

Processamento Digital de Imagens

Percepção de Imagens

Eduardo A. B. da Silva

Programa de Engenharia Elétrica - COPPE/UFRJ

Laboratório de Sinais, Multimídia e Telecomunicações

eduardo@smt.ufrj.br

Sergio L. Netto

Programa de Engenharia Elétrica - COPPE/UFRJ

Laboratório de Sinais, Multimídia e Telecomunicações

sergioln@smt.ufrj.br

Abril de 2017



Sumário

① Percepção de Imagens

- Olho Humano
- Formação Fotométrica de Imagens
- Um Pouco de Ótica
- Resposta em Freqüência
- Luminância × Brilho
- Bandas de Mach
- Mascaramento
- Quantização
- Modelo para Visão Monocromática

② Avaliação de Qualidade

- Critérios de Fidelidade
- Critérios Subjetivos
- Critérios Objetivos

- Medidas Baseadas no Erro Médio Quadrático
- Medidas Baseadas em Similaridade Estrutural

③ Avaliação de Qualidade de Vídeo

- VQEG - Video Quality Experts Group
- ITU-T J.144

④ Cor

- Cor
- Casamento e Reprodução de Cores
- Leis do Casamento de Cores
- Diagrama De Cromaticidade
- Sistemas de Coordenadas de Cores
- Medidas de Diferença de Cor
- Propriedades Temporais da Visão

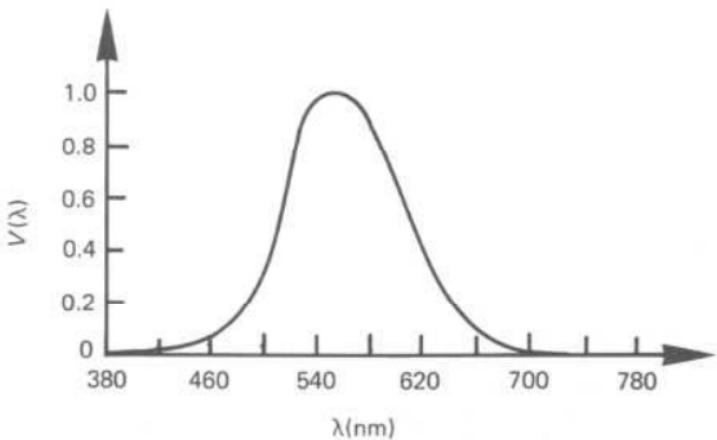


Perepção de Imagens

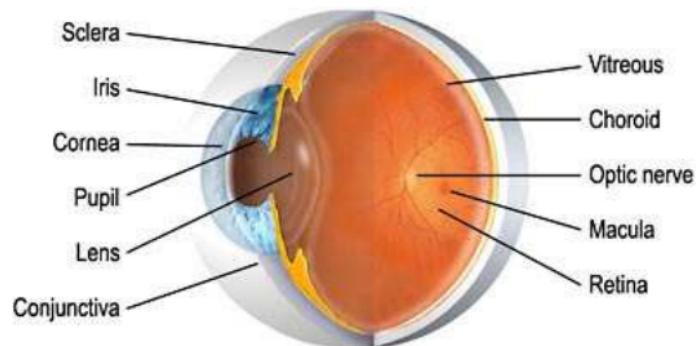
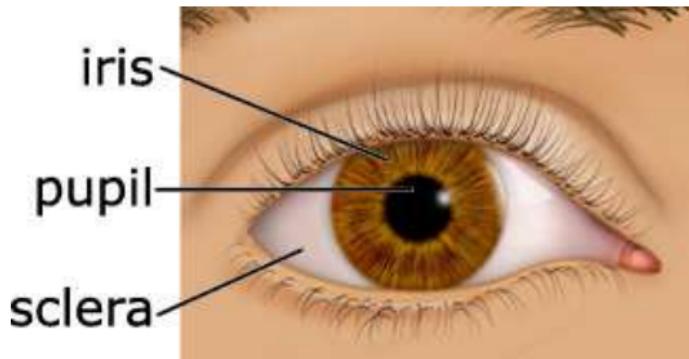
- **Luz:** Densidade espectral de energia $L(\lambda)$, $350\text{nm} \leq \lambda \leq 780\text{nm}$.
- **Luz recebida de um objeto:** $I(\lambda) = \rho(\lambda)L(\lambda)$.
- **Faixa dinâmica do sistema visual:** 10 ordens de grandeza (10^{10}).
- Estamos interessados na visão fotópica (displays eletrônicos são sempre bem iluminados).
- **Luminância:** Proporcional à energia recebida pelo olho. Para um objeto com distribuição de luz $I(x, y, \lambda)$:

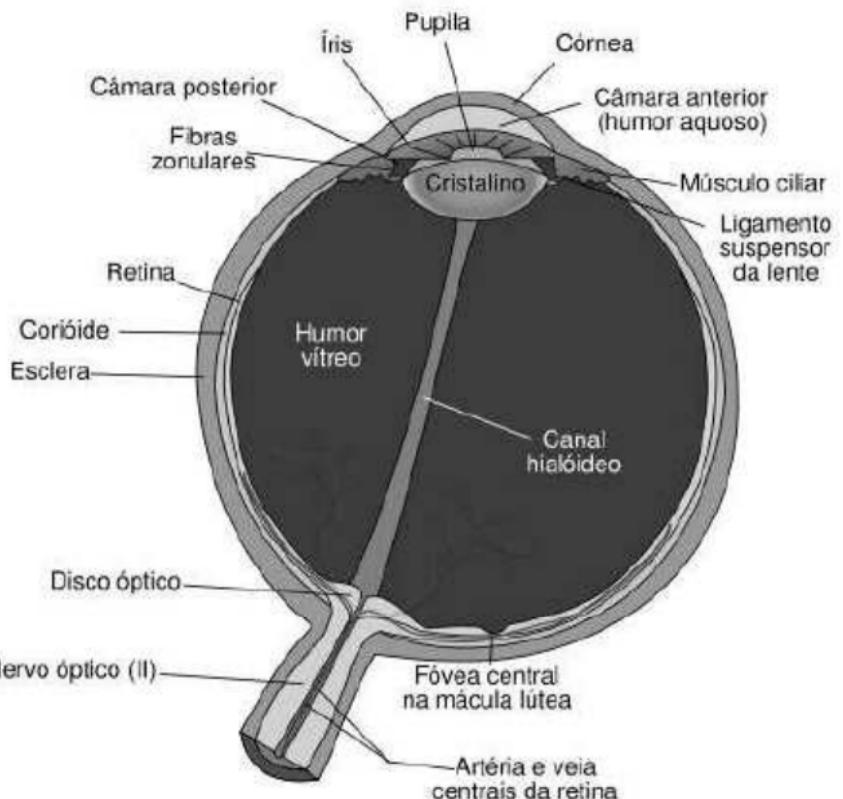
$$f(x, y) = \int_0^{\infty} I(x, y, \lambda) v(\lambda) d\lambda$$

$v(\lambda)$ é a eficiência luminosa relativa do sistema visual (diferente para visão fotópica e escotópica)

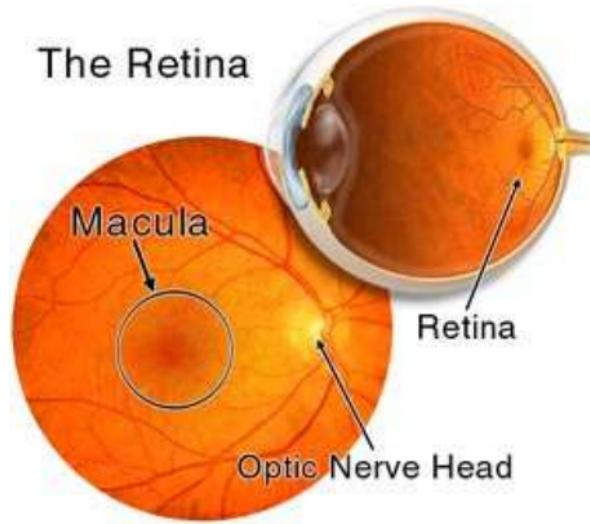


Olho Humano



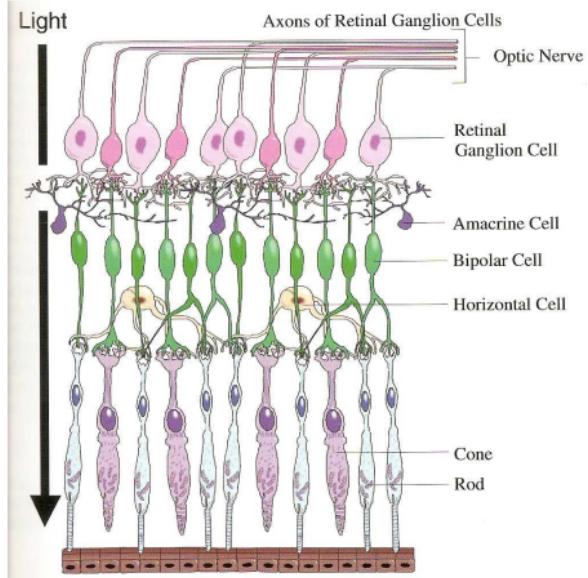


- **Retina:** rede de sensores óticos que transformam a informação luminosa em algo que possa ser transmitido para o cérebro através dos neurônios. Possui 2 tipos de sensores:



- **Cones:** \approx 6.5 milhões – são os responsáveis pela visão das cores. Visão fotópica (altos níveis de iluminação). Bem densos na fóvea ($\simeq 120$ cones/grau $\simeq 0.5'$ ou $2\mu m$). Rara densidade depois de 1° .
- **Bastonetes:** \approx 100 milhões – só distinguem detalhes em “preto e branco”. Funcionam a partir de baixos níveis de iluminação.

Bastonetes

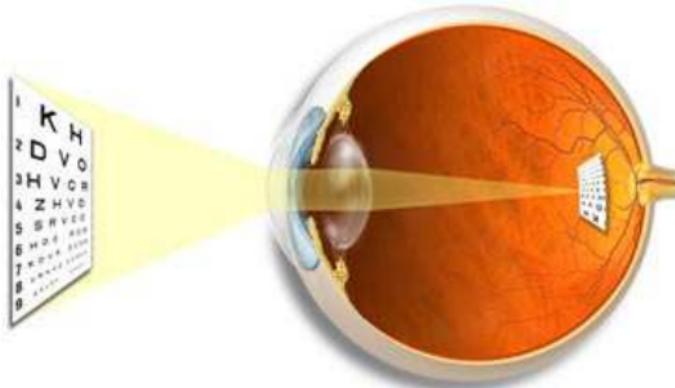


- Sensores usados para visão escotópica (noturna);
- Percebem apenas brilho (não cor);
- Resolução média para baixa, pois múltiplos bastonetes se conectam a um mesmo nervo;
- Em torno de 100 milhões.

Cones

- Sensores usados para visão fotópica (diária);
- Sensíveis à cor;
- Alta resolução pois cada cone se conecta a um nervo;
- Em torno de 10 milhões;
- A Fovea é a região de maior concentração de cones, em torno de 300,000 num círculo de diâmetro 1.5 mm;
- Três tipos: Vermelho (RED), Verde (GREEN) e Azul (BLUE).

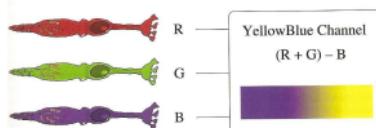
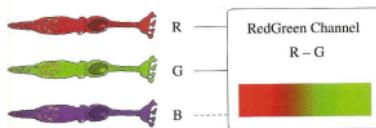
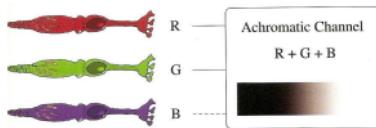
- **Córnea, pupila, íris e cristalino** – “focalizam” a imagem na retina.
- **Pupila:** $\simeq 2\text{mm}$ de diâmetro, atua como filtro passa-baixas, com corte em 60 ciclos/grau.



- **Brilho:** Intensidade luminosa (luminância) “percebida” pelo sistema de visão humano.
- O brilho depende não somente da luminância de um objeto, mas da luminância à volta do objeto.

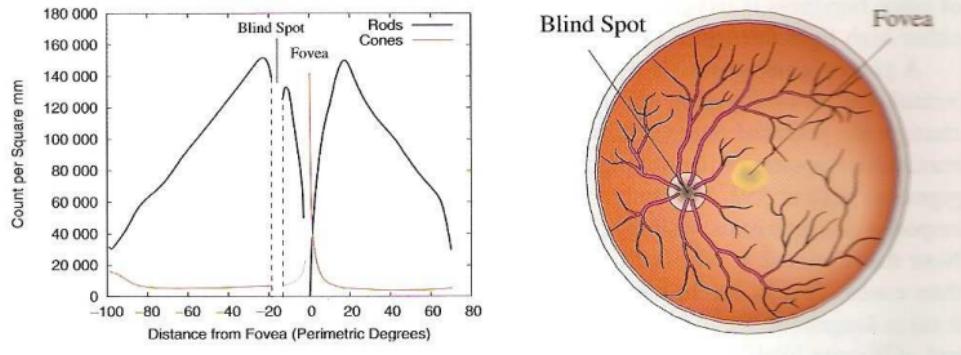
Canais de Cores

- Três informações passam para o cérebro.



- As linhas pontilhadas representam “controvérsias” quanto à influência dessa cor “original” (RGB) no canal.

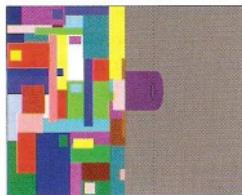
O Ponto Cego



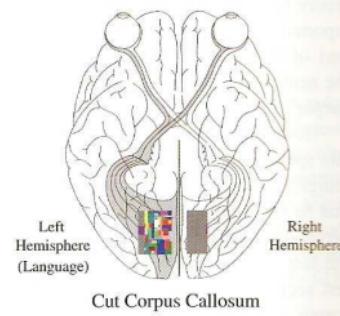
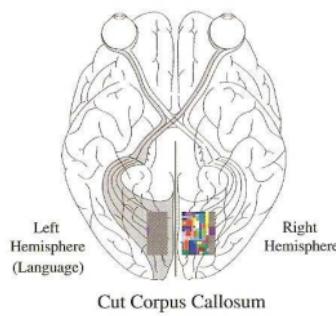
Olhando essa figura responda ainda: por que é mais fácil ver uma estrela à noite olhando “um pouco para o lado” de onde ela está?

Formação de Imagens no Cérebro

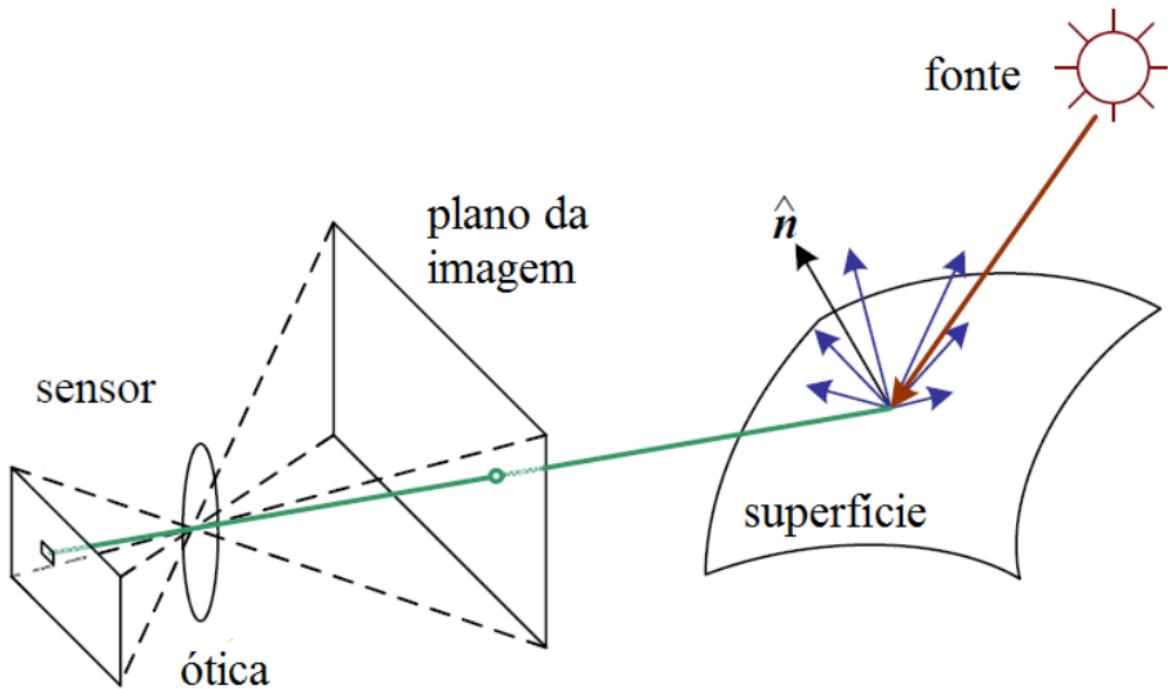
Direct View



Mirror Image



Formação Fotométrica de Imagens



Fontes de Luz

- Fontes Pontuais:

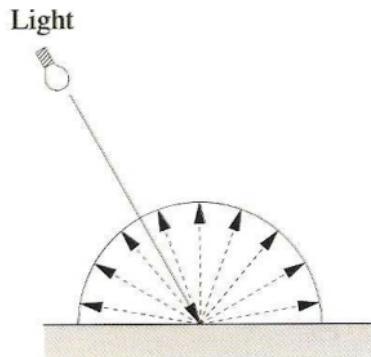
- Localização única no espaço (ex: lâmpada), podendo estar no infinito (ex: sol);
- Caracterizadas pela sua intensidade e espectro de cor;

- Fontes de Área:

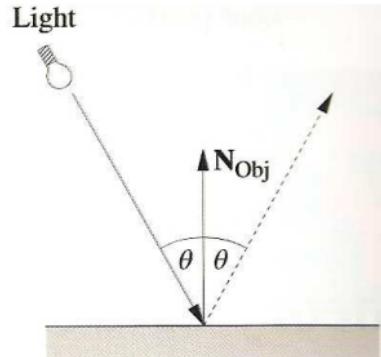
- Modelagem mais complexa. Em alguns casos pode ser modeladas como áreas retangulares finitas emitindo luz uniforme em todas as direções (ex: lâmpada fluorescente);

Luz e Superfícies

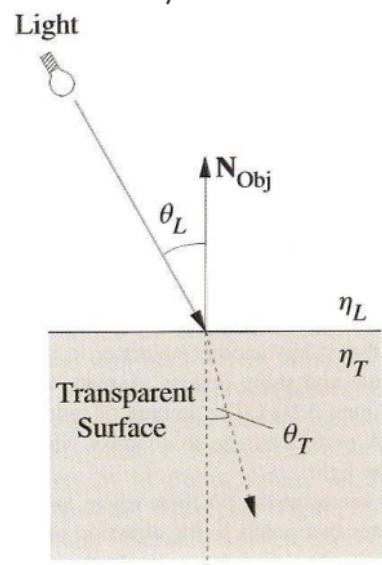
Difusão da Luz



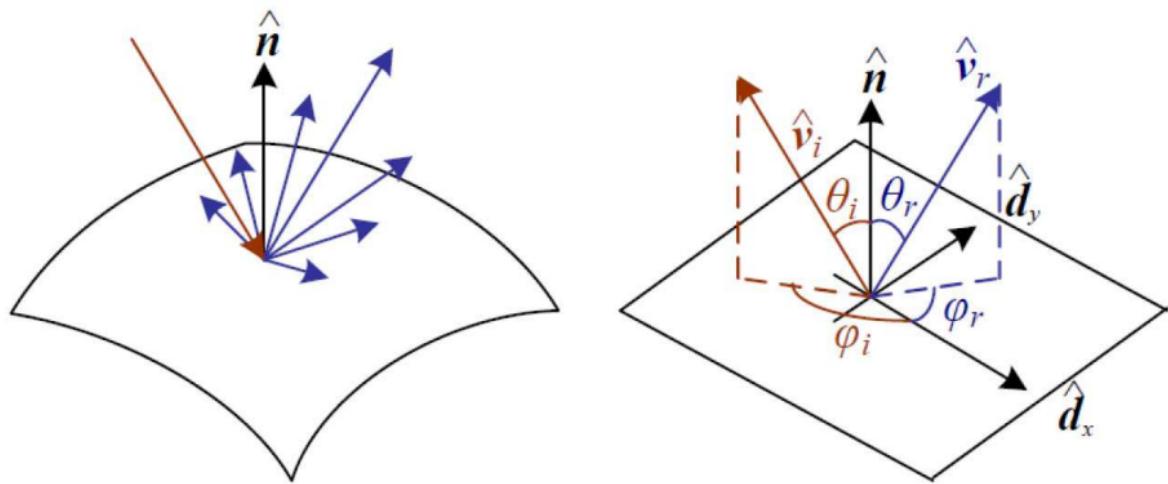
Reflexão da Luz



Refração da Luz



Reflectância e Sombreamento



⇒ BRDF: Bidirectional Reflectance Distribution Function

$$f_r(\theta_i, \varphi_i, \theta_r, \varphi_r; \lambda)$$

BRDF

- a BRDF é recíproca:

$$f_r(\theta_i, \varphi_i, \theta_r, \varphi_r; \lambda) = f_r(\theta_r, \varphi_r, \theta_i, \varphi_i; \lambda)$$

- Para uma superfície isotrópica (não há direções preferenciais), a BRDF pode ser simplificada para:

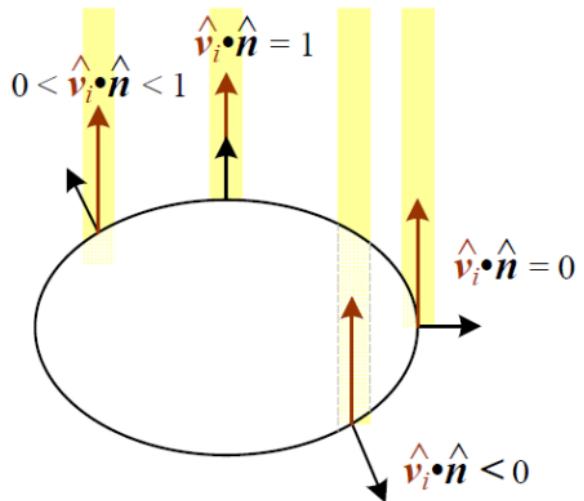
$$f_r(\theta_i, \theta_r, |\varphi_i - \varphi_r|; \lambda) \text{ ou } f_r(\hat{v}_i, \hat{v}_r, \hat{n}; \lambda)$$

- Para a luz incidente $L_i(\hat{v}_i; \lambda)$, podemos calcular a quantidade de luz emitida de um ponto p na direção \hat{v}_r da superfície por:

$$L_r(\hat{v}_r; \lambda) = \int L_i(\hat{v}_i; \lambda) f_r(\hat{v}_i, \hat{v}_r, \hat{n}; \lambda) \cos^+ \theta_i d\hat{v}_i$$

onde

$$\cos^+ \theta_i = \max(0, \cos \theta_i) \quad (\text{foreshortening factor})$$



Reflexão Difusa

- Também chamada de *Lambertiana* ou *mate*;
- A luz é espalhada uniformemente em todas as direções, ou seja:

$$f_d(\hat{v}_i, \hat{v}_r, \hat{n}; \lambda) = f_d(\lambda)$$

$$L_d(\hat{v}_r; \lambda) = \sum_i L_i(\lambda) f_d(\lambda) \cos^+ \theta_i = \sum_i L_i(\lambda) f_d(\lambda) [\hat{v}_i \cdot \hat{n}]^+$$

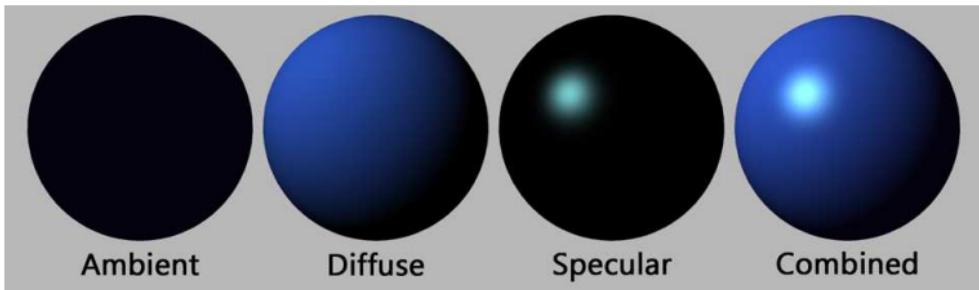
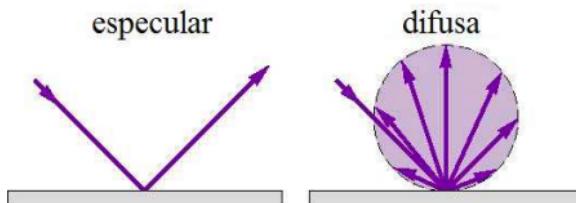
onde

$$[\hat{v}_i \cdot \hat{n}]^+ = \max(0, \hat{v}_i \cdot \hat{n})$$



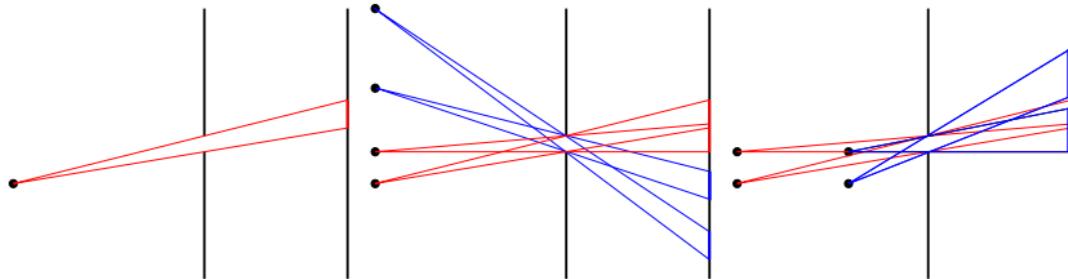
Reflexão Especular

- Depende fortemente da direção da luz de saída;



Um Pouco de Ótica

- O que ocorre quando a luz irradiada por um “ponto” luminoso passa por um orifício e “choca-se” num anteparo?

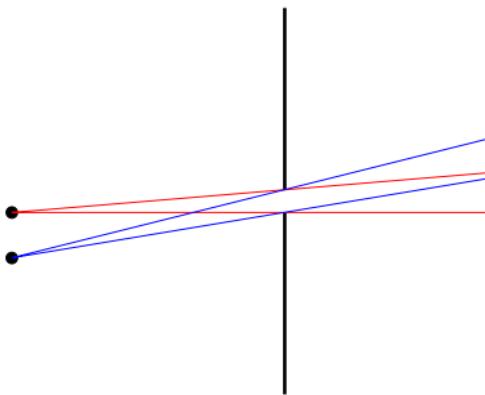


- Nossa pupila é um orifício como esse.

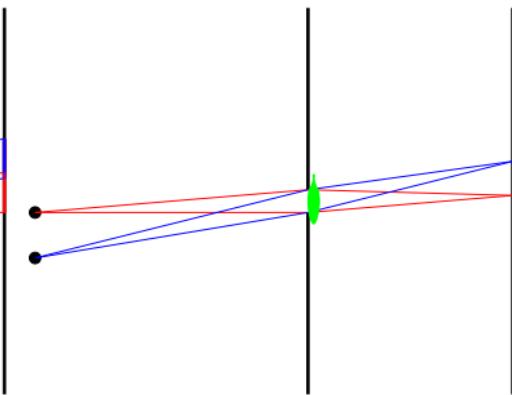
Lentes

- Baseiam-se na refração da luz.

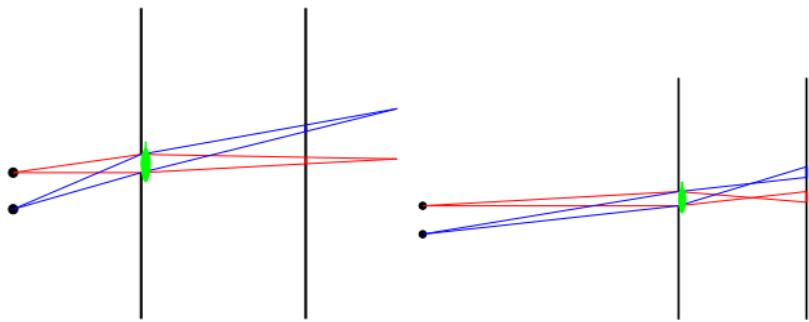
Sem lente



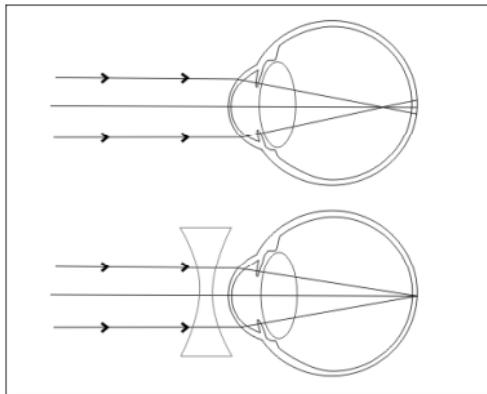
Com lente



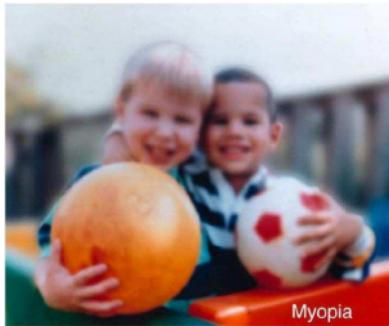
- Lentes desajustadas.



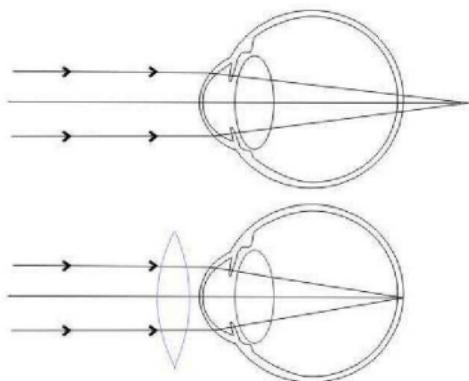
Miopia



- O que vê um míope?

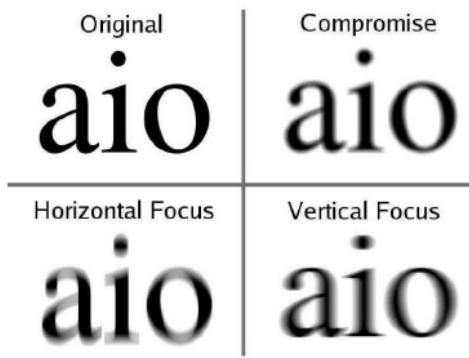


Hipermetropia



Astigmatismo

- Vários tipos.
- Um problema do cristalino e não da lente.



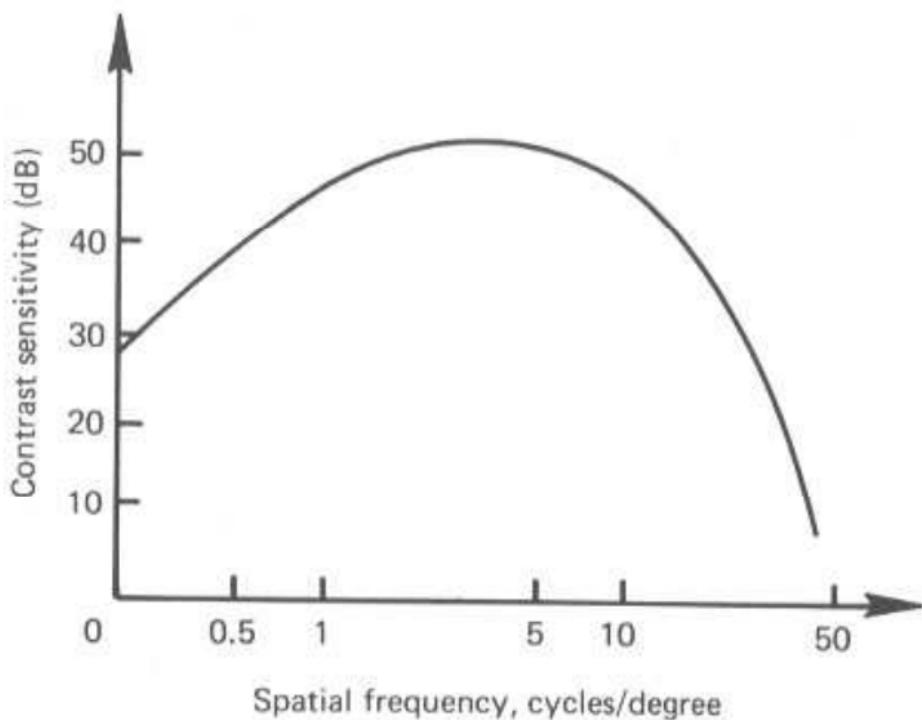
Resposta em Freqüência

- Tanto o número de sensores na retina quanto o sistema de focalização do olho humano limitam a resolução do sistema ótico humano.
- Em outras palavras, o olho humano irá ter respostas diferentes para diferentes faixas de freqüências espaciais.
- O olho humano equivale a um filtro passa-banda.
- O formato é resultado da influência combinada da ótica do olho humano, da retina e dos neurônios que transportam a informação para o cérebro e possivelmente do próprio cérebro.

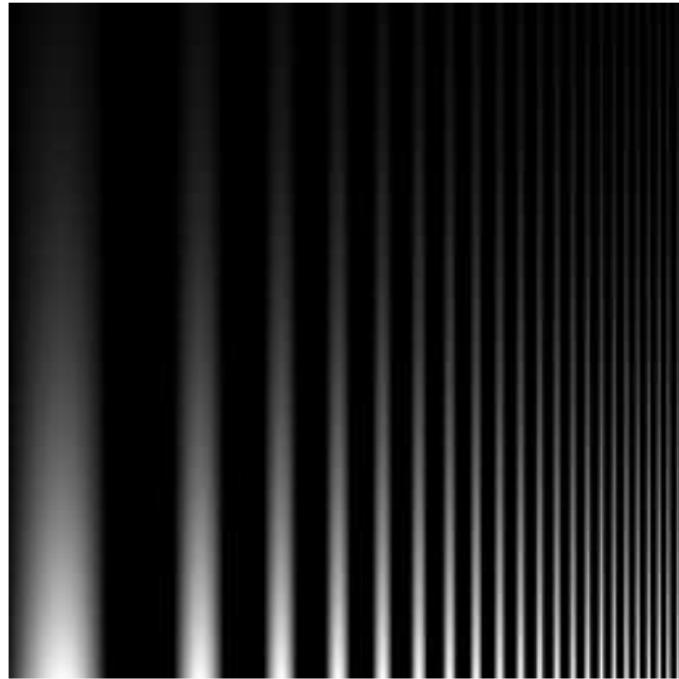
$$H(\omega_1, \omega_2) = H_\rho(\rho) = A \left[\alpha + \left(\frac{\rho}{\rho_0} \right) \right] e^{-\left(\frac{\rho}{\rho_0} \right)^\beta}, \quad \rho = \sqrt{\omega_1^2 + \omega_2^2}$$

Ex: $A = 2.6$, $\alpha = 0.0192$, $\rho_0 = (0.114)^{-1} = 8.772$, $\beta = 1.1$

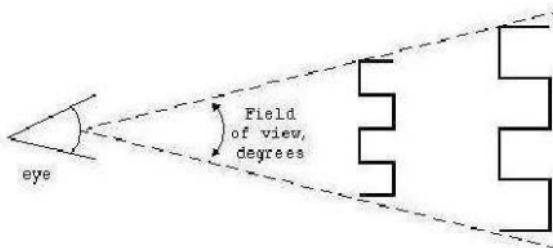




- Observe o padrão a seguir. O que você conclui?



Ciclos por Grau

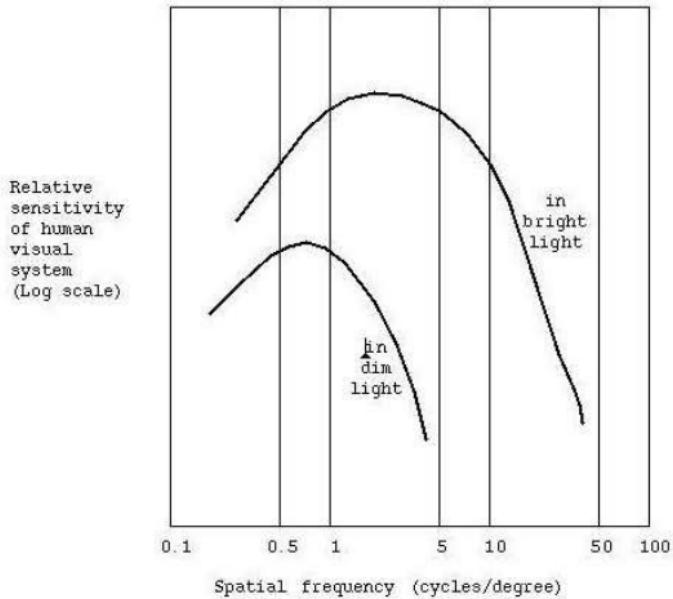


- a) With a fixed field of view of a given number of cycles, the farther from the eye, the larger each cycle must be.



- b) A larger, more distant object can appear to be the same size as a smaller, closer object.

Influência da Iluminação



⇒ Por que?

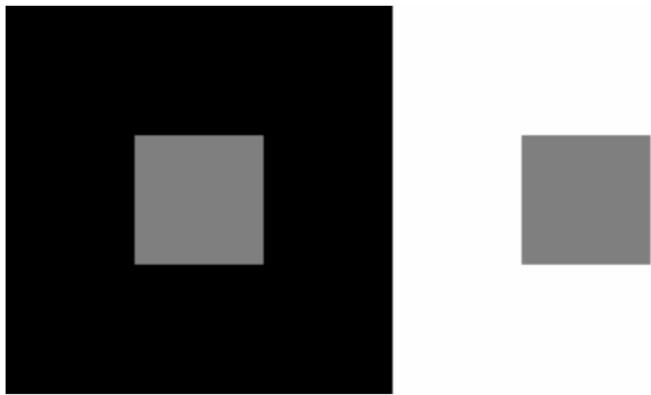


Luminância × Brilho

Relembrando:

- **Luminância:** Proporcional à energia recebida pelo olho.
- **Brilho:** Intensidade luminosa “percebida” pelo sistema de visão humano.
- O brilho depende não somente da luminância de um objeto, mas da luminância à volta do objeto.
- Observe os quadrados a seguir. O que se conclui?







Contraste Simultâneo



- O valor do incremento de luminância que está no limiar da percepção é tanto maior quanto maior for a luminância à volta (Lei de Weber).
- Correção gama:** “casamento” das características dos sistemas de aquisição e reprodução de imagens com as características do olho humano. Deve atingir 2 objetivos:
 - Um objeto “ao vivo” deve ter o mesmo “contraste” do próprio reproduzido numa tela, num papel ou outro meio.
 - Os valores dos pixels de uma imagem devem ser proporcionais à sua luminância.

Intensidade linear	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Com correção gama	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

- A correção gama é na prática aproximada por uma função do tipo $Y = X^\gamma$.

A relação do brilho com a luminância é bem mais complexa do que enuncia a lei de Weber.

Modelo:

$$\begin{cases} \text{ambiente: } f_s \\ \text{objeto: } f_o \end{cases}$$

f_o começa a ser notado como diferente de f_s

$$\frac{|f_s - f_o|}{f_o} = \text{cte} \quad \frac{\Delta f}{f} \simeq d(\log f) = \text{cte} \simeq 0.02$$

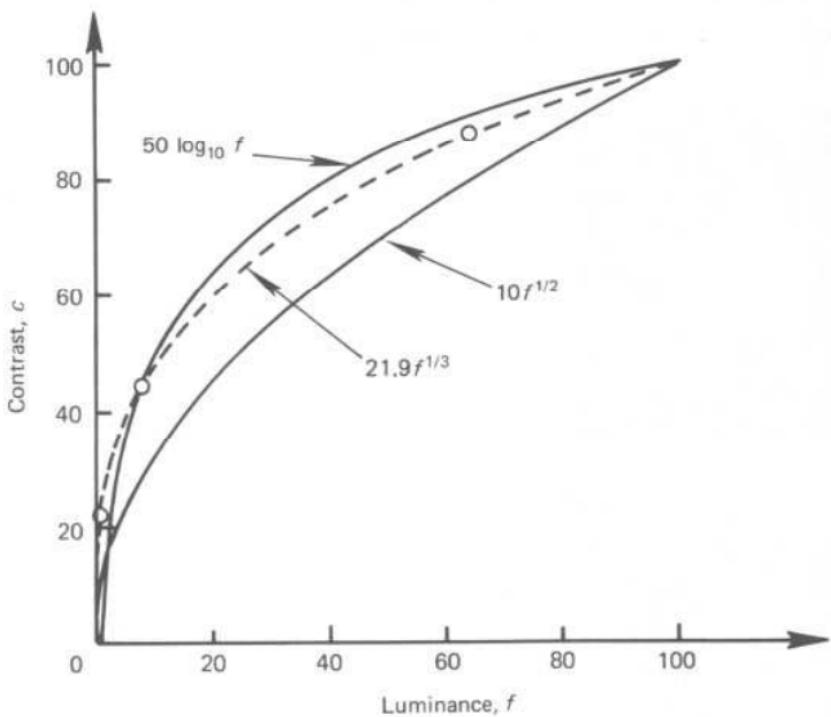
Incrementos iguais no $(\log f)$ equivalem a incrementos iguais no contraste

$$\frac{df}{f} = dc \quad \Rightarrow \quad c = a_1 + a_2 \log f$$

Outros modelos:

- ① $c = 50 \log f, \quad 1 \leq f \leq 100$
- ② $c = \alpha_n f^{\frac{1}{n}}, \quad n = 2, 3, \dots, \quad \alpha_2 = 10, \quad \alpha_3 = 21.9$
- ③ $c = \frac{f(f_B+100)}{f_B+f}, \quad f_B : \text{luminância do background}$



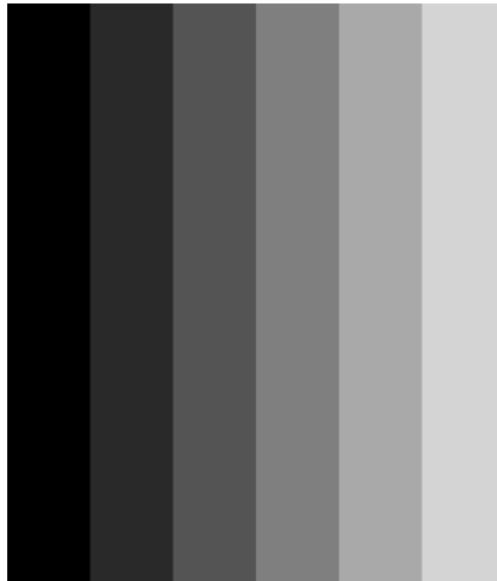


Vejamos alguns exemplos.

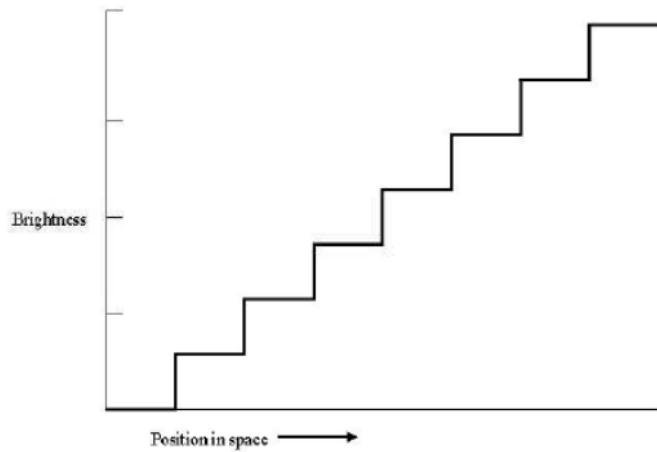


Bandas de Mach

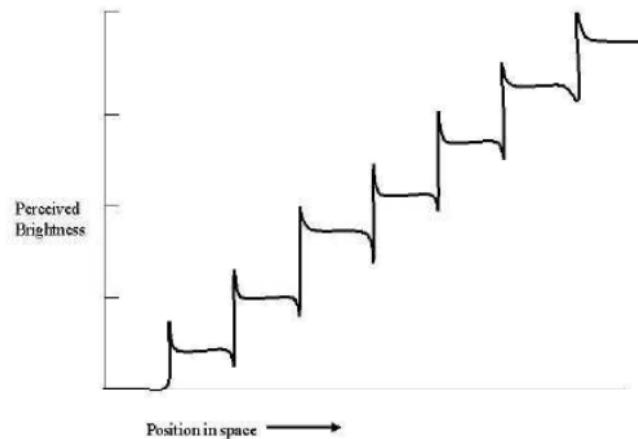
- Seja a imagem a seguir, onde cada faixa possui luminância constante. O que você observa?



Brilho Real

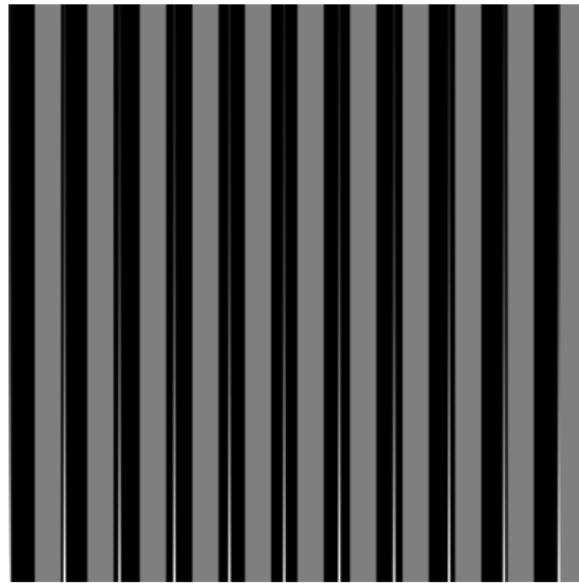


Brilho Percebido



Mascaramento

- Detalhes (ou defeitos) em uma imagem são menos visíveis perto de grandes variações de luminância. Observe a imagem a seguir.



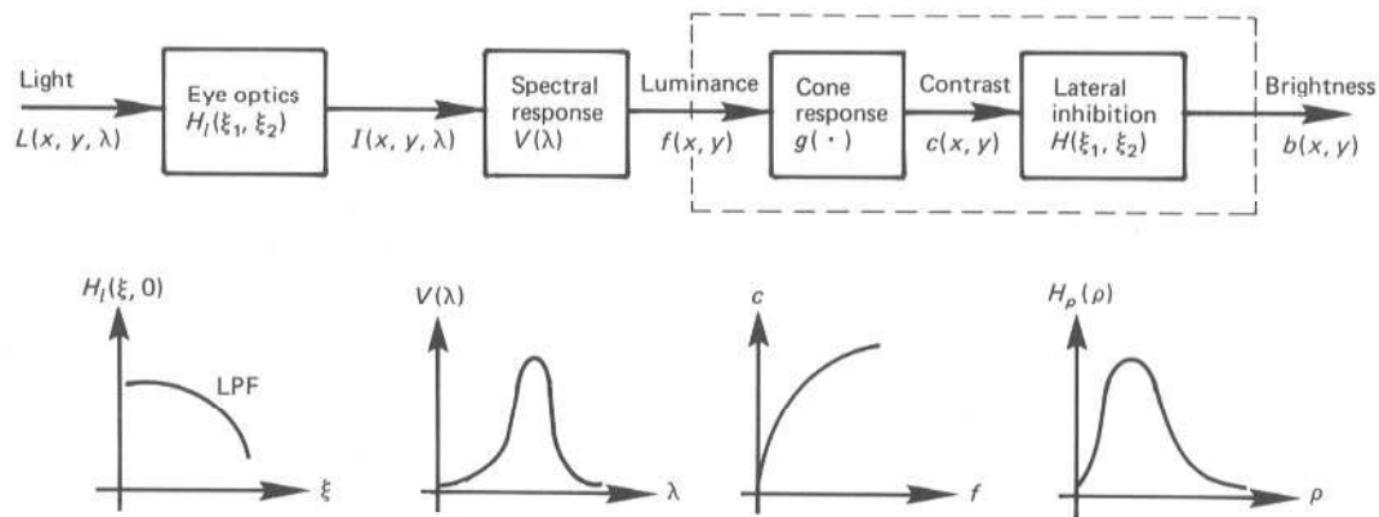
Quantização

- Em um computador, os números não têm precisão infinita.
⇒ Cada amostra só pode assumir um número finito de valores.
- Olho humano: somente consegue distinguir ≈ 256 níveis de brilho.
⇒ Basta usar 8 bits ($2^8 = 256$) para representar cada amostra de vídeo.
- Qual o efeito de se usar um número menor do que 8 bits/amostra?
- Observar “falsos contornos”



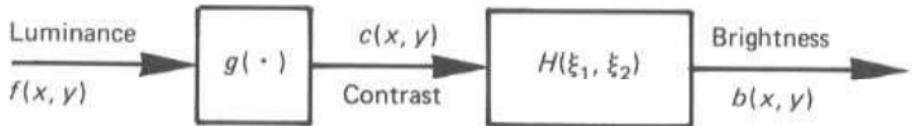


Modelo para Visão Monocromática



(a) Overall monochrome vision model.

Modelo simplificado: $H_1(\omega_1, \omega_2) \cdot G(\omega_1, \omega_2) \simeq G(\omega_1, \omega_2)$



(b) Simplified monochrome vision model.

Avaliação de Qualidade / Critério de Fidelidade de Imagens

- Determinação de qual informação é importante na imagem/vídeo:
 - Permite saber qual informação descartar;
 - Otimização Taxa x Distorção.
- Muitas vezes depende da aplicação:
 - Tipo de mídia;
 - Objetivo da mídia;
 - Visor;
 - Condições de visão, etc.
- Algumas recomendações ITU relativas a avaliação de qualidade:
 - **ITU BT-500:** Methodology for the subjective assessment of the quality of television pictures.
 - **ITU BT-601:** Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios.
 - **ITU BT-709:** Parameter Values for the HDTV Standards for Production and International Programme Exchange.
 - **ITU BT-1210:** Test materials to be used in subjective assessment.
- Duas classes:
 - Critérios Subjetivos;
 - Critérios Objetivos.



Critérios Subjetivos

- Testes realizados com pessoas.
 - Assim refletem com acurácia nossa percepção visual.
- Metodologia básica:
 - Construção de um banco de dados de imagens/vídeos de teste.
 - ▶ Imagens originais.
 - ▶ Imagens processadas/reconstruídas.
 - Grupo de pessoas avaliadoras representativo da população desejada.
 - Avaliação das imagens pelos avaliadores usando um critério predefinido.
- Três Categorias:
 - Testes de Imperfeição: os avaliadores atribuem notas às imagens em função de quão ruins elas são.
 - Testes de Qualidade: os avaliadores atribuem notas às imagens em função de quão boas elas são.
 - Testes Comparativos: os avaliadores atribuem notas às imagens em função comparando a original e a processada/reconstruída.
 - Observações:
 - ▶ Os testes comparativos são considerados como os mais confiáveis pois fornecem uma medida comparativa, que é a mais fácil de ser conferida pela maioria das pessoas.
 - ▶ Os outros requerem uma medida absoluta, mais difícil de ser conferida sem “polarização”/“desvio”.



- Exemplos de atribuição de notas.

Imperfeição	Qualidade	Comparação
5-Imperceptible	A-Excellent	+2 much better
4-Perceptible, not annoying	B-Good	+1 better
3-Somewhat annoying	C-Fair	0 the same
2-Severely annoying	D-Poor	-1 worse
1-Unusable	E-Bad	-2 much worse

Imperfeição

Give your mark! (DSIS method)

Please, vote on the impaired video, keeping in mind the reference one.

Impairments are:

- 5, Imperceptible
- 4, Perceptible, but not annoying
- 3, Slightly annoying
- 2, Annoying
- 1, Very annoying

Your choice: 3

Watch again OK

Qualidade

Give your mark! (SAMVID method)

Please, vote on the video that you have just watched (reference video is not evaluated).

Note: If you have evaluated all the videos you can finish this task. If there are still some sequences left, if you want to change your opinion on some sequence, choose a response by pressing one of the buttons below.

Current sequence B

Reference Watch video Change task

Reference	A	[B]
mark not set		new mark 34

Comparação

Give your mark! (SCACI method)

Please, choose your opinion about the quality of the LEFT picture compared to the quality of the RIGHT picture (For example, choosing 2 or -3 means that the LEFT picture is slightly worse than the RIGHT one).

Circle symbolize your opinion on left and right video correspondingly. Red circle means that video is bad, and green means that video is good.

Your choice: 2

Watch again OK

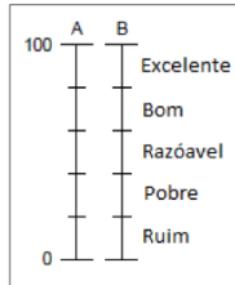
Testes subjetivos mais usados

- DSCQS - Double Stimulus Continuous Quality Scale

- Usa clipes curtos, da ordem de 10s.
- Ordem de apresentação:



- Ou A ou B é escolhida como referência, a outra é a testada.
 - ▶ A escolha da referência é aleatória
- O trecho cinza é uma imagem cinza que é mostrada por alguns segundos
- A cada imagem é dada uma nota entre zero e 100 usando um gabarito como abaixo:

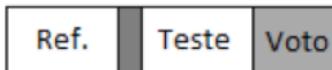


- É o método preferido quando as qualidades das sequências original e degradada são semelhantes, por ser mais sensível a pequenas diferenças de qualidade.



- DSIS - Double Stimulus Impairment Scale (ou Degradation Category Rating)

- Usa clipes curtos, da ordem de 10s.
- Ordem de apresentação:



- A referência é sempre mostrada antes da sequência de teste.
- A quantidade de distorção da sequência é classificada em uma escala de “muito incômodo” a “imperceptível”, conforme abaixo.

<input type="checkbox"/> Imperceptível
<input type="checkbox"/> Perceptível, mas não incômodo
<input type="checkbox"/> Ligeiramente incômodo
<input type="checkbox"/> Incômodo
<input type="checkbox"/> Muito incômodo

- É adequado para avaliar deficiências claramente visíveis, como artefatos causados por erros de transmissão.



- **SSCQE - Single Stimulus Continuous Quality Evaluation**

- Usa programas longos, da 20 a 30 minutos de duração.
 - A sequência de referência não é mostrada.
 - Usando uma barra deslizante, indivíduos avaliam continuamente a qualidade percebida numa escala de ruim a excelente (como a do DSCQS).

- **ACR - Absolute Category Rating**

- Usa clipes curtos, da ordem de 10s.
 - A sequência de referência não é mostrada.
 - As notas são dadas em cinco níveis discretos, de ruim a excelente.
 - É um método mais rápido que o DSCQS ou DSIS.

- **PC - Pair Comparison**

- Usa sequências de uma mesma cena, com diferentes degradações.
 - São exibidas em pares, com todas as combinações possíveis.
 - Os observadores indicam a preferência em cada par.
 - Permite uma discriminação muito fina entre as sequências.

- Esses testes devem ser realizados em situações controladas:
 - Escala de pontuação;
 - Visor empregado;
 - Configuração do visor;
 - Condições de iluminação;
 - Distância de visualização;
 - Tempo de exibição;
 - Quantidade de participantes;
 - Perfis dos participantes;
 - Obtenção de estatísticas dos resultados.

- Algumas recomendações gerais:

- A escolha do conteúdo espacial e temporal das sequências deve levar em conta o objetivo do experimento.
- O conjunto das cenas de teste deve se estender pela gama de qualidade geralmente encontrada para as condições de ensaio.
- Para comparação entre diferentes laboratórios, é obrigatório usar o mesmo conjunto de sequências de origem.
- As sequências devem ser apresentadas numa ordem pseudo-aleatória, tentando evitar que sequências geradas a partir de uma mesma referência sejam mostradas uma logo depois da outra.
- As condições de visualização devem seguir as normas
- O tamanho e o tipo do monitor devem ser adequados para a aplicação sobre teste.
- Pode ser necessário calibrar o monitor.
- É melhor usar a tela inteira para a exibição das sequências.
 - ▶ Se não for possível, exibir em uma janela, com um fundo 50% cinza ao redor ($Y=Cb=Cr=128$).

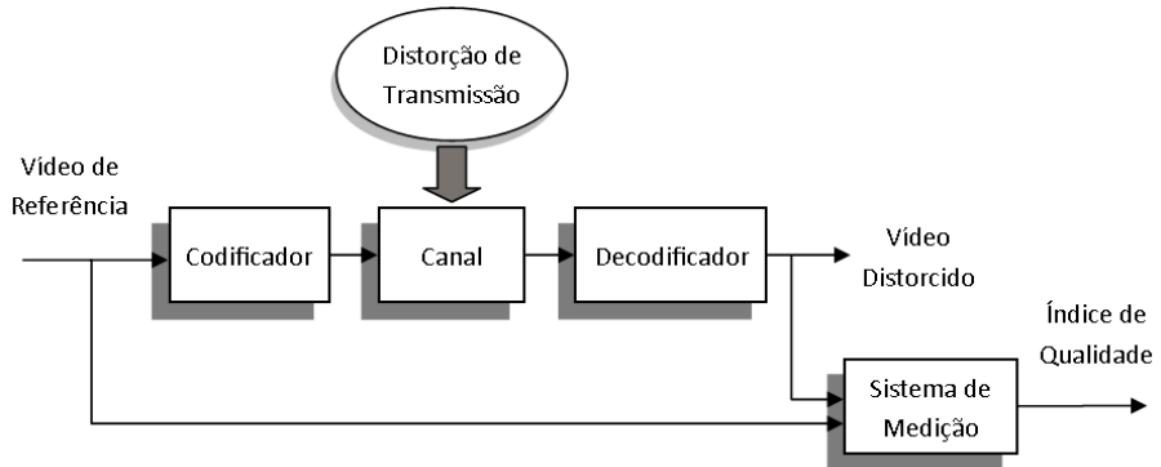


Critérios Objetivos

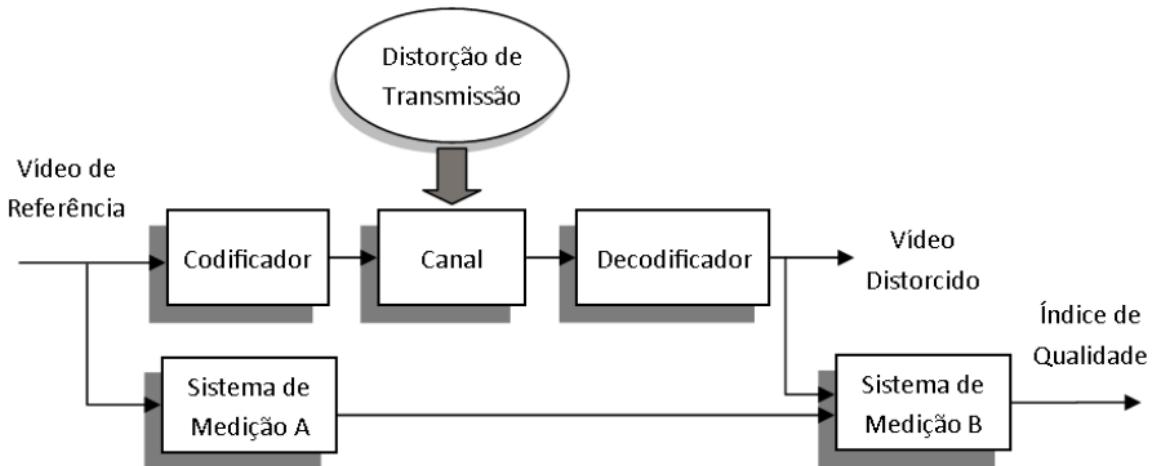
- Consistem do emprego de equações (métricas objetivas) para avaliar a distorção de uma imagem processada / reconstruída.
 - Não estão necessariamente vinculadas ao SVH, apesar de haver várias tentativas nessa área.
 - Podem ser realizadas por computadores, processadores, máquinas em geral.
 - ⇒ Implementação em tempo real.
- Muito úteis para comparar os mais diversos processamentos a aplicar numa imagem.
- Podem ser de 3 tipos:



- Referência completa (full-reference): a imagem/vídeo original e a degradada estão disponíveis para o cálculo da métrica.



- Referência reduzida (reduced reference): apenas a imagem degradada e algumas medidas sobre a imagem original estão disponíveis para o cálculo da métrica (p.ex., histograma).



- Sem referência (no-reference): apenas a imagem degradada está disponível para o cálculo da métrica.



- Erro / Distorção de um pixel:

$$\text{erro}(c, l) = \hat{l}(c, l) - l(c, l),$$

onde:

- (c, l) é a posição do pixel;
- l é o valor original do pixel e
- \hat{l} é o valor do pixel na imagem processada / reconstruída;
- l pode ser: luminância ou crominância: em geral utiliza-se só a luminância.



A partir dessa medida de distorção podemos definir as mais variadas métricas:

- Erro total:

$$\text{erro_total} = \sum_{c=1}^C \sum_{l=1}^L (\hat{I}(c, l) - I(c, l))$$

- Soma dos erros absolutos (SAD):

$$\text{SAD} = \sum_{c=1}^C \sum_{l=1}^L |\hat{I}(c, l) - I(c, l)|$$

- Média dos erros absolutos (MSAD):

$$\text{MSAD} = \frac{\sum_{c=1}^C \sum_{l=1}^L |\hat{I}(c, l) - I(c, l)|}{CL}$$

- Erro quadrático médio (MSE):

$$\text{MSE} = \frac{\sum_{c=1}^C \sum_{l=1}^L [\hat{l}(c, l) - l(c, l)]^2}{CL}$$

- Raiz do erro médio quadrático (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{c=1}^C \sum_{l=1}^L [\hat{l}(c, l) - l(c, l)]^2}{CL}}$$

- Razão sinal ruído (SNR):

$$\text{SNR} = \frac{\sum_{c=1}^C \sum_{l=1}^L [\hat{l}(c, l)]^2}{\sum_{c=1}^C \sum_{l=1}^L [\hat{l}(c, l) - l(c, l)]^2}$$

- Razão sinal ruído (SNR_{RMS}):

$$\text{SNR}_{\text{RMS}} = \sqrt{\frac{\sum_{c=1}^C \sum_{l=1}^L [\hat{l}(c, l)]^2}{\sum_{c=1}^C \sum_{l=1}^L [\hat{l}(c, l) - l(c, l)]^2}}$$

- Razão sinal ruído de pico (PSNR):

$$\text{PSNR} = 10 \log_{10} \left\{ \frac{(N - 1)^2}{\frac{1}{CL} \sum_{c=1}^C \sum_{l=1}^L [\hat{I}(c, l) - I(c, l)]^2} \right\},$$

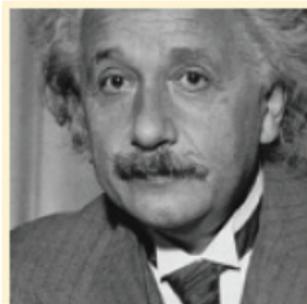
onde N é o valor máximo que o pixel pode assumir.

- Essas métricas são muito utilizadas para comparação de resultados, em especial o PSNR.
- Para implementação de escolhas em algoritmos a SAD é mais usada (complexidade reduzida).

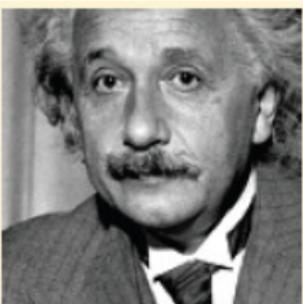
Limitações do Erro Médio Quadrático

- Como o erro médio quadrático leva em conta apenas os valores dos pixels e não a relação entre os pixels (estrutura) em uma imagem, em muitos casos ele não é uma métrica adequada para medir a qualidade de imagens.
- Nas imagens a seguir, a de cima à esquerda é a original, e todas as demais possuem o mesmo MSE (ou PSNR), exceto a de baixo à direita, em que o MSE é o dobro.
 - a) é a original
 - b) o contraste é maior
 - c) tem um brilho (nível DC) diferente
 - d) contaminada com ruído gaussiano
 - e) contaminada com ruído impulsivo
 - f) comprimida em JPEG
 - g) borrada
 - h) deslocada para a esquerda

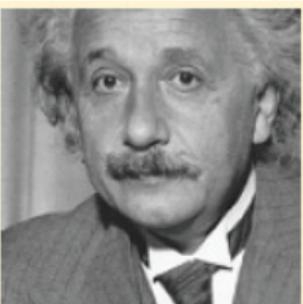




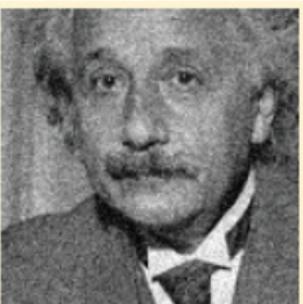
MSE=0, SSIM=1
CW-SSIM=1
(a)



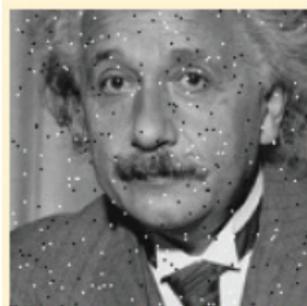
MSE=306, SSIM=0.928
CW-SSIM=0.938
(b)



MSE=309, SSIM=0.987
CW-SSIM=1.000
(c)



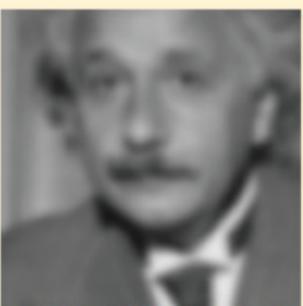
MSE=309, SSIM=0.576
CW-SSIM=0.814
(d)



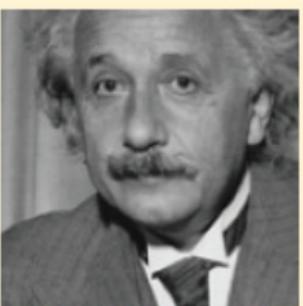
MSE=313, SSIM=0.730
CW-SSIM=0.811
(e)



MSE=309, SSIM=0.580
CW-SSIM=0.633
(f)

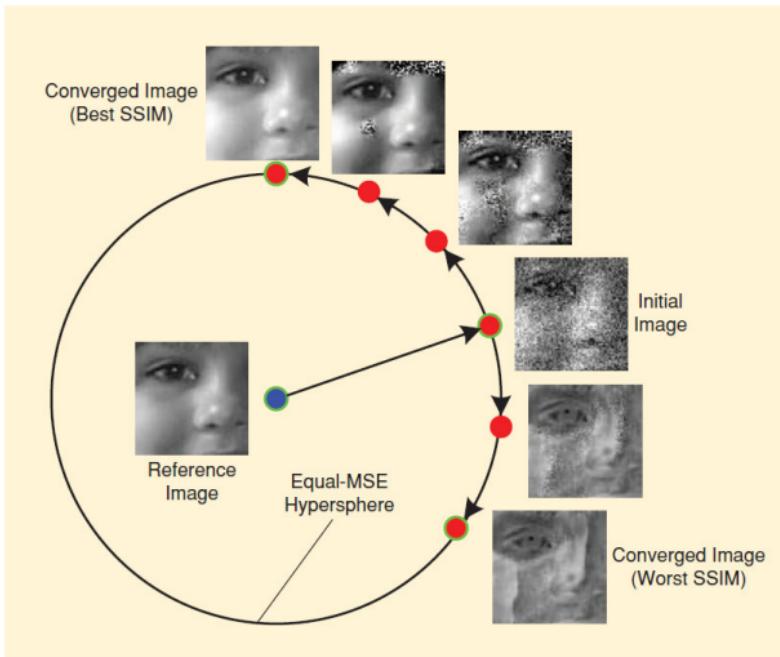


MSE=308, SSIM=0.641
CW-SSIM=0.603
(g)

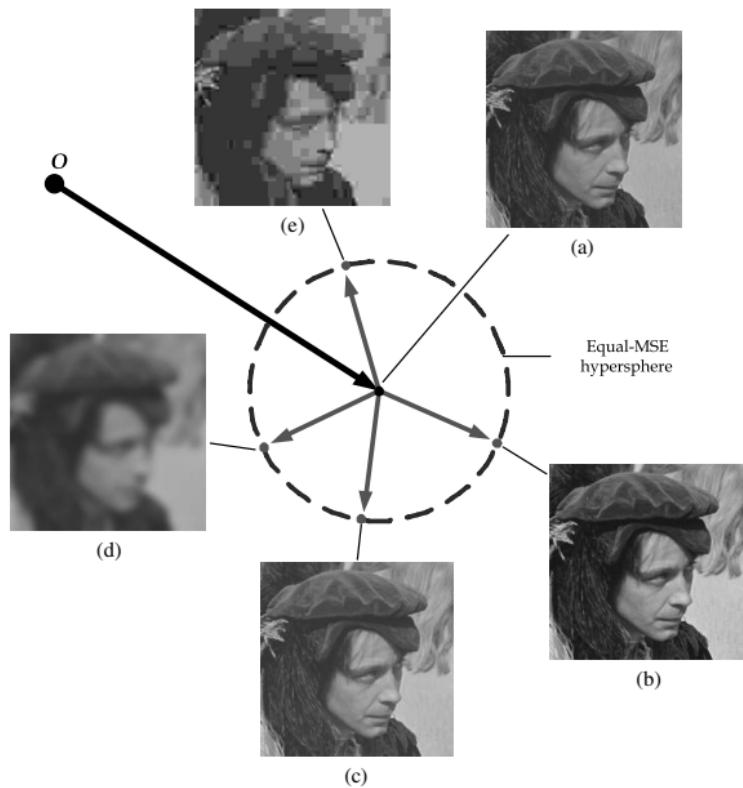


MSE=694, SSIM=0.505
CW-SSIM=0.925
(h)

- Já neste caso, a imagem original está no centro e todas as imagens em volta possuem o mesmo MSE (PSNR).



- Neste caso, a) é a original, b) com contraste maior ($MSE=181$), c) com brilho maior ($MSD=181$), d) borrada ($MSE=181$) e e) comprimida em JPEG ($MSE=181$).



- Nota-se que as medidas objetivas derivadas do MSE ainda deixam muito a desejar.
- Há intensa pesquisa hoje para se achar medidas objetivas alternativas (p.ex., SSIM - Structural Similarity Index).
- Entretanto, o MSE (PSNR) ainda é muito usado.

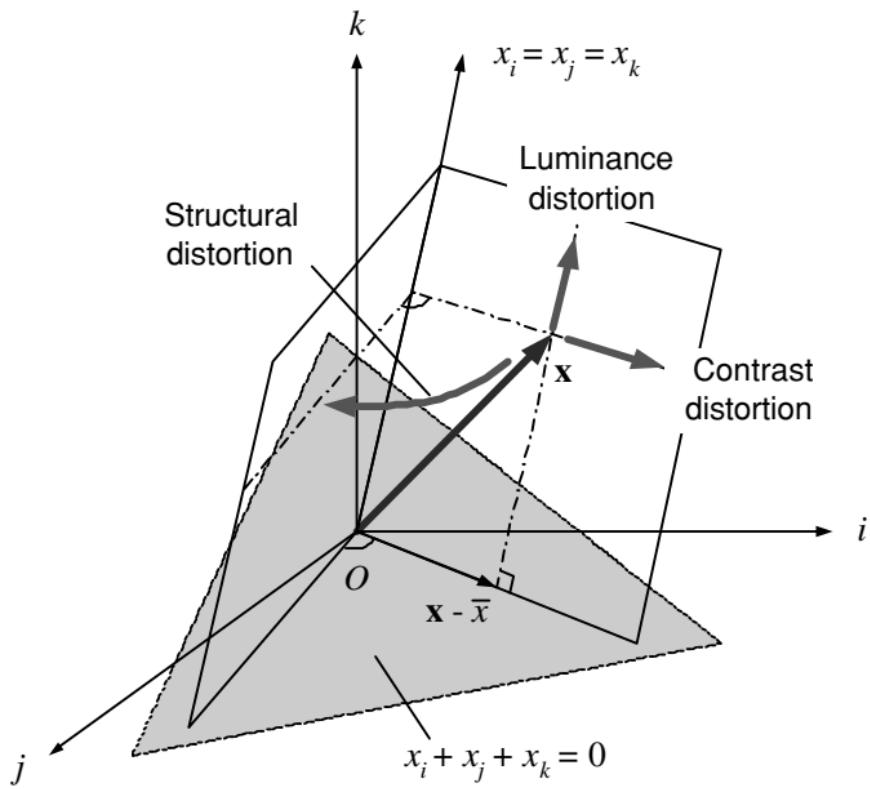


SSIM - Structural Similarity Metric

- A relação do SSIM com a qualidade subjetiva é melhor que a do PSNR.
- É uma métrica que vem ganhando força nos últimos anos, apesar do PSNR ainda ser mais usado.
- Divide a semelhança entre imagens em três fatores distintos:
 - Brilho (nível DC)
 - Contraste (faixa dinâmica)
 - Estrutura
- Os dois primeiros fatores não são estruturais, isto é, uma diferença entre eles não modifica a estrutura da imagem:
 - Apesar de importantes para a percepção subjetiva, em geral mudanças no brilho e no contraste podem ser facilmente compensadas com a soma de um valor constante ou um remapeamento simples dos pixels.
 - A estrutura é usada com base na suposição que o sistema visual humano é altamente adaptado para extrair informações estruturais de uma imagem.



- Uma interpretação geométrica de distorções nestes 3 fatores (brilho, contraste e estrutura) seria:



- Uma medida do brilho (nível DC) de uma imagem \mathbf{x} é o seu valor médio,

$$\mu_{\mathbf{x}} = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N x(m, n)$$

- Uma vez calculada a sua média, o seu contraste pode ser medido pelo valor RMS da imagem subtraída da média,

$$\begin{aligned}\sigma_{\mathbf{x}} &= \frac{1}{MN} \sqrt{\sum_{m=1}^M \sum_{n=1}^N [x(m, n) - \mu_{\mathbf{x}}(m, n)]^2} \\ &= \frac{1}{MN} \|\mathbf{x} - \mu_{\mathbf{x}}\|\end{aligned}$$

- A “estrutura pura” de uma imagem seria então a imagem com média zero e contraste unitário, isto é,

$$\tilde{\mathbf{x}} = \frac{\mathbf{x} - \mu_{\mathbf{x}}}{\sigma_{\mathbf{x}}}$$

- Devemos notar que duas medidas acima são escalares, isto é, cada uma é representada por apenas um número (brilho/contraste) e a outra é vetorial, pois a estrutura de uma imagem $M \times N$ possui MN pixels, podendo ser vista como um vetor no \mathbb{R}^{MN} .
- Uma maneira de comparar dois números (escalares) se baseia no fato de que a média aritmética é sempre maior que a média geométrica, e elas só são iguais quando os dois números são iguais:

$$(a - b)^2 \geq 0 \quad \Rightarrow \quad a^2 - 2ab + b^2 \geq 0 \quad \Rightarrow \quad \frac{a^2 + b^2}{2} \geq ab$$

- Assim, a razão $\frac{2ab}{a^2 + b^2}$ é uma boa medida de similaridade de a e b , pois:
 - Seu valor é sempre menor ou igual a 1, só sendo igual a 1 quando $a = b$.
 - Dividindo o numerador e o denominador por ab , vemos que ela é igual a $\frac{2}{\frac{a}{b} + \frac{b}{a}}$. Assim ela não depende dos valores absolutos de a e b , mas apenas da sua razão.
 - Quando mais diferentes forem a e b , maior será $(a - b)^2$, e portanto menor será o seu valor.



- Assim, a medida de diferença de brilho entre duas imagens \mathbf{x} e \mathbf{y} seria então

$$b(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2}$$

- No caso da comparação de duas regiões pretas, para evitar o problema de dividir 0/0, fazemos

$$b(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

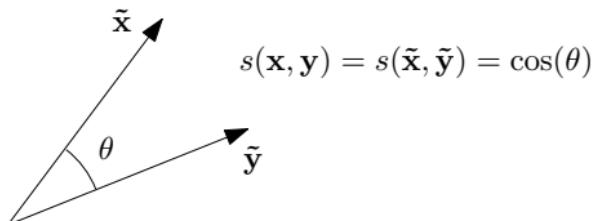
onde C_1 é uma constante pequena.

- Analogamente, no caso da diferença entre contrastes, temos:

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

onde C_2 é uma constante pequena para evitar o 0/0 no caso de regiões uniformes (contraste zero)

- Já no caso da comparação entre as estruturas, o que se faz é considerar as imagens x e y como vetores no \mathbb{R}^{MN} e considerar que as imagens são tão mais parecidas quanto menor for o ângulo entre elas.



- Quanto mais parecidas as imagens, menor o ângulo e mais próximo de 1 o cosseno.
- O ângulo entre x e y é o mesmo que entre as imagens com a estrutura apenas, \tilde{x} e \tilde{y} . Como elas possuem comprimento unitário, o ângulo é dado por

$$\begin{aligned}\cos(\theta) &= \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \tilde{x}(m, n) \tilde{y}(m, n) \\ &= \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \left(\frac{x(m, n) - \mu_x(m, n)}{\sigma_x} \right) \left(\frac{y(m, n) - \mu_y(m, n)}{\sigma_y} \right) \\ &= \frac{\sigma_{xy}}{\sigma_x \sigma_y}\end{aligned}$$

- Evitando a divisão 0/0 que ocorreria no caso de imagens de baixo contraste, a medida de similaridade entre estruturas seria então

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{\mathbf{xy}} + C_3}{\sigma_x \sigma_y + C_3}$$

onde C_3 é uma constante pequena.

- Então, a medida de similaridade entre duas imagens, conhecida como Structural Similarity Metric (SSIM) é dada pelo produto dos três fatores - diferença entre brilho, contraste e forma:

$$\begin{aligned} S(\mathbf{x}, \mathbf{y}) &= b(\mathbf{x}, \mathbf{y})c(\mathbf{x}, \mathbf{y})s(\mathbf{x}, \mathbf{y}) \\ &= \left[\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right] \left[\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right] \left[\frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \right] \end{aligned}$$

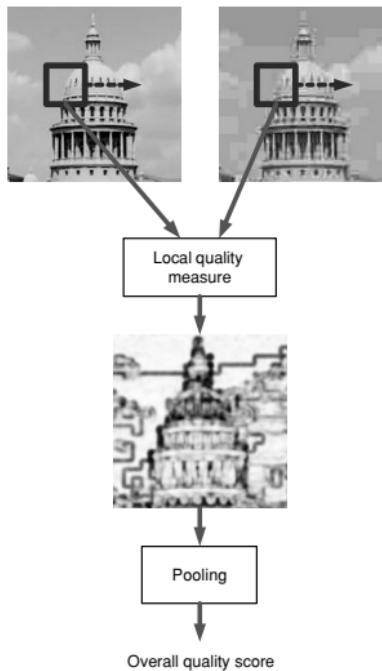
- Se fizermos $C_3 = C_2/2$, o numerador do 2º termo cancela com o denominador do 3º termo e o valor do SSIM fica

$$S(\mathbf{x}, \mathbf{y}) = \left[\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right] \left[\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right]$$

- Note que todos os termos do SSIM estão entre zero e um, então o valor do SSIM está sempre entre zero e 1, sendo igual a 1 para imagens idênticas.



- Na prática, não se calcula o SSIM para a imagem toda, mas se calcula para janelas da imagem:
 - Se desloca uma janela por cima da imagem, e se calcula o SSIM para cada posição da janela.
 - Depois, se faz uma média (que pode ser ponderada) de todos os valores de SSIM para todas as posições das janelas (polling).



Exemplos

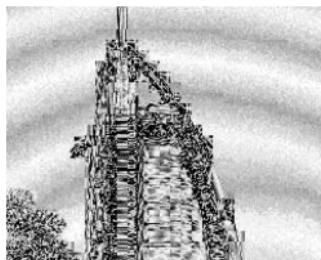
- A seguir vemos, para imagens com diversos tipos de distorção, a comparação entre o mapa gerado pelo SSIM de todas as janelas (direita) com o erro absoluto entre todas as janelas (esquerda). (a) contém a imagem original, (b) a imagem distorcida, (c) o mapa de erro absoluto e (d) o mapa dos valores de SSIM.
- Distorção: Compressão JPEG



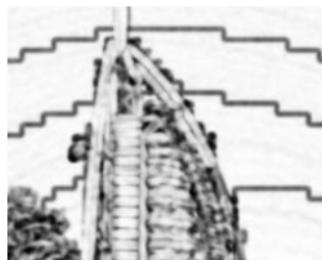
(a)



(b)



(c)



(d)

Exemplos

- Distorção: Compressão JPEG2000



(a)



(b)



(c)



(d)

Exemplos

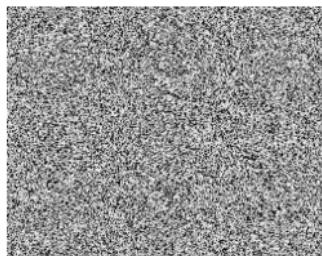
- Distorção: Ruído Branco Gaussiano



(a)



(b)



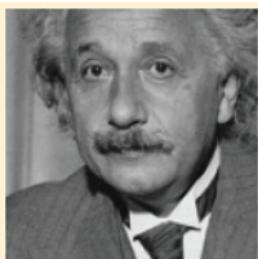
(c)



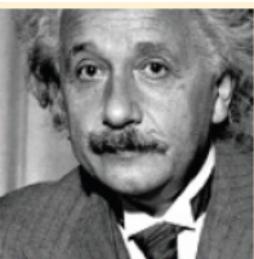
(d)

Exemplos

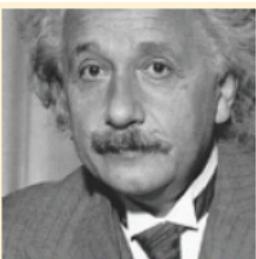
- Repetimos aqui as imagens de Einstein mostradas anteriormente com diversos tipos de degradação, para que possamos observar os valores de SSIM e PSNR:



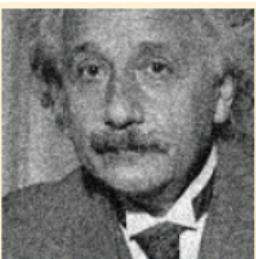
MSE=0, SSIM=1
CW-SSIM=1
(a)



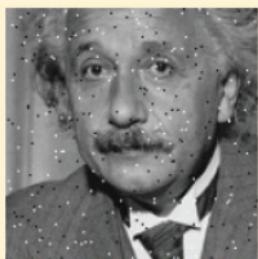
MSE=306, SSIM=0.928
CW-SSIM=0.938
(b)



MSE=309, SSIM=0.987
CW-SSIM=1.000
(c)



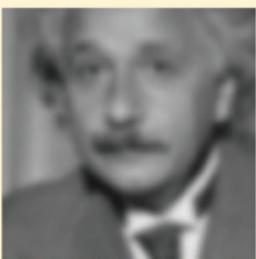
MSE=309, SSIM=0.576
CW-SSIM=0.814
(d)



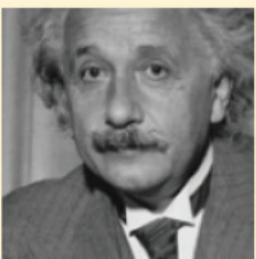
MSE=313, SSIM=0.730
CW-SSIM=0.811
(e)



MSE=309, SSIM=0.580
CW-SSIM=0.633
(f)



MSE=308, SSIM=0.641
CW-SSIM=0.603
(g)



MSE=694, SSIM=0.505
CW-SSIM=0.925
(h)

Limitações do SSIM

- O SSIM falha quando há uma diferença significativa na estrutura.
- Entretanto, algumas vezes imagens com estruturas muito diferentes podem ser parecidas:
 - Se uma imagem está deslocada em relação à outra, o valor de SSIM tende a ser pequeno - vide imagem (h) do Einstein - ela é a imagem do Einstein apenas deslocada de alguns pixels, e o valor de SSIM é 0.505.
 - O mesmo ocorre para pequenas rotações e escalamentos.
- Uma solução para isto é o uso do SSIM no domínio da transformada wavelet complexa (CW-SSIM), que é robusta ao deslocamento, ao escalamento e à rotação.



Video Quality Experts Group - VQEG

- Grupo da ITU-T formado em 1997 com o objetivo de criar um método para avaliação objetiva de qualidade de vídeo.
- Fase 1:
 - Para sequências codificadas em MPEG-2.
 - ▶ Diferentes perfis e níveis
 - ▶ Concatenação de codificadores
 - ▶ Conversões entre vídeos analógicos e digitais
 - ▶ Erros de transmissão
 - Propostas de: CPqD (Brasil), EPFL (Suíça), KDD (Japão), a KPN Research / Swisscom (Holanda / Suíça), a NASA (EUA), NHK / Mitsubishi (Japão), NTIA / ITS (EUA), TAPETES (UE) Technische Universität Braunschweig (Alemanha), e Tektronix / Sarnoff (EUA).
 - Atributos avaliados:
 - ▶ Precisão
 - ▶ Motonicidade
 - ▶ Consistência das previsões das métricas.
 - Todas as propostas se mostraram estatisticamente equivalentes ao PSNR.
 - ▶ Correlação de aproximadamente 70% com notas subjetivas.

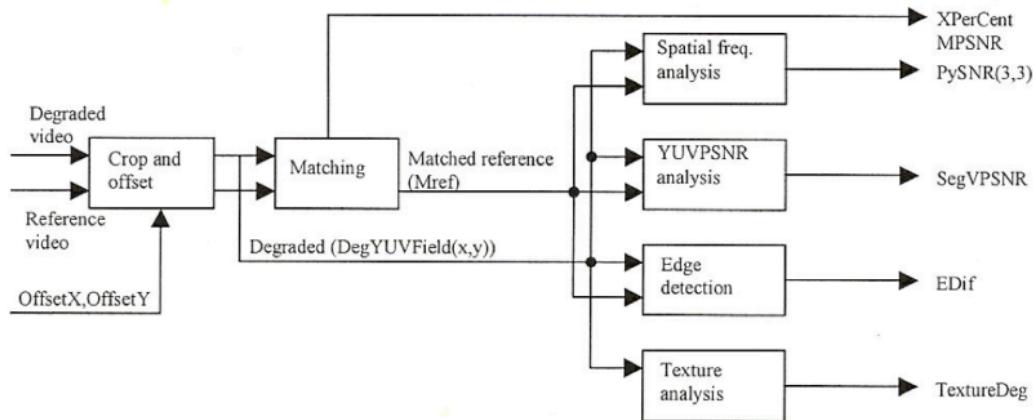
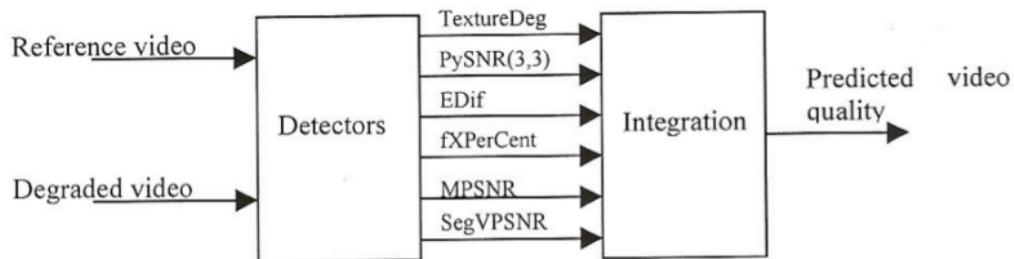


- Fase 2:

- Concluída em 2003
- Forte foco em vídeos digitais codificados, com vasta gama de distorções.
- Foram produzidas 128 sequências de teste.
- O ground truth da qualidade das sequências foi obtido usando o método DSCQS.
- Propostas de: British Telecom (Reino Unido), Chiba University (Japão), o CPqD (Brasil), a NASA (EUA), NTIA / ITS (EUA) e Yonsei University (Coréia).
- Correlação de até 94% com avaliação subjetiva.
- São a base da recomendação ITU-T J.144.



British Telecom Full Reference Metric - BTFR



British Telecom Full Reference Metric - BTFR

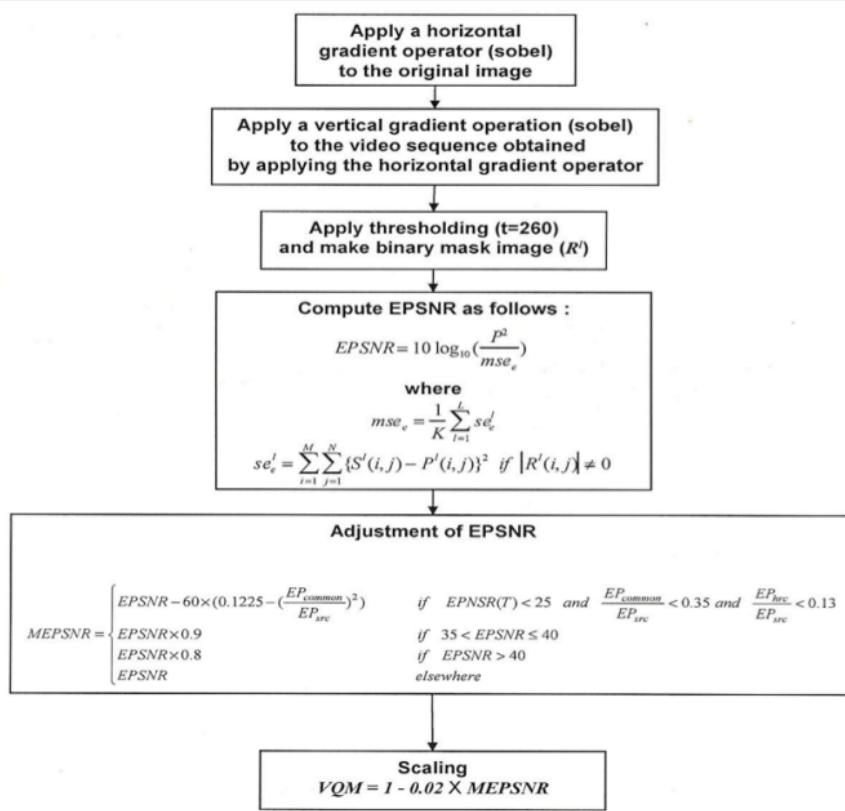
- TextureDeg: número de vezes que a variação de luminância ao longo de uma linha muda de direção.
- PySRN: PSNR após uma transformação piramidal (equivalente a uma filtragem passa-banda em oitavas).
- EDif: diferença entre os mapas de arestas dos quadros original e degradado.
- FXPerCent: estatísticas de casamento entre blocos dos quadros original e degradados.
- MPSNR: PSNR dos blocos casados.
- SegVPSNR: PSNR segmentado baseado em componentes de cores.

Métrica da Yonsei University - Edge Peak Signal to Noise Ratio - EPSNR

- O processamento é feito apenas em regiões de interesse (em volta de arestas).
 - Ela calcula uma relação sinal ruído apenas em volta das arestas.
- O casamento das sequências é feito baseado em versões subamostradas das sequências e apenas nas regiões de interesse (áreas com muito movimento).



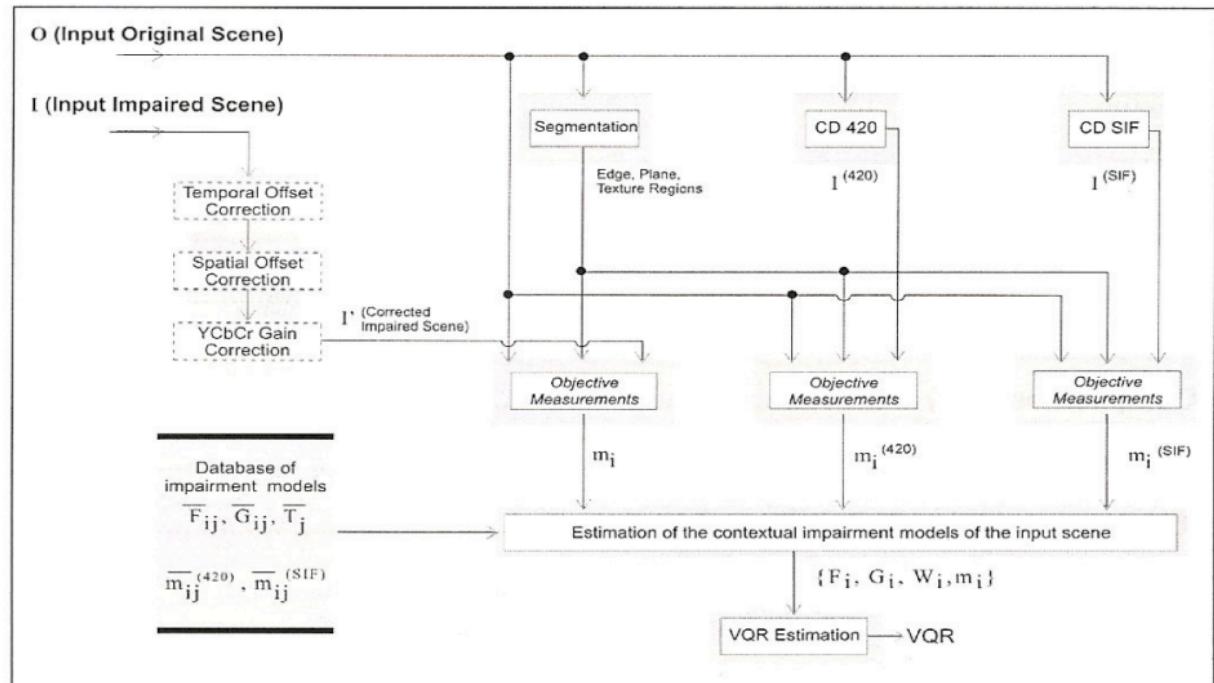
Edge Peak Signal to Noise Ratio - EPSNR



Métrica do CPqD- Image Evaluation Based on Segmentation - IES

- ① Caso as entradas original e degradada no espaço e no tempo.
- ② As sequências são segmentadas em regiões Planas, Areastas e Texturas.
- ③ Para cada quadro, os MSEs para cada região e cada componente de vídeo são calculadas, o que dá 9 medidas por quadro.
- ④ É feita uma média das medidas no espaço e no tempo, e pesos calculados com base em um conjunto de treinamento são usados para gerar a nota objetiva final.

Métrica do CPqD- Image Evaluation Based on Segmentation - IES



Métrica do NTIA - Video Quality Metrics - VQM

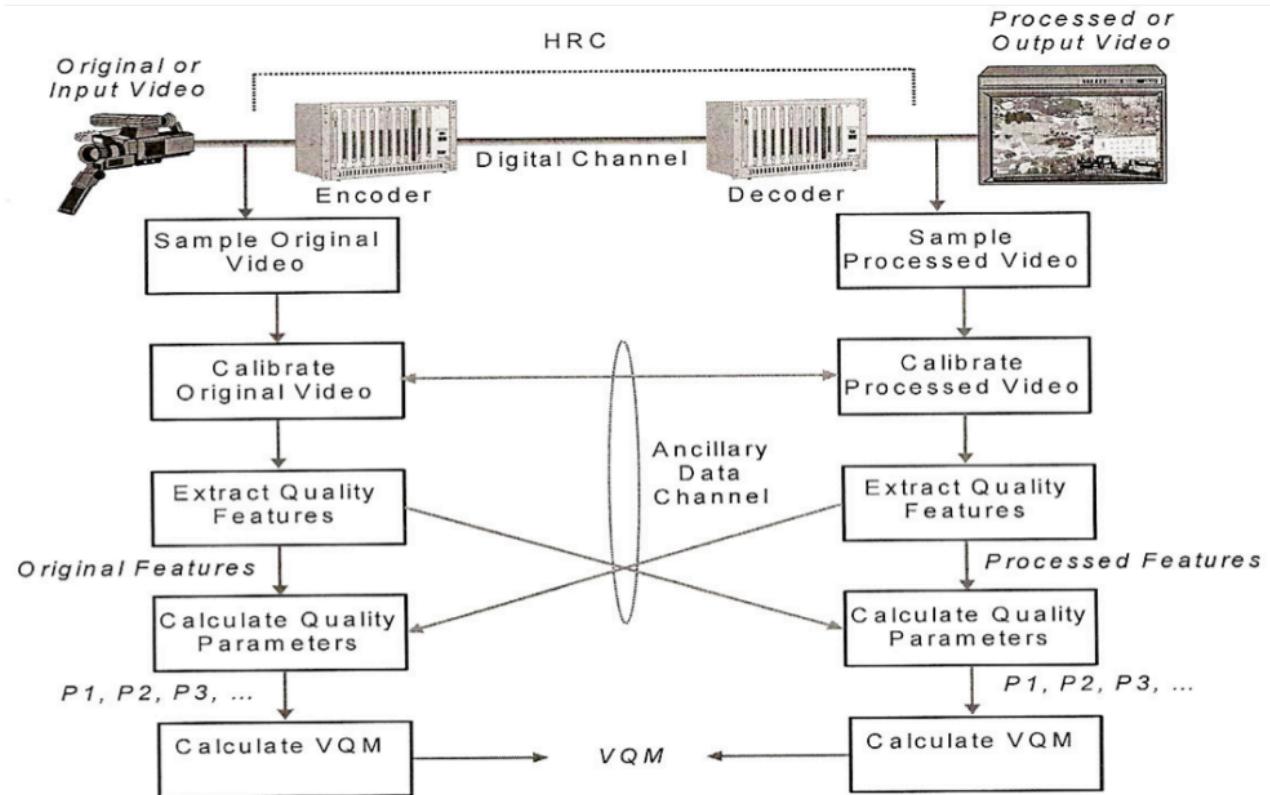
- NTIA - National Telecommunications and Information Administration.
- Métrica de propósito geral para sistemas de vídeo que abrangem uma vasta gama de qualidade e taxa de bits.
 - Funciona para vários tipos de codificação.
 - Atende a taxas de 10kbit/s a 45Mbit/s.
 - Atende também a vídeo analógico e sistemas baseados em fita.
 - Possui também uma versão especializada para televisão e outra para sistemas de vídeo conferência.
 - Similar ao BTFR, também combina uma série de métricas para obter uma nota final
 - Usa ROIs (regiões de interesse) tanto determinadas automaticamente quanto especificadas pelo usuário.
 - De todas as métricas na ITU-T J.144, é o que possui maior correlação com a qualidade subjetiva dos vídeos (94%).
 - Possui software disponível em www.its.bldrdoc.gov/n3/video/



Métrica do NTIA - Video Quality Metrics - VQM

- O cálculo da métrica compreende os seguintes passos:
 - Amostra os vídeos original e degradado.
 - Calibra os dois vídeos (ganho e casamento espaço-temporal)
 - Extrai características perceptuais
 - Calcula os parâmetros de qualidade (combinando as características perceptuais) para obter uma nota final.

Métrica do NTIA - Video Quality Metrics - VQM

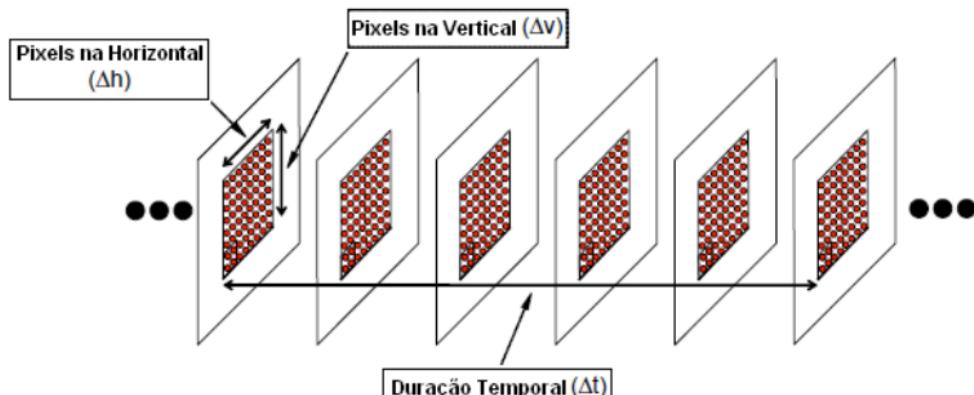


Métrica do NTIA - Video Quality Metrics - VQM

- Calibração:
 - Registro espacial
 - ▶ Deslocamento horizontal e vertical
 - Determinação da região válida
 - ▶ Elimina pixels “over-scanned” e efeitos de transição devido à filtragem.
 - Offset de ganho
 - ▶ Equaliza brilho e contraste entre as sequências.
 - Registro temporal
 - ▶ Remove os atrasos entre as sequências.

Métrica do NTIA - Video Quality Metrics - VQM

- Características perceptuais:
 - Caracterizam as mudanças perceptuais nas propriedades espaciais, temporais e de crominância dos vídeos.
 - As características são medidas através do seguinte:
 - ▶ Filtro perceptual + função matemática + limiar de perceptibilidade.
 - Os filtros perceptuais realçam alguns aspectos da qualidade do vídeo.
 - ▶ Um realce de bordas identifica com maior precisão embaçamento e efeito de blocos.
 - ▶ Componentes de cor são úteis para identificar distorções de matiz e erros de transmissão.
 - ▶ A diferença no tempo entre quadros destaca movimentos não naturais e bruscos.
 - A sequência de vídeo é dividida em regiões espaço-temporais, $b(s, t)$.
 - ▶ A figura a seguir ilustra um bloco $b(s, t)$ de 8 pixels na horizontal, 8 pixels na vertical e 6 quadros de vídeo, num total de 384 pixels.

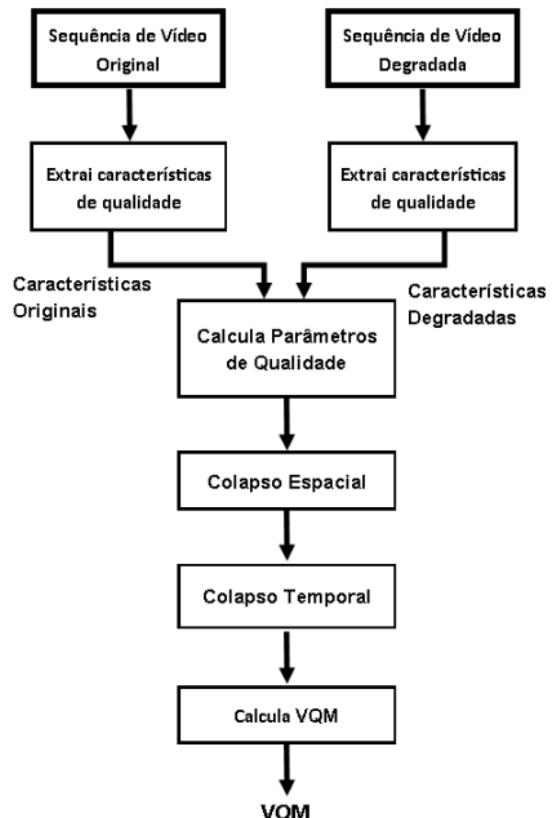


- Características extraídas:
 - Baseadas no gradiente espacial: embaçamento e blocagem.
 - Baseadas na crominância: médias de C_b e C_r .
 - Baseadas no contraste: desvio padrão do componente de luminância (Y).
 - Baseadas no tempo: desvio padrão da diferença absoluta entre quadros consecutivos.
- As características são extraídas por região (blocos espaço-temporais)
 - Características $f_o(s, t)$ e $f_d(s, t)$ do sinal original e degradado são extraídas de cada bloco $b(s, t)$ usando uma função matemática simples.
 - ▶ As duas funções mais utilizadas são a média e o desvio padrão.
 - Opcionalmente pode ser aplicado um limiar de perceptibilidade para reduzir a sensibilidade do parâmetro às distorções imperceptíveis - o valor “satura” se abaixo ou acima de um limiar.

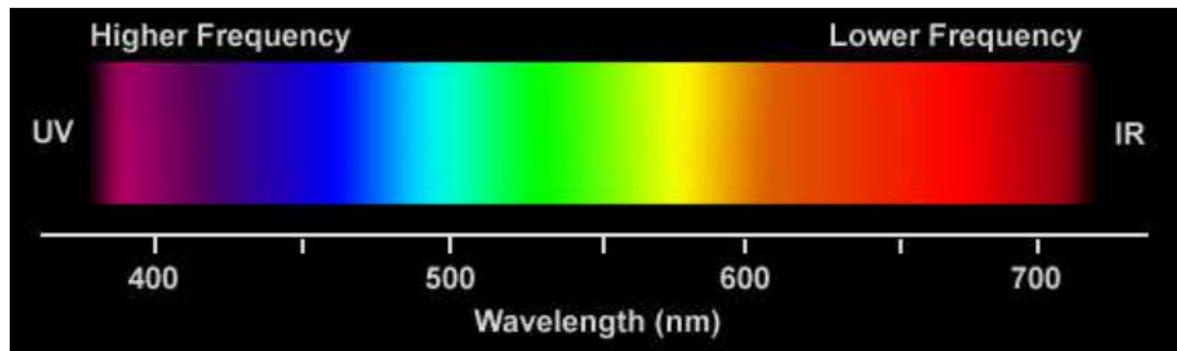
Métrica do NTIA - Video Quality Metrics - VQM

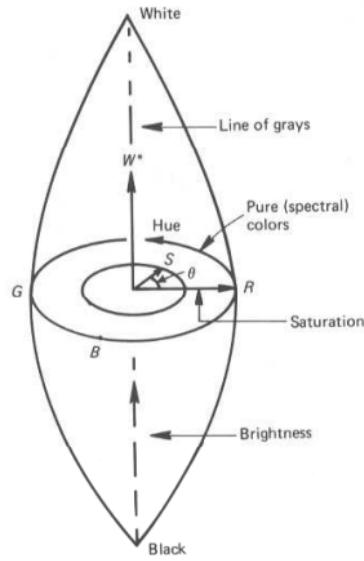
- Cálculo dos parâmetros de qualidade (segue os passos abaixo):
 - Compara os valores do vídeo original e degradado (taxa de erro e distância euclidiana ao longo do tempo) $\rightarrow c(s, t)$;
 - Executa o colapso espacial: $c(s, t) \rightarrow p(t)$
 - ▶ Pode ser feito com média, desvio padrão, média dos menores 5%, média dos maiores 5%. etc.
 - Executa o colapso temporal: $p(t) \rightarrow q$.
 - Executa escalamento não linear e “clipping”.
- A nota final é computada como uma combinação linear de parâmetros baseados em um modelo perceptual geral projetado para maximizar a correlação entre as notas objetivas e subjetivas.

Métrica do NTIA - Video Quality Metrics - VQM



Cor



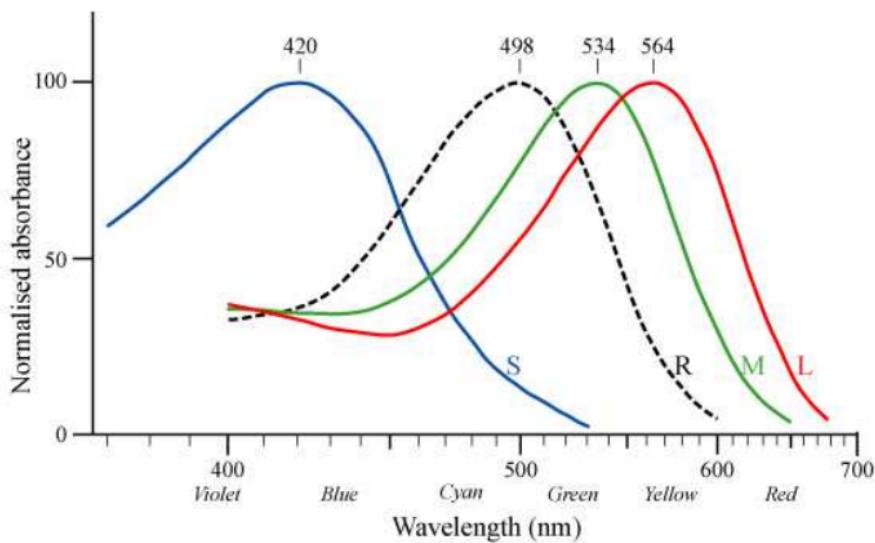


W : brilho

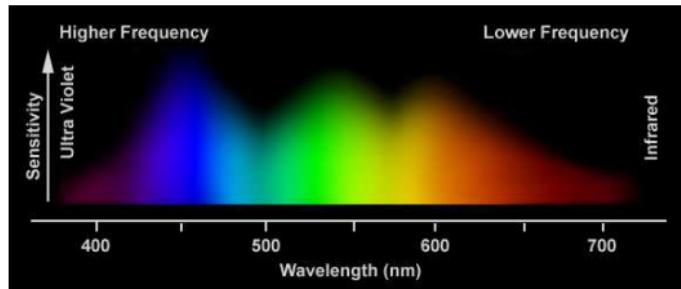
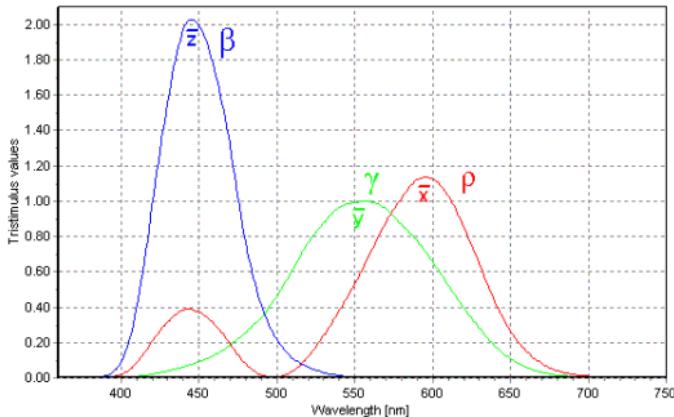
θ : matiz

S : Saturação (o quanto de “branco” possui uma cor)

- O olho humano possui 3 tipos de cones, sensíveis a comprimentos de onda nas faixas do verde-amarelado, verde e azul.
- Curvas “Tristimulus”: todas as cores que percebemos são uma combinação das respostas dos cones aos diferentes comprimentos de onda.



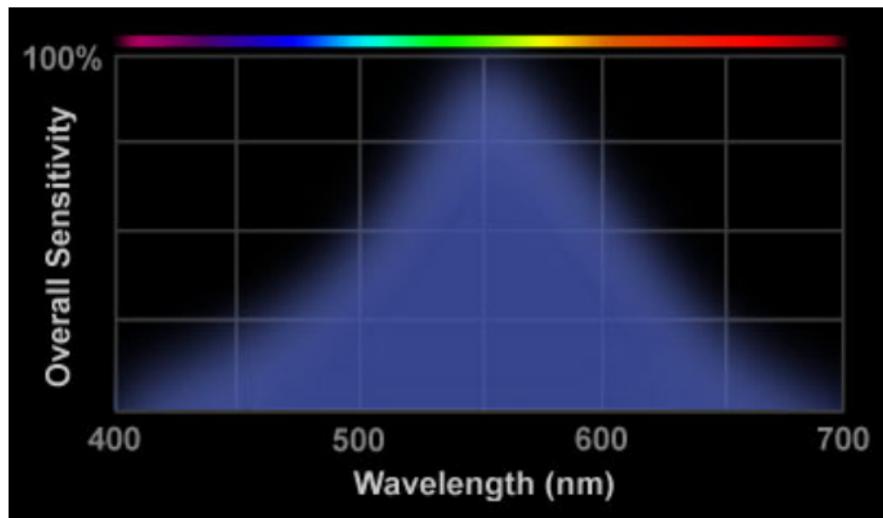
CIE 1964 Color Matching Functions (10-degree observer)



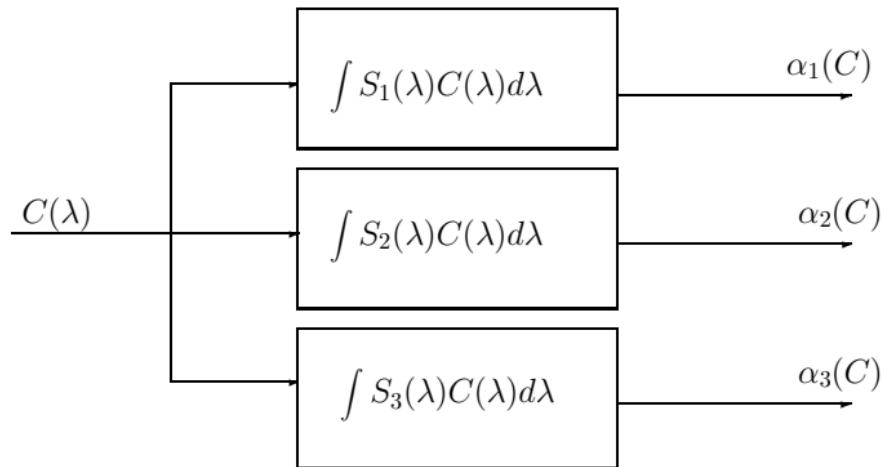
- O cérebro combina as saídas dos 3 cones (cor) e produz a sensação correspondente a uma dada cor.



- Eficiência luminosa relativa e luminância



- Por este motivo, todas as cores percebidas pelo ser humano podem ser **simuladas** com uma combinação de 3 cores básicas (Thomas Young, 1802).



$$\alpha_i(C) = \int_{\lambda_{min}}^{\lambda_{max}} S_i(\lambda)C(\lambda)d\lambda$$

- Sejam C_1 e C_2 tal que $\alpha_i(C_1) = \alpha_i(C_2)$, $i = 1, 2, 3$. Então C_1 e C_2 são percebidas como idênticas.

Metameria de Cores

- Duas cores com distribuições espectrais diferentes podem parecer a mesma
 - Isto é, possuir os mesmos valores RGB
- Duas cores metaméricas irão parecer idênticas para o Sistema Visual Humano (SVH).
- A maioria das câmeras e visores são projetados com sensores que imitam a resposta do SVH.

Casamento e Reprodução de Cores

Problema Básico: reprodução de cores usando um conjunto de fontes de luz.

Número mínimo: 3 fontes, $p_k(\lambda)$, $k = 1, 2, 3$.

Seja $\int_{\lambda_{min}}^{\lambda_{max}} p_k(\lambda) d\lambda = 1$, $k = 1, 2, 3$. $p_k(\lambda)$ são linearmente independentes

Cor $C(\lambda) \Rightarrow$ 3 primárias com pesos β_k , $k = 1, 2, 3$

$\Rightarrow \sum_{k=1}^3 \beta_k p_k(\lambda)$ deve ser percebido como $C(\lambda)$

$$\Rightarrow \alpha_i(C) = \int \sum_{k=1}^3 \beta_k p_k(\lambda) S_i(\lambda) d\lambda = \sum_{k=1}^3 \beta_k \int p_k(\lambda) S_i(\lambda) d\lambda$$



Definindo a resposta do cone i a uma unidade primária p_k :

$$a_{ik} = \alpha_i(p_k) = \int S_i(\lambda)p_k(\lambda)d\lambda, \quad i, k = 1, 2, 3$$

$$\Rightarrow \sum_{k=1}^3 \beta_k a_{ik} = \alpha_i(C) = \int S_i(\lambda)C(\lambda)d\lambda, \quad i = 1, 2, 3$$

Resolvo o sistema de equações e acho as proporções β_k de cada primária:

$$\begin{aligned} \boldsymbol{\beta} &= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} & \boldsymbol{\alpha}(C) &= \begin{bmatrix} \alpha_1(C) \\ \alpha_2(C) \\ \alpha_3(C) \end{bmatrix} & \mathbf{S}(\lambda) &= \begin{bmatrix} S_1(\lambda) \\ S_2(\lambda) \\ S_3(\lambda) \end{bmatrix} & \mathbf{p}(\lambda) &= \begin{bmatrix} p_1(\lambda) \\ p_2(\lambda) \\ p_3(\lambda) \end{bmatrix} \\ \boldsymbol{\beta} &= \left[\int_{\lambda_{min}}^{\lambda_{max}} \mathbf{S}(\lambda)\mathbf{p}^t(\lambda)d\lambda \right]^{-1} \boldsymbol{\alpha}(C) \end{aligned}$$



- Na prática, as fontes primárias são calibradas com relação a um branco de referência com distribuição de energia conhecida $W(\lambda)$.
- Defino $T_k(C) = \frac{\beta_k}{w_k}$, $k = 1, 2, 3 \Rightarrow$ três estímulos da cor C
- Para o branco de referência, $T_k(C) = 1$, $k = 1, 2, 3$.
- Seja uma cor espectral de energia unitária $C(\lambda) = \delta(\lambda - \lambda')$ com comprimento de onda λ' .
- Designo três estímulos correspondentes de $T_k(\lambda') \Rightarrow$ **curvas de casamento espectral**.
- Como $\sum_{k=1}^3 \beta_k a_{ik} = \int S_i(\lambda) C(\lambda) d\lambda$, para uma cor espectral $\delta(\lambda - \lambda')$, $\beta_k = T_k(\lambda') w_k$.

$$\Rightarrow \sum_{k=1}^3 w_k a_{ik} T_k(\lambda') = S_i(\lambda')$$

- Substituindo $S_i(\lambda)$ na expressão acima,

$$w_k T_k(C) = \int C(\lambda) T_k(\lambda) d\lambda, \quad k = 1, 2, 3.$$



Usando a forma matricial,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1(C) \\ \beta_2(C) \\ \beta_3(C) \end{bmatrix} \quad \mathbf{S}(\lambda) = \begin{bmatrix} S_1(\lambda) \\ S_2(\lambda) \\ S_3(\lambda) \end{bmatrix}$$

$$\mathbf{T}(\lambda) = \begin{bmatrix} T_1(\lambda) \\ T_2(\lambda) \\ T_3(\lambda) \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{bmatrix} \quad \mathbf{p}(\lambda) = \begin{bmatrix} p_1(\lambda) \\ p_2(\lambda) \\ p_3(\lambda) \end{bmatrix}$$

$$\mathbf{A}\boldsymbol{\beta} = \int \mathbf{S}(\lambda)C(\lambda)d\lambda \quad (1)$$

$$\mathbf{AWT}(\lambda) = \int \mathbf{S}(\lambda')\delta(\lambda - \lambda')d\lambda' = \mathbf{S}(\lambda) \quad (2)$$

(2) → (1)

$$\mathbf{A}\boldsymbol{\beta} = \int \mathbf{AWT}(\lambda)C(\lambda)d\lambda$$

$$\int \mathbf{T}(\lambda)C(\lambda)d\lambda = \mathbf{W}^{-1}\boldsymbol{\beta} \Rightarrow \frac{\beta_k}{w_k} = T_k(C) = \int T_k(\lambda)C(\lambda)d\lambda$$



Ex: CIE

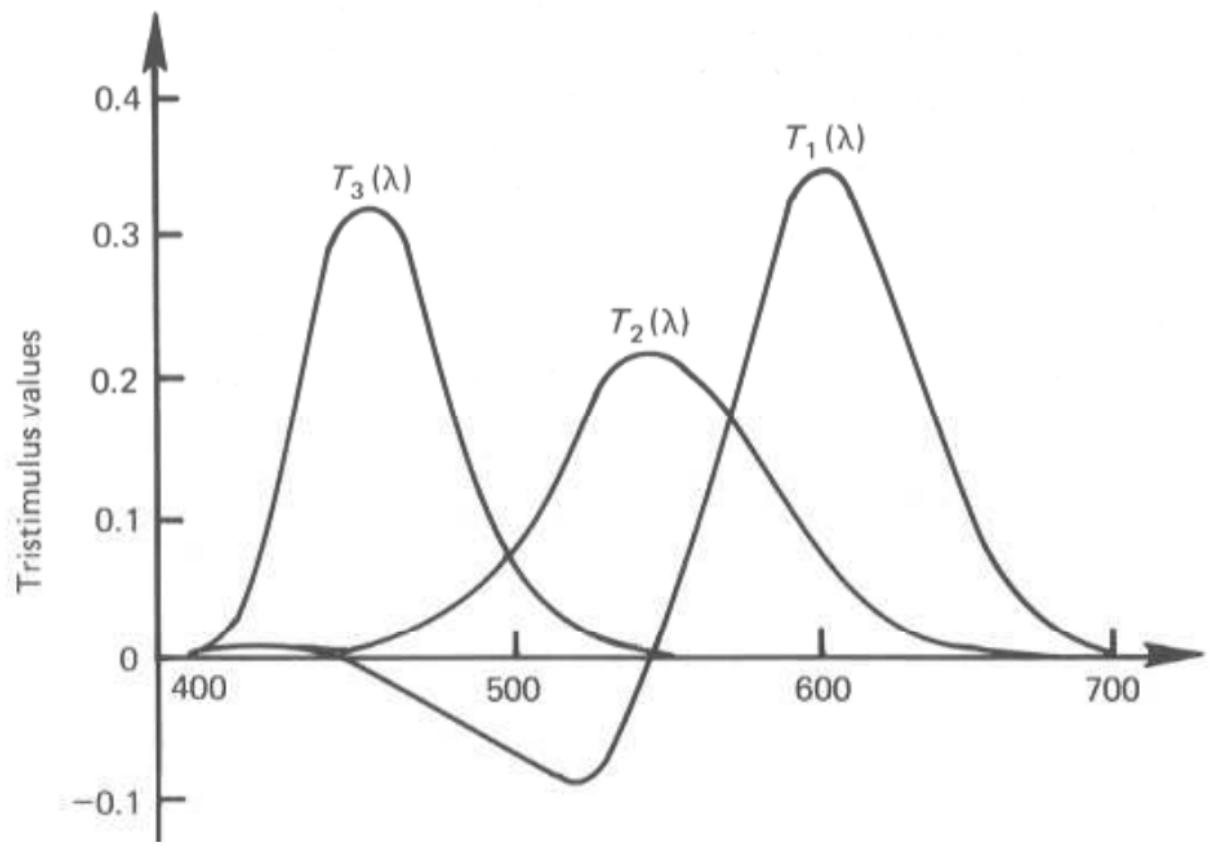
$$p_1 = \delta(\lambda - \lambda_1), \quad \lambda_1 = 700\text{nm}, \text{ vermelho}$$

$$p_2 = \delta(\lambda - \lambda_2), \quad \lambda_2 = 546.1\text{nm}, \text{ verde}$$

$$p_3 = \delta(\lambda - \lambda_3), \quad \lambda_3 = 435.8\text{nm}, \text{ azul}$$

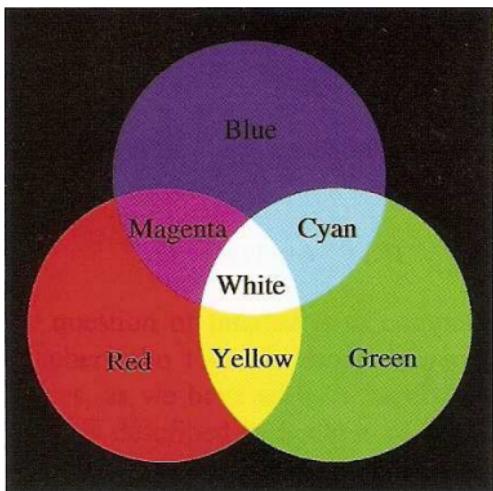
$$a_{ik} = \int S_i(\lambda) p_k(\lambda) d\lambda = S_i(\lambda_k), \quad i, k = 1, 2, 3.$$

Branco CIE: espectro plano $\Rightarrow \alpha_i(w) = \int S_i(\lambda) d\lambda$



Geração de Cores

Aditivo



Corpos Emissores de Luz

Subtrativo



Corpos Absorvedores de Luz

Leis do Casamento de Cores

1) Qualquer cor pode ser reproduzida pela mistura de no máximo três cores.

\exists 3 primárias tal que a matriz $\{a_{ik}\} = \left\{ \int S_i(\lambda)p_k(\lambda)d\lambda \right\}$, com

$$\sum_{k=1}^3 a_{ik}\beta_k = \int S_i(\lambda)C(\lambda)d\lambda, i = 1, 2, 3 \text{ é } \underline{\text{não-singular}}$$

2) A luminância de uma mistura de cores é igual à soma das luminâncias de seus componentes.

$$Y = Y(C) = \int C(\lambda)V(\lambda)d\lambda,$$

onde $V(\lambda)$ é a eficiência luminosa relativa.

$$\text{Luminância da } k\text{-ésima primária } Y(p_k) = \int p_k(\lambda)V(\lambda)d\lambda$$

$$\Rightarrow p_k \text{ com proporção } \beta_k = w_k T_k$$



\Rightarrow contribuição da k -ésima primária = $T_k w_k \int p_k(\lambda) V(\lambda) d\lambda$

Esta lei é equivalente a $Y = \sum_{k=1}^3 T_k \int w_k p_k(\lambda) V(\lambda) d\lambda \triangleq \sum_{k=1}^3 T_k \ell_k$,

onde ℓ_k é o coeficiente de luminosidade da k -ésima primária.

Em geral, $C(\lambda) \neq \sum w_k T_k p_k(\lambda)$

Isto só vai acontecer se $V(\lambda) = \sum_{k=1}^3 \theta_k S_k(\lambda) d\lambda = \boldsymbol{\theta}^t \mathbf{S}(\lambda)$,

o que é razoável, pois são os cones os responsáveis pela percepção das cores e da luminância em altos níveis de iluminação.



Prova:

$$Y = \int C(\lambda) V(\lambda) d\lambda \quad \mathbf{A}\beta = \int \mathbf{S}(\lambda) C(\lambda) d\lambda$$

$$\text{mas } \mathbf{A} = \int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \Rightarrow \int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \beta = \int \mathbf{S}(\lambda) C(\lambda) d\lambda$$

$$\Rightarrow \mathbf{A} = \int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) \beta d\lambda = \int \mathbf{S}(\lambda) C(\lambda) d\lambda \quad (1)$$

$$\Rightarrow Y_k = \int p(\lambda) V(\lambda) d\lambda \Rightarrow \sum \beta_k Y_k = \int \mathbf{p}^t(\lambda) \beta V(\lambda) d\lambda, \quad Y = \int C(\lambda) V(\lambda) d\lambda$$

\Rightarrow Se $V(\lambda) = \sum \theta_k S_k(\lambda)$, vale $Y = \sum Y_k \beta_k$. Caso contrário, não.

\Rightarrow Na prática, vale $Y = \sum Y_k \beta_k$.

Alternativamente, $p(\lambda) = \sum \beta_k p_k(\lambda) = \mathbf{p}^t \beta$,



$$\beta = \left[\int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \right]^{-1} \int \mathbf{S}(\lambda) C(\lambda) d\lambda$$

$$\Rightarrow p(\lambda) = \mathbf{p}^t \left[\int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \right]^{-1} \int \mathbf{S}(\lambda) C(\lambda) d\lambda$$

$$Y(p) = \int \mathbf{p}^t V(\lambda) d\lambda \left[\int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \right]^{-1} \int \mathbf{S}(\lambda) \mathbf{C}(\lambda) d\lambda$$

$$Y(C) = \int C(\lambda) V(\lambda) d\lambda$$

Seja $V(\lambda) = \sum_{k=1}^3 \theta_k S_k(\lambda) d\lambda = \boldsymbol{\theta}^t \mathbf{S}(\lambda)$

$$\Rightarrow Y(C) = \int C(\lambda) \boldsymbol{\theta}^t \mathbf{S}(\lambda) d\lambda = \boldsymbol{\theta}^t \int C(\lambda) \mathbf{S}(\lambda) d\lambda$$

$$\begin{aligned}
 Y(p) &= \int \mathbf{p}^t \theta^t \mathbf{S}(\lambda) d\lambda \left[\int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \right]^{-1} \int \mathbf{S}(\lambda) \mathbf{C}(\lambda) d\lambda \\
 &= \int \theta^t \mathbf{S}(\lambda) \mathbf{p}^t d\lambda \left[\int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \right]^{-1} \int \mathbf{S}(\lambda) \mathbf{C}(\lambda) d\lambda \\
 &= \underbrace{\theta^t \int \mathbf{S}(\lambda) \mathbf{p}^t d\lambda}_{=I} \left[\int \mathbf{S}(\lambda) \mathbf{p}^t(\lambda) d\lambda \right]^{-1} \int \mathbf{S}(\lambda) \mathbf{C}(\lambda) d\lambda \\
 &= \theta^t \int C(\lambda) \mathbf{S}(\lambda) d\lambda \\
 &= Y(C)
 \end{aligned}$$

Se $V(\lambda) = \sum_{k=1}^3 \theta_k S_k(\lambda) \Rightarrow Y(p) = Y(C)$



- 3) O olho humano não consegue distinguir os componentes de uma mistura de cor.
- 4) Um casamento de cores válido para um dado nível de luminância é também válido para uma larga faixa de luminâncias.
- 5) Adição de cores (OBS: $[C] \Rightarrow$ “percepção” da cor C).

$$[C_1] = [C'_1], \quad [C_2] = [C'_2]$$

$$\Rightarrow \alpha_1[C_1] + \alpha_2[C_2] = \alpha_1[C'_1] + \alpha_2[C'_2]$$

6) Subtração de cores.

$$[C_1] + [C_2] = [C'_1] + [C'_2], \quad [C_1] = [C'_1]$$

$$\Rightarrow [C_2] = [C'_2]$$

7) Lei Transitiva.

$$[C_1] = [C_2], \quad [C_2] = [C_3] \Rightarrow [C_1] = [C_3]$$

8) Casamento de Cores.

Casamento “direto”:

$$1. \alpha[C] = \alpha_1[C_1] + \alpha_2[C_2] + \alpha_3[C_3]$$

Casamentos “indiretos”:

$$2. \alpha[C] + \alpha_1[C_1] = \alpha_2[C_2] + \alpha_3[C_3]$$

$$3. \alpha[C] + \alpha_1[C_1] + \alpha_2[C_2] = \alpha_3[C_3]$$

Leis de Grassman \Rightarrow Só não valem para níveis muito baixos ou muito altos de luminância.



Diagrama De Cromaticidade

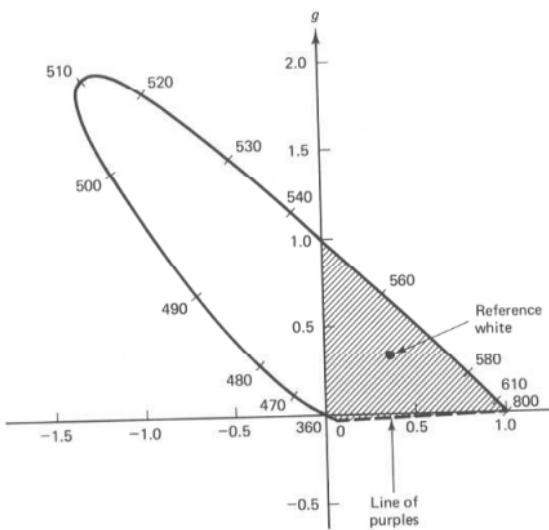
Cromaticidades: $t_k \triangleq \frac{T_k}{T_1 + T_2 + T_3} \quad k = 1, 2, 3.$

Como $t_1 + t_2 + t_3 = 1 \Rightarrow$ projeção de um sólido de cor em um plano

Se t_1 e t_2 são coordenadas de cor (e.g., matiz e saturação):

$\Rightarrow (t_1, t_2, Y)$ no espaço e $Y = \text{cte} \Rightarrow$ plano de crominância.

Diagrama de Cromaticidade: Subespaço de cores no plano de crominâncias (projeção no plano $(T_1 \times T_2)$ da interseção do sólido de cor com o plano $T_1 + T_2 + T_3 = 1$)



- Cores espectrais \Rightarrow cores visíveis “dentro”;
- Linha ligando 360nm (azul) e 700nm (vermelho) \Rightarrow Linha de Púrpuras;
- Triângulo dentro de $(0,0)$, $(0,1)$, e $(1,0)$ \Rightarrow cores reproduzidas com as 3 primárias – “Color Gamut”;
- Branco de referência $(\frac{1}{3}, \frac{1}{3})$ – $t_k = 1, k = 1, 2, 3$.
 - Cores saturadas: longe deste ponto;
 - Cores não-saturadas: perto deste ponto;

Sistemas de Coordenadas de Cores

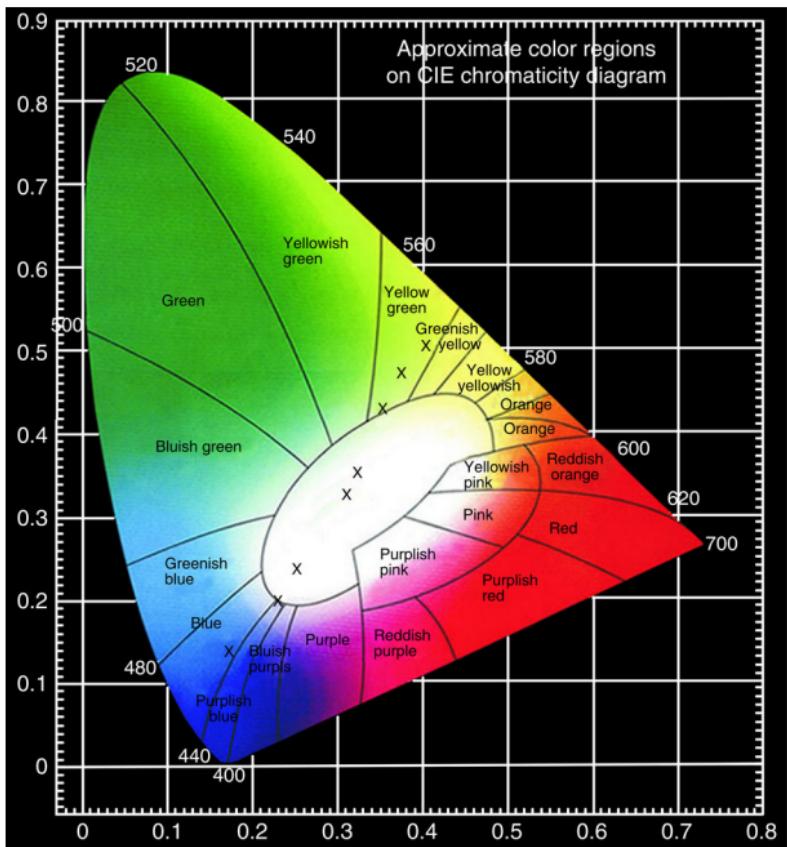
- Primárias espectrais (RGB) da CIE não conseguem reproduzir todas as cores

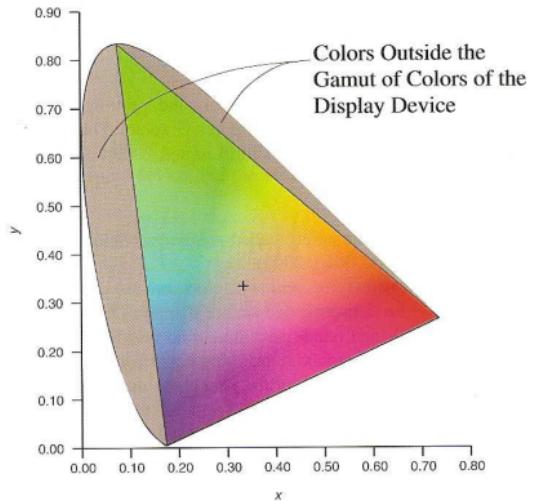
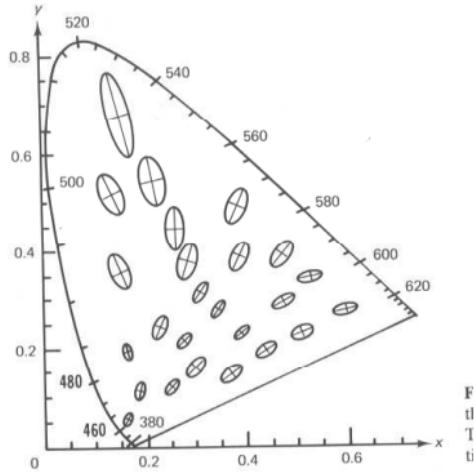
CIE
XYZ

⇒

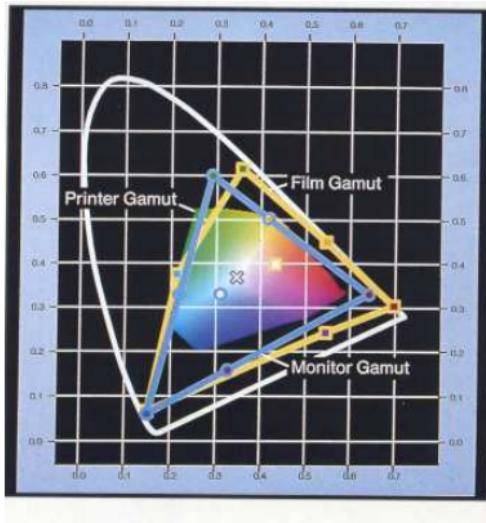
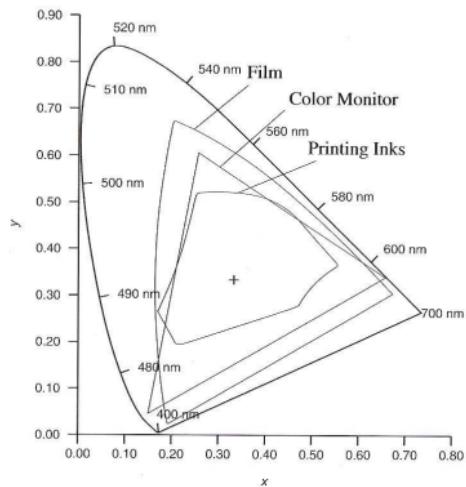
Espaço XYZ

- primárias hipotéticas com valores de três estímulos sempre positivos;
- Tamanhos e excentricidades diferentes (até 20% de variação de excentricidade);

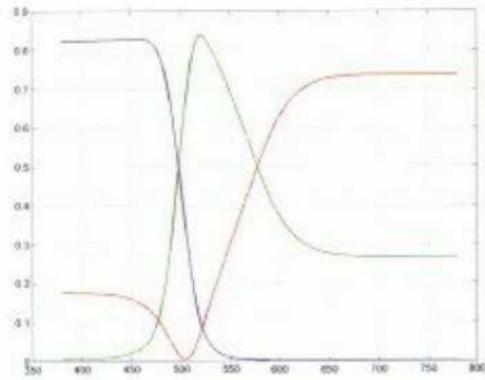
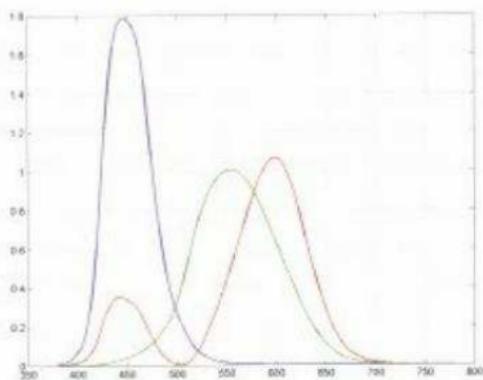


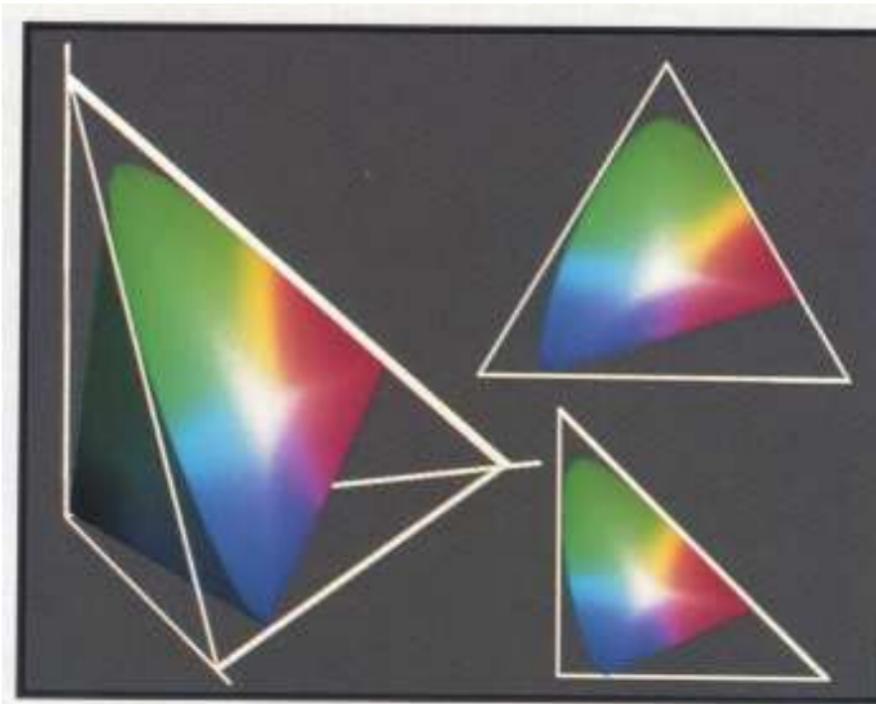


- Exemplos de limitação de faixas cromáticas.



Componentes e coeficientes:





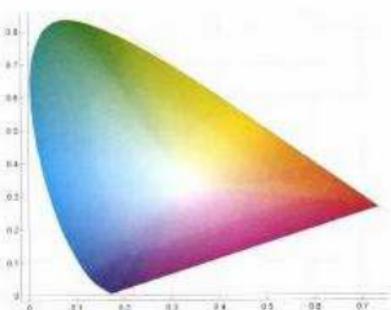


Fig. C15. Diagrama x - y

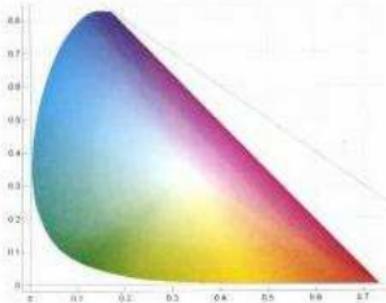


Fig. C16. Diagrama x - z

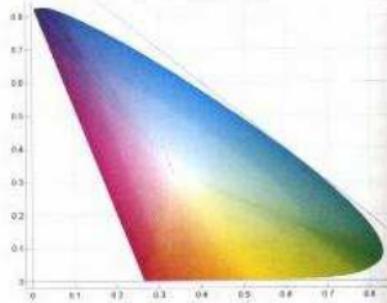
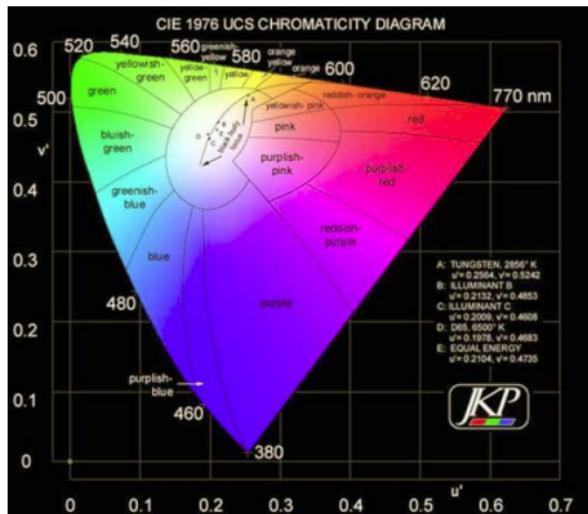


Fig. C17. Diagrama y - z

Espaço UCS

UCS
Uniform Chromaticity Scale $\Rightarrow u, v, f \Rightarrow$ elipses com excentricidade variando em até 2:1



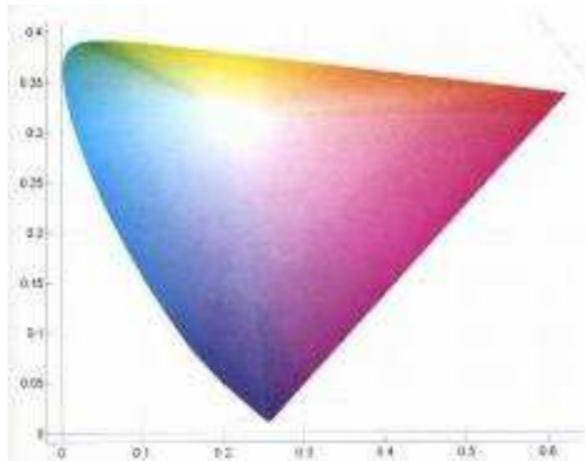


Fig. C28. Diagrama uv UCS 1960

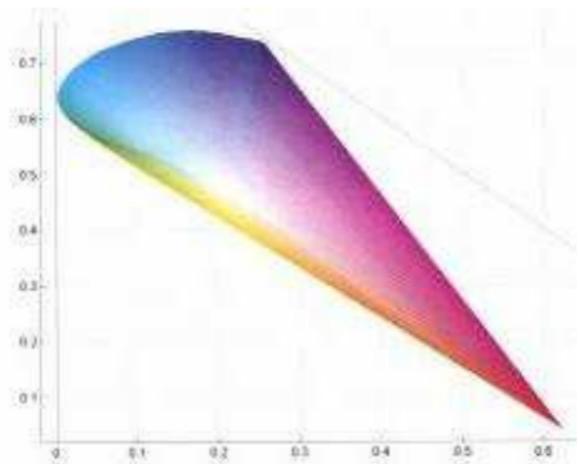


Fig. C29. Diagrama uw UCS 1960

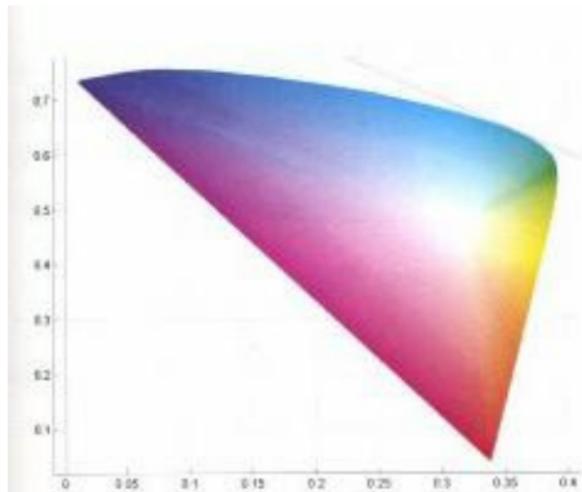


Fig. C30. Diagrama vw UCS 1960

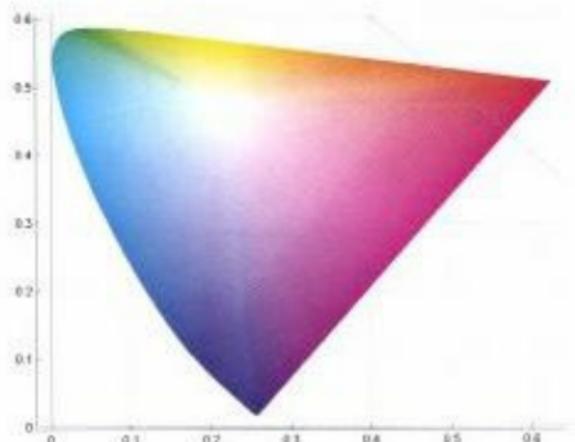


Fig. C31. Diagrama $u'v'$ UCS 1976

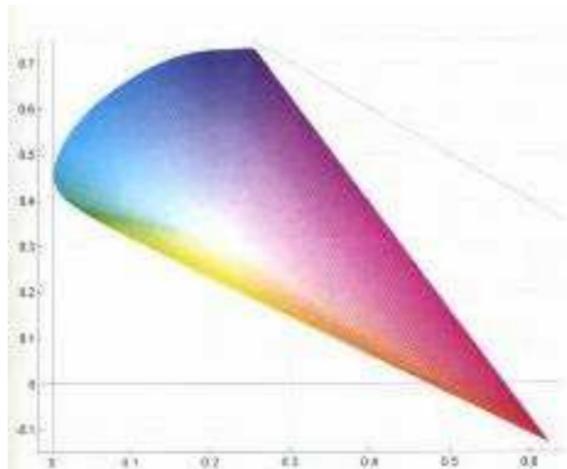


Fig. C32. Diagrama $u'w'$ UCS 1976

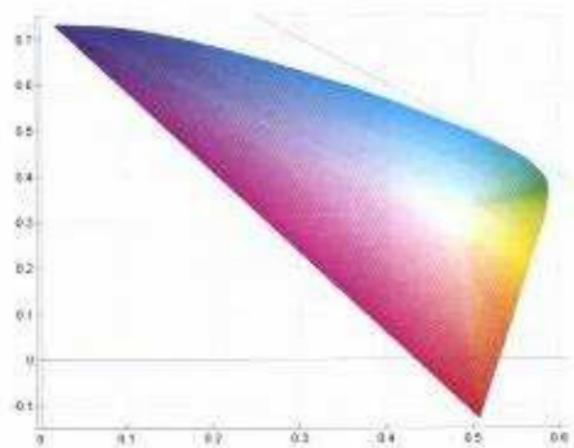
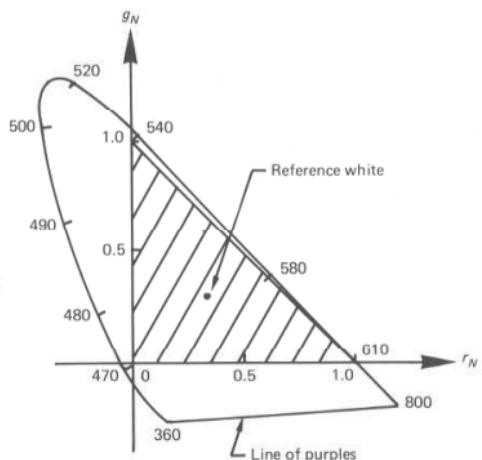
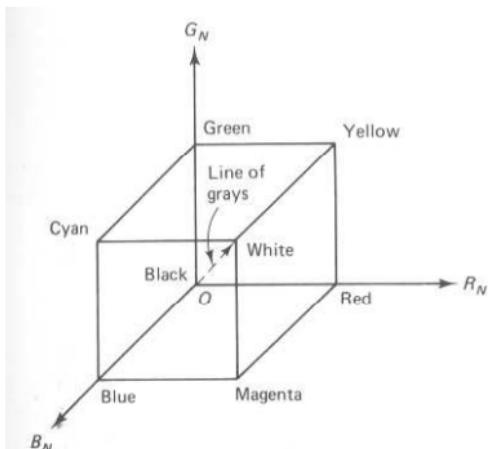


Fig. C33. Diagrama $v'w'$ UCS 1976

Espaço NTSC

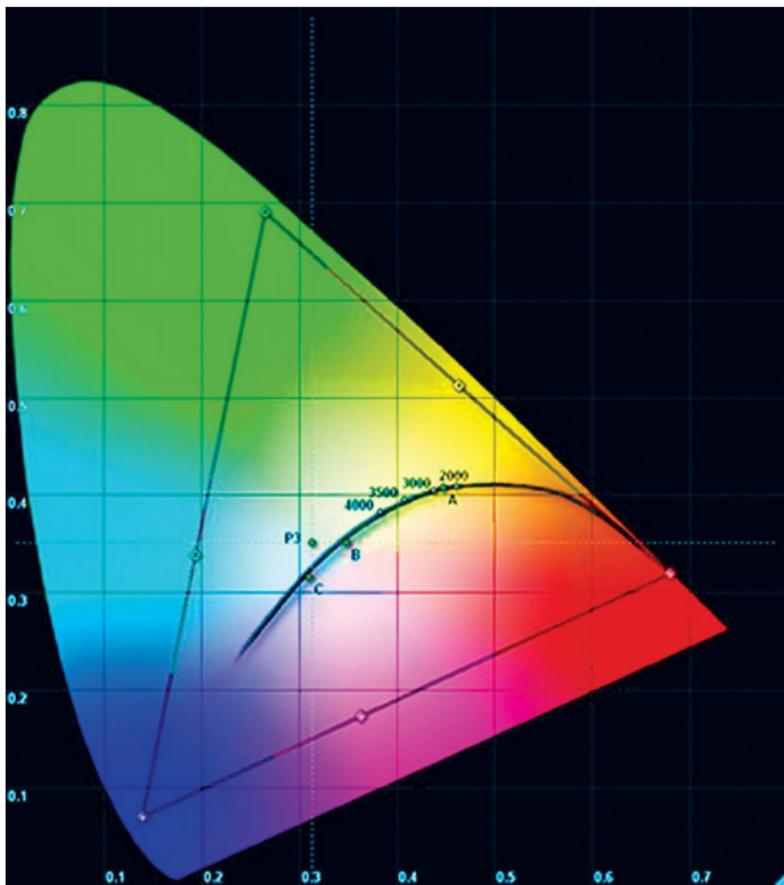
NTSC
(Recepção)

\Rightarrow Primárias $R_N, G_N, B_N \Rightarrow$ Branco de referência: iluminante C

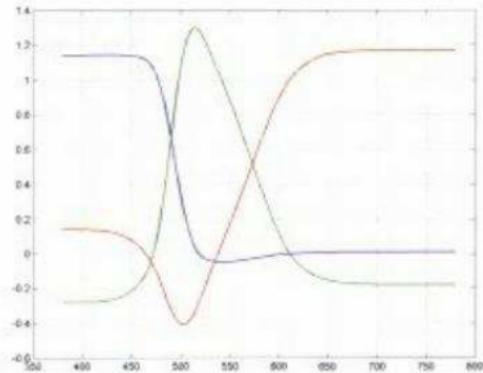
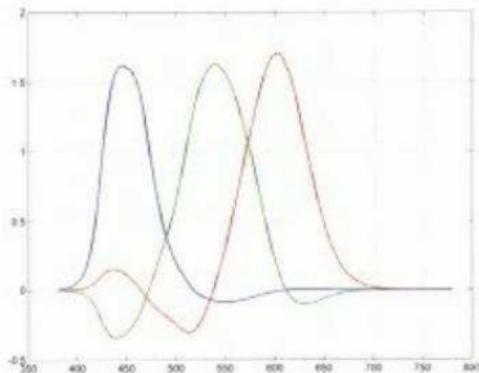


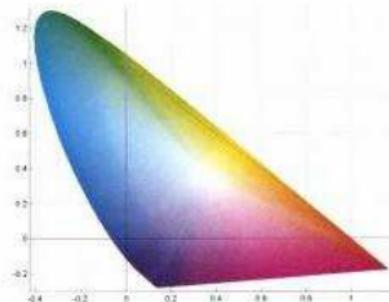
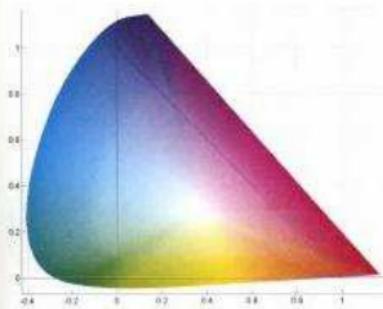
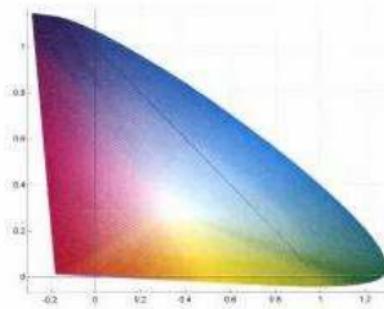
NTSC
(Transmissão)
(SMT - COPPE/UFRJ)

$\Rightarrow Y, I, Q$



Componentes e coeficientes:



Fig. C20. Diagrama $r' - g'$ Fig. C21. Diagrama $r' - b'$ Fig. C22. Diagrama $g' - b'$

Outros Espaços

- UCS modificado com Y transformado com uma raiz cúbica, de modo a ficar equivalente ao contraste;

U^*, V^*, W^* \Rightarrow ● Origem deslocada para o branco de referência;
 ● Bom para medir diferenças de cores – diferença entre duas cores proporcional ao tamanho do segmento de reta as unindo;

S, Θ, W^* \Rightarrow Representação polar de U^*, V^*, W^* ● S : saturação;
 ● Θ : Matiz;
 ● W^* : Contraste;

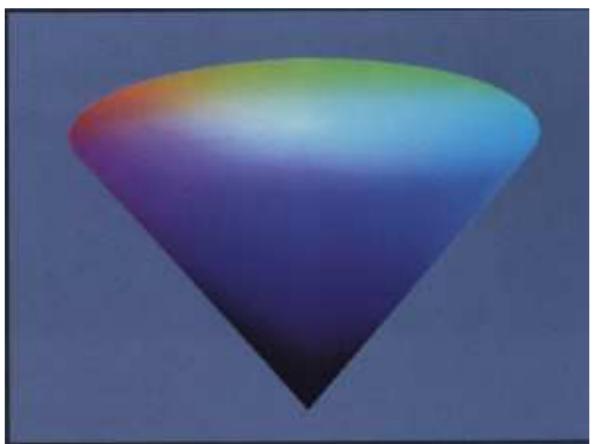
L^*, a^*, b^* \Rightarrow Expressão quantitativa do sistema de Munsell de classificação de cores.



Espaço RGB



Espaço HSV



Color coordinate system	Description
1. C.I.E. spectral primary system: R, G, B	Monochromatic primary sources P_1 , red = 700 nm, P_2 , green = 546.1 nm, P_3 , blue = 435.8 nm. Reference white has flat spectrum and $R = G = B = 1$. See Figs. 3.13 and 3.14 for spectral matching curves and chromaticity diagram.
2. C.I.E. X, Y, Z system Y = luminance	$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$
3. C.I.E. uniform chromaticity scale (UCS) system: u, v, Y u, v = chromaticities Y = luminance U, V, W = tristimulus values corresponding to u, v, w	$u = \frac{4X}{X + 15Y + 3Z} = \frac{4x}{-2x + 12y + 3}$ $v = \frac{6Y}{X + 15Y + 3Z} = \frac{6y}{-2x + 12y + 3}$ $U = \frac{2X}{3}, V = Y, W = \frac{-X + 3Y + Z}{2}$
4. U^*, V^*, W^* system (modified UCS system) Y = luminance [0.01, 1]	$U^* = 13W^*(u - u_0)$ $V^* = 13W^*(v - v_0)$ $W^* = 25(100Y)^{1/3} - 17, 1 \leq 100Y \leq 100$ u_0, v_0 = chromaticities of reference white W^* = contrast or brightness



5. S, θ, W^* system: $S = \text{saturation}$ $\theta = \text{hue}$ $W^* = \text{brightness}$	$S = [(U^*)^2 + (V^*)^2]^{1/2} = 13W^*[(u - u_0)^2 + (v - v_0)^2]^{1/2}$ $\theta = \tan^{-1}\left(\frac{V^*}{U^*}\right) = \tan^{-1}[(v - v_0)/(u - u_0)], 0 \leq \theta \leq 2\pi$
6. NTSC receiver primary system R_N, G_N, B_N	Linear transformation of X, Y, Z . Is based on television phosphor primaries. Reference white is illuminant C for which $R_N = G_N = B_N = 1$. $\begin{bmatrix} R_N \\ G_N \\ B_N \end{bmatrix} = \begin{bmatrix} 1.910 & -0.533 & -0.288 \\ -0.985 & 2.000 & -0.028 \\ 0.058 & -0.118 & 0.896 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$
7. NTSC transmission system: $Y = \text{luminance}$ $I, Q = \text{chrominances}$	$Y = 0.299R_N + 0.587G_N + 0.114B_N$ $I = 0.596R_N - 0.274G_N - 0.322B_N$ $Q = 0.211R_N - 0.523G_N + 0.312B_N$
8. L^*, a^*, b^* system: $L^* = \text{brightness}$ $a^* = \text{red-green content}$ $b^* = \text{yellow-blue content}$	$L^* = 25\left(\frac{100Y}{Y_0}\right)^{1/3} - 16, 1 \leq 100Y \leq 100$ $a^* = 500\left[\left(\frac{X}{X_0}\right)^{1/3} - \left(\frac{Y}{Y_0}\right)^{1/3}\right]$ $b^* = 200\left[\left(\frac{Y}{Y_0}\right)^{1/3} - \left(\frac{Z}{Z_0}\right)^{1/3}\right]$ $X_0, Y_0, Z_0 = \text{tristimulus values of the reference white}$



Medidas de Diferença de Cor

Medidas quantitativas de diferença de cor \Rightarrow codificação, melhoramento e análise

Espaço de cor dos três estímulos \Rightarrow Espaço Reimanniano com métrica $(ds)^2 = \sum_{i=1}^3 \sum_{j=1}^3 c_{ij} dX_i dX_j$

$ds \Rightarrow$ distância entre cores X_i e $X_i + dX_j$

$c_{ij} \Rightarrow$ relacionado à percepção humana.

JND: equivale a um $(ds)^2 = \varepsilon^2 \Rightarrow \varepsilon^2 = \sum_{i=1}^3 \sum_{j=1}^3 c_{ij} dX_i dX_j \Rightarrow$ equação de um elipsóide

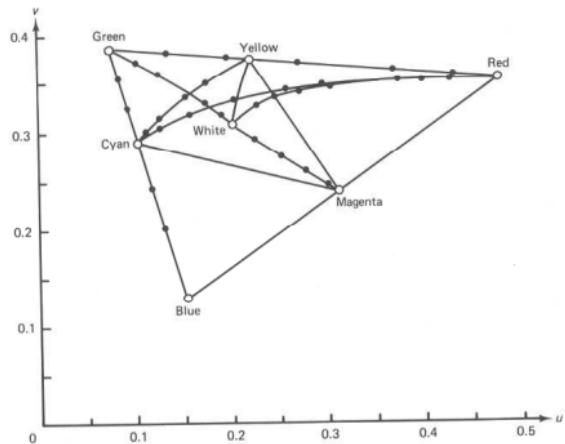
$\Rightarrow c_{ij}$ constantes \Rightarrow espaço Euclidiano.



$\Rightarrow c_{ij}$ variáveis \Rightarrow distância através da geodésica:

$$d(C_1, C_2) \triangleq \mathcal{C} \int_{C_1(X_i)}^{C_2(X_i)} dS \quad \mathcal{C} \quad \text{tal que } d(C_1, C_2) \text{ é mínimo}$$

Geodésicas projetadas no plano UCS das primárias do NTSC:



Procedimento complexo

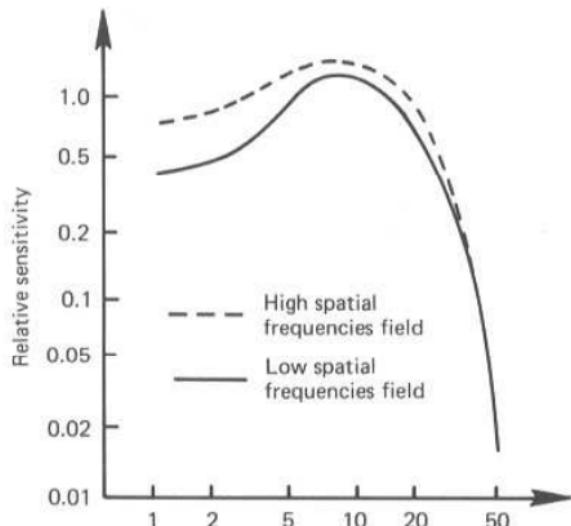


Propriedades Temporais da Visão

Lei de Bloch: “ Flashes de luz de mesma energia e durações diferentes são indistinguíveis abaixo de uma dada duração”

- tipicamente 30ms
- Quanto mais escuro, maior este tempo (ex: cinema × televisão)

Critical Fusion Frequency (CFF): Flashes periódicos são indistinguíveis entre si se $f > \text{CFF}$.



⇒ O olho é mais sensível a flicker em altas frequências espaciais.