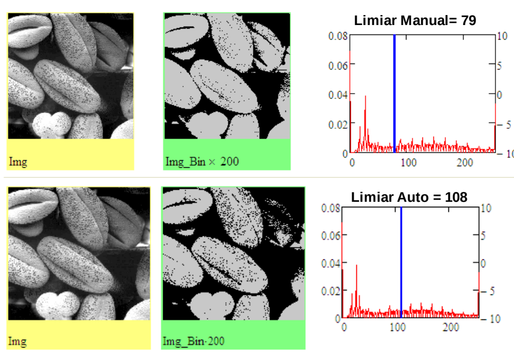
$$L_{opt} \equiv \text{ArgMax}\left[\omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2\right]$$

Método de Otsu



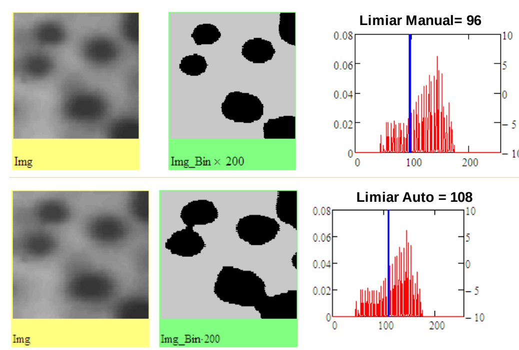
$$L_{opt} \equiv \text{ArgMax}\left[\omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2\right]$$

Método de Otsu



$$L_{opt} \equiv \text{ArgMax}\left[\omega_A \cdot (\mu_A - \mu_T)^2 + \omega_B \cdot (\mu_B - \mu_T)^2\right]$$

Método de Otsu



Limiarização – método de Pun(1980)

- Baseado nas entropias a posteriori
  - Calcula-se as entropias a posteriori das classes separadas com um limiar arbitrário
  - Determina-se o limiar que maximiza a entropia total a posteriori
- Exercício: (para casa)
  - Compare (em palavras) a idéia deste método com a do método de Otsu

Limiarização – método de Pun (1980)

entropia da classe A

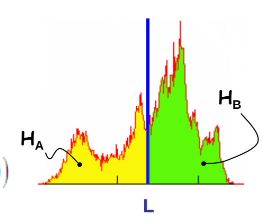
$$H_A(L) := -\sum_{v=0}^{L} \left(h_v \cdot \ln(h_v)\right)$$

entropia da classe B

$$H_B(L) := -\sum_{v=L+1}^{255} \left(h_v \cdot \ln(h_v)\right)$$

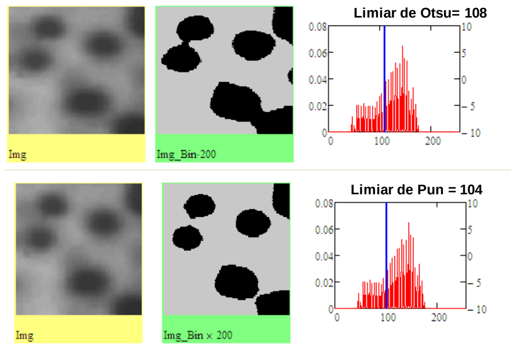
entropia total a posteriori

$$H_T(L) := H_A(L) + H_B(L)$$



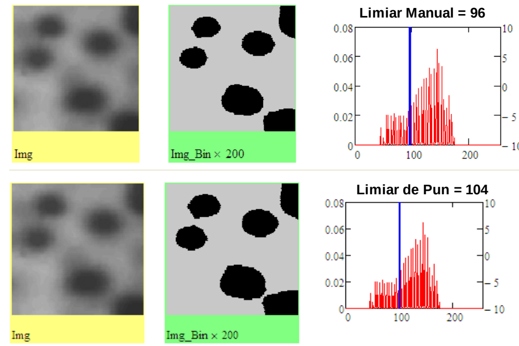
$H_T(L) := H_A(L) + H_B(L)$

Método de Pun



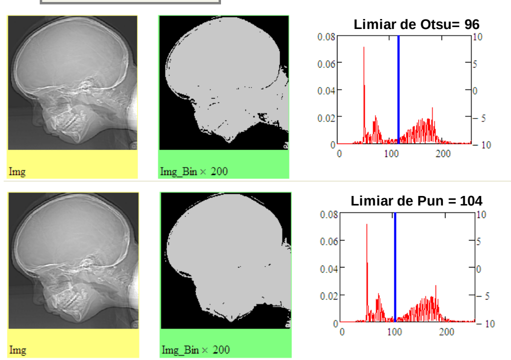
$H_T(L) := H_A(L) + H_B(L)$

Método de Pun



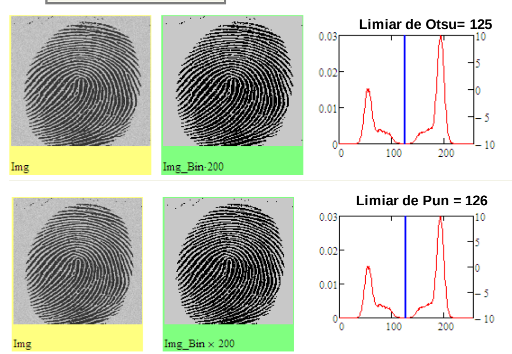
$H_T(L) := H_A(L) + H_B(L)$

Método de Pun



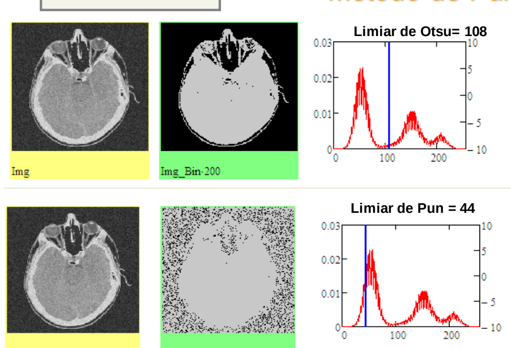
$H_T(L) := H_A(L) + H_B(L)$

Método de Pun



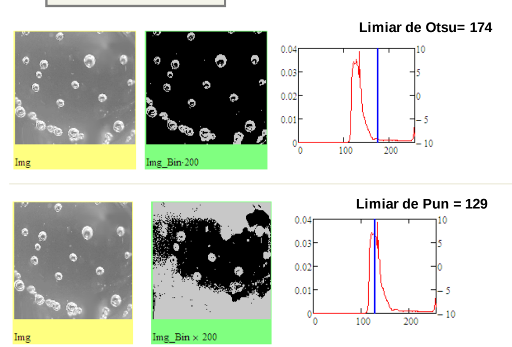
$H_T(L) := H_A(L) + H_B(L)$

Método de Pun



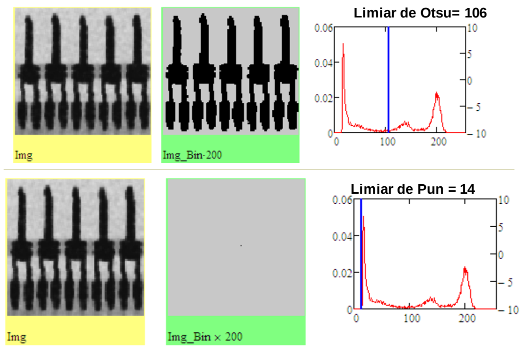
$H_T(L) := H_A(L) + H_B(L)$

Método de Pun



$$H_T(L) := H_A(L) + H_B(L)$$

Método de Pun



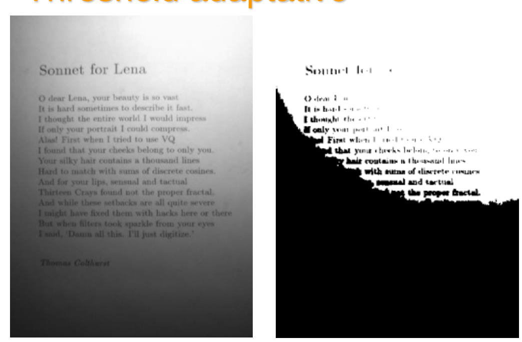
Comparação

- O método baseado nas entropias a posteriori parece ter desempenho menor que o método baseado na variância inter-classes
- Porém ele é superior em alguns casos e empata em outros
- Os métodos de segmentação baseados em entropia estão entre os melhores. Entretanto, o método de Pun parece ser fraco
- A razão está no fato de ele não levar em consideração a correlação entre os pixels
- Como considerar isso ?
  - **Requer conhecimento**
    - da estrutura topológica da imagem
    - da regularidade da imagem

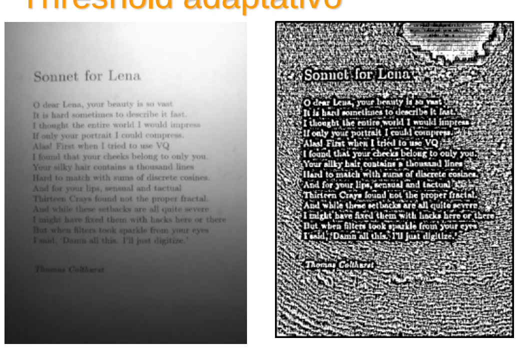
Threshold adaptativo

- O método é baseado nas estatísticas locais da vizinhança de um ponto.
- Por exemplo, toma-se a média, ou a mediana locais
- Outra forma é tomar a média, ou a mediana locais, subtraídas de um certo valor.

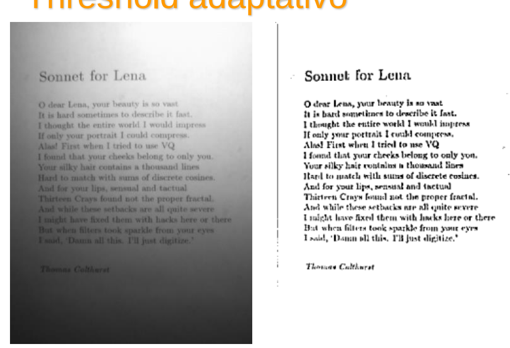
Threshold adaptativo



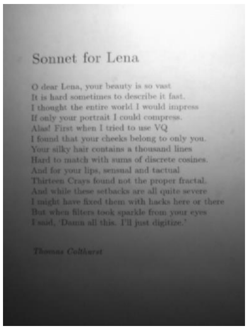
Threshold adaptativo

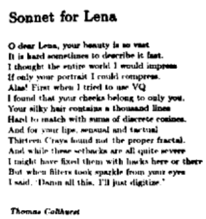


Threshold adaptativo



Threshold adaptativo





Threshold adaptativo