

Quantização Vetorial Classificada Adaptada Perceptivamente

Cláudio A. Fleury

Departamento de Engenharia, UCG,
Av. Universitária, 1.440, Setor Universitário, Área III, 74605-010, Goiânia - GO
E-mail: kaw@ucg.br

Gilberto A. Carrijo

Faculdade de Engenharia Elétrica, UFU,
Av. João Naves de Ávila, 2160, Bloco 3N, Campus Santa Mônica, 38400-902, Uberlândia - MG
E-mail: gilberto@ufu.br

Resumo – Uma nova técnica de codificação digital de imagens, baseada em quantização vetorial adaptada perceptivamente é apresentada. Ela explora as deficiências visuais humana e a importância da atração visual das regiões de interesse em uma imagem estática. A técnica é fundamentada em um modelo de percepção visual. A partir de simulações mostrou-se a eficiência desta nova abordagem, principalmente em termos qualitativos.

Palavras Chaves: *Codificação de Imagem, Percepção Visual, Quantização Vetorial.*

Abstract – A new approach on digital image coding using perceptual adaptive vector quantization is presented. It explores the eyes deficiencies and the visual attention atractors importance of each region of interest on static images. The technique is based on a computacional perceptual model. Some preliminaris simulations were developed and showed the efficiency of this new approach, mainly in qualitative aspects.

Keywords – *Image Coding, Vector Quantization, Visual Perception.*

I. INTRODUÇÃO

Definitivamente a informação baseada em imagens estáticas e dinâmicas (vídeos) passou a fazer parte do mundo globalizado, promovido em grande parte, pela disponibilidade, nos últimos anos, de acesso aos serviços da Internet.

Por serem grandes consumidoras de recursos das redes de comunicações, as imagens precisam ser codificadas de modo eficiente para que se adequem às limitações dos canais de transmissão existentes, ou à demanda crescente de transmissão e armazenamento no formato digital.

As buscas por novas técnicas de codificação fonte tem-se intensificado, e a Quantização Vetorial de aliada à adaptação, tem sido uma opção de pesquisa científica.

A codificação perceptiva apresenta como principal vantagem o alto fator de compressão, pois ignorando as informações não percebidas pelos elementos sensoriais humano, a taxa de codificação

pode ser diminuída a valores baixos sem perda significativa de qualidade.

Denomina-se, então, o modelo perceptivo aplicado à quantização vetorial, de técnica de codificação fonte: *Quantização Vetorial Adaptada Perceptivamente (QV-AP)*, a qual será apresentada neste trabalho.

Na Seção II descreve-se as características básicas da *Quantização Vetorial*, e na Seção III discute-se a *Quantização Vetorial Adaptativa*, e na Seção IV apresenta-se o *Modelo de Percepção Visual* adotado na técnica proposta, a QV-AP, mostrada na Seção V. A Seção VI encerra o artigo com algumas conclusões sobre a nova técnica apresentada.

II. QUANTIZAÇÃO VETORIAL (QV)

A aplicação da QV a uma imagem pode ser resumida da seguinte forma: supondo uma imagem particionada em células ou blocos de $p \times p$ pixels, cada célula é considerada como um vetor de dimensão p e é codificada pela busca de um vetor representativo (aquele que mais se aproxima da célula) em um conjunto de vetores (livro-código ou *codebook*). Uma palavra binária identificando o vetor representativo selecionado é atribuída como sendo a palavra-código (*codeword*) que descreve a célula original. O decodificador usa esta palavra-código para endereçar o livro-código (cópia fiel do existente no codificador). Cada entrada do livro-código contém uma representação digital precisa de um dos N vetores representativos. O projeto do livro-código é baseado em uma técnica de agrupamento para projeto de quantizadores vetoriais (algoritmo LBG) [6].

A Quantização Vetorial (QV) também pode ser considerada como uma extensão da quantização escalar para espaços vetoriais, definida como um mapeamento M , de um espaço vetorial de dimensão K , das amostras da entrada para um conjunto finito de vetores de reconstrução, vetores-código ou palavras-código. O conjunto de vetores de reconstrução é o livro-código (*codebook*). O espaço vetorial de entrada de dimensão N é dividido em um conjunto C de N regiões \mathfrak{R}_i de dimensão K , chamadas de partições ou células, e um vetor de

reconstrução y_i é associado a cada uma das \mathcal{R}_i regiões. As partições são delimitadas por hiperplanos de dimensões $K-1$, e todos os vetores de entrada na partição \mathcal{R}_i são mapeados para a palavra código y_i

$$M: \mathcal{R}^k \rightarrow C \quad (1)$$

onde $C = \{y_1, \dots, y_N\}$, $y_i \in \mathcal{R}^k$,
e $\mathcal{R}_i = \{x \in \mathcal{R}^k: M(x) = y_i\}$.

Verifica-se uma nítida vantagem da quantização vetorial sobre a escalar: diferentemente dos quantizadores escalares, os vetoriais não estão restritos às taxas de bits inteiras, isto é possível porque um VQ quantifica amostras da entrada a uma taxa de bits fracionada. Isto é útil por duas razões. Primeira, a alocação inteira de bits pode levar um quantizador a performances sub-ótimas. Segunda: torna-se possível alcançar taxas de bit abaixo de 1 bit/amostra (bits per sample, bps), o que é impossível de se obter com um único quantizador escalar [4].

III. QUANTIZAÇÃO VETORIAL ADAPTATIVA

As técnicas de codificação com taxa constante de bits são mais usuais e comumente encontradas nas aplicações didáticas, por serem de implementação mais fácil; enquanto que as técnicas de codificação com taxa variável de bits são mais comumente encontradas nas aplicações práticas, por serem de melhor resultado, embora, sejam de implementação mais difícil e complexa [3].

A QV ordinária prevê a extração seqüencial de vetores do sinal original e a codificação individual de tais vetores. Em geral estes vetores não são distribuídos identicamente e nem são independentes. Na busca de um maior conhecimento das características dos sinais, várias técnicas tem sido aplicadas, dentre elas o conhecimento estatístico dos vetores (distribuições condicionais de frequência) ou a determinação da porção essencial da informação (percebida pelo cérebro) existente no sinal real.

A performance da codificação pode ser melhorada se o quantizador de alguma forma adapta-se no tempo, ou no espaço, às características (estatísticas ou perceptivas) da fonte vetorial [4].

Um quantizador vetorial é considerado adaptativo se o seu livro de códigos (codebook) ou a sua regra de codificação é mudada no tempo de modo a tirar proveito da informação explorada acerca da vizinhança do vetor a ser codificado ou da informação realmente percebida pelo cérebro humano. A palavra “adaptativa” é usualmente associada aos sistemas em que o codebook muda lentamente em relação à taxa de vetores ao invés de mudar substancialmente para cada vetor sucessivo [4].

IV. CODIFICAÇÃO PERCEPTIVA

IV.a Modelos

A Codificação Perceptiva prevê a incorporação de modelos de visão, parciais ou completos, em algoritmos de compressão de imagens. Os modelos que fornecem informação localizada (espacial e/ou temporal) sobre as distorções ou sobre o conteúdo das cenas, são mais adequados à adaptação dos algoritmos de codificação [7].

Depois da difusão dos padrões de compressão baseados na técnica DCT, qualquer técnica nova a ser adotada tem que provar sua superioridade em relação aos padrões correntes por uma considerável quantia. Por isto, as pesquisas caminham atualmente em direção à sintonia das técnicas existentes ao invés da busca de novos paradigmas.

Embora as técnicas convencionais de compressão modelem, com pouca precisão, algumas propriedades do Sistema Visual Humano (SVH), elas têm sido otimizadas geralmente em termos de métricas objetivas tais como o Erro Quadrático Médio (EQM). Este contra-senso explica-se pela facilidade no uso das métricas objetivas e pela complexidade e pouca disponibilidade de métricas subjetivas.

A percepção geral da qualidade de uma imagem por um observador humano é fortemente dependente das áreas com as piores distorções existentes em toda a cena, isto é, a degradação gerada pelos codificadores em algumas regiões, mesmo que de pequena envergadura, provocam um decréscimo significativo na avaliação subjetiva da imagem (baixa qualidade).

As principais características perceptivas inerentes aos Padrões de Compressão existentes atualmente são as seguintes: matriz de quantização, tamanho do bloco DCT, taxa de quadros, sub-amostragem da crominância, quantização espacialmente adaptativa, mudanças de cena.

As duas áreas mais comumente estudadas para a aplicação dos algoritmos da codificação perceptiva são o projeto da matriz de quantização, e o controle da quantização espacialmente adaptada, e nesta última categoria, é que se insere a técnica proposta neste trabalho.

IV.b Técnicas de Codificação Perceptiva

As técnicas de codificação com aplicação de conceitos perceptivos podem ser agrupadas em três níveis, considerando as diversas etapas de tratamento pelas quais passam imagens a serem codificadas em um determinado sistema de processamento digital de imagem: pré-processamento, processamento perceptivo, e pós-processamento.

- *Pré-processamento Perceptivo*

Estas técnicas promovem a remoção de qualquer conteúdo da cena que não seja perceptível ou desnecessário aos propósitos da aplicação. A

remoção de ruídos de alta frequência tem sido uma prática comum, mas cuidados têm sido tomados para não embaçar as bordas dos objetos da cena. A redução significativa da largura de banda da imagem pela diminuição da resolução nas áreas que não são Regiões de Interesse (RI) tem justificado o uso deste tipo de processamento preliminar.

- *Quantização Perceptiva*

Os codificadores baseados em Transformação Discreta do Co-seno (DCT) usam a codificação perceptiva no projeto da matriz de quantização para modelar a Função de Sensibilidade ao Contraste (FSC) do olho humano, e no controle espacial adaptativo da quantização (para modelar o mascaramento espacial). O padrão MPEG sugere um processo incipiente de quantização espacialmente adaptativa no algoritmo TM5 (H.26x no algoritmo RM8). Ele não modela os efeitos de mascaramento. Uma abordagem usual tem sido classificar as cenas em diversas categorias: planas (flats), bordas (edge) e texturas, e então atribuir quantizações de resolução maior para as áreas de bordas. Outra abordagem, computacionalmente mais cara, tem sido buscar iterativamente o ajuste da quantização espacial adaptativa após a avaliação da imagem por um modelo de visão inicial.

- *Pós-processamento Perceptivo*

Após a aplicação das técnicas de um, ou de ambos tipos citados, pode-se ainda tentar remover ou reduzir quaisquer erros de codificação notáveis perceptivamente, através da mudança da própria imagem codificada.

IV. c Modelo de Percepção Visual Adotado

Baseado no mascaramento espacial e nas propriedades de atenção visual do SVH [7], o quantizador proposto substitui a estratégia de quantização adaptativa usada pelo TM5. A estratégia de mascaramento espacial é implementada pela parte superior do diagrama mostrado na *Figura 1*.

A imagem é dividida em blocos de 8×8 pixels, e cada bloco é classificado em *plano*, *borda* ou *textura*, usando a mesma técnica do modelo de visão inicial apresentado em [7]. A atividade é medida em cada bloco através do cálculo da variância (como feito em TM5), e tem seu valor ajustado de acordo com a classificação de cada bloco.

A graduação do parâmetro que avalia e agrega os fatores de importância é feita pela atividade ajustada da seguinte forma:

$$ativ'_j = \begin{cases} \min(ativ_j, ativ_{lim}) & \text{p/ regiões planas} \\ ativ_{lim} \left(\frac{ativ_j}{ativ_{lim}} \right)^\epsilon & \text{p/ outras regiões} \end{cases} \quad (1)$$

onde:

$ativ'_j$ - atividade ajustada para o bloco j ;
 $ativ_j$ - variância do bloco j ;
 $ativ_{lim}$ - variância do limiar de visib.(usado 5,0);
 ϵ - constante: 0,7 para bordas; 1,0 para texturas.

A atividade normalizada está restrita à faixa [0,5 – 2,0], o que assegura um erro de quantização mínimo nas regiões planas e uma maior quantização com a atividade gradual nas bordas.

$$Nativ_j = \frac{2ativ'_j + ativ_m}{ativ'_j + 2ativ_m} \quad (2)$$

onde:

$Nativ_j$ - atividade normalizada para o bloco j
 $ativ_m$ - valor médio da $ativ'_j$ do quadro anterior

V. QUANTIZAÇÃO VETORIAL ADAPTADA PERCEPTIVAMENTE (QV-AP)

Visando obter uma maior performance na codificação de imagens em baixas taxas de bits por pixel é que buscou-se a Quantização Vetorial como técnica fundamental. Esta técnica apresenta algumas deficiências em relação à codificação de regiões da imagem com bordas. Estas deficiências são corrigidas ou amenizadas por várias técnicas variantes da Quantização Vetorial Ordinária (QVO), tais como a Quantização Vetorial Classificada (QVC) [9], ou Quantização Vetorial Estruturada em Árvore [1].

O uso da classificação na QVC permite a eliminação do problema mais sério da quantização vetorial no que diz respeito à percepção visual em relação ao seu resultado, que é a degradação das bordas das imagens, pois a técnica prevê a criação de um ou vários codebooks específicos para blocos deste tipo, ou seja, para blocos com bordas. E ainda como vantagem adicional, esta técnica diminui a complexidade computacional do QVO, pois diminui o universo de pesquisa dos codevectors, restringindo-a apenas ao codebook classificado [9] ao qual pertence o bloco em codificação.

A partir da identificação dos pontos de maior importância da imagem em termos de atração visual, adota-se a geração de regiões circulares, concêntricas ou não, que englobam os pontos mais expressivos indicados pelo mapa de importância para a imagem a ser codificada. A escolha da forma geométrica circular para as regiões de importância foi devida, principalmente, à facilidade de sua descrição, que necessita apenas de um par de coordenadas para o centro do círculo e um número inteiro para representar seu raio, todos os valores são considerados em unidades de pixels.

Para a codificação das imagens que apresentam um único local de convergência da atenção visual, pode-se usar regiões circulares concêntricas, as quais são descritas com o envio ao decodificador de algumas poucas informações adicionais em relação

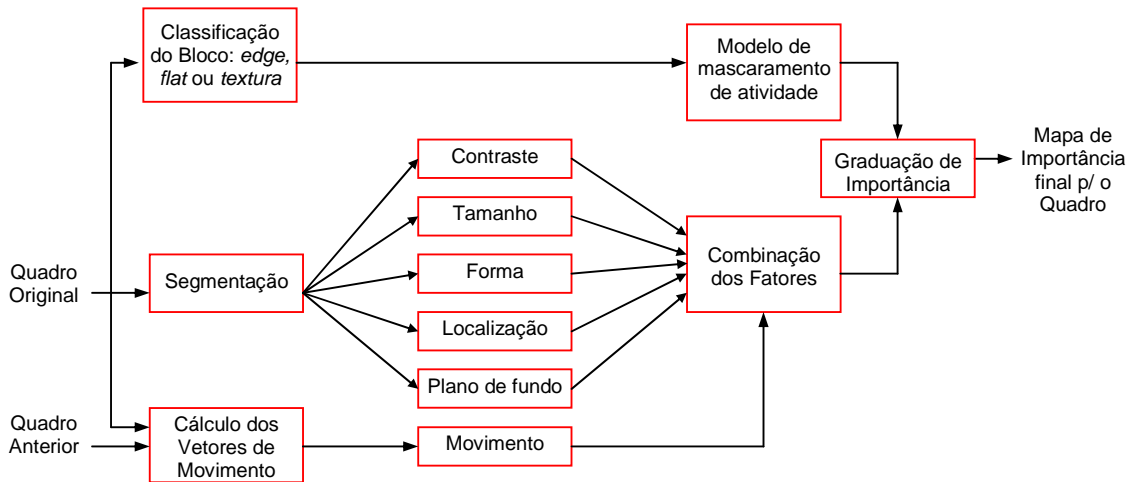


Figura 1 - Diagrama de Blocos do Modelo de Percepção Visual

àquelas da codificação padrão prevista na QVC: o raio de cada uma das regiões ou estipula um incremento fixo ao raio do primeiro círculo e a quantidade de círculos.

Definidas as regiões de maior atração da atenção visual humana, a codificação e a decodificação de cada bloco de pixels da imagem fica sujeita à localização espacial do mesmo para a determinação da resolução a ser empregada, ou seja, da quantidade de níveis do codebook para cada região espacial. A região mais interna (com centro no foco principal da atenção da visão humana) tem seus blocos codificados com base no codebook maior, e a mais externa (região complementar, no caso do uso de apenas duas regiões), com base no codebook menor [3].

Os resultados experimentais preliminares mostram projetos baseados em duas regiões, sendo que em um trabalho investigativo posterior pretende-se avaliar se três ou mais níveis de importância (regiões) redundariam em performances subjetivas melhores para codificação em taxas baixas (em torno de 0.4 bpp).

É razoável que algumas imagens apresentem mais de um foco com forte atração visual, e para estes casos pode-se adotar regiões circulares não concêntricas. Novos estudos acerca do raio destas regiões precisam ser realizados.

Nas Figura 2 e Figura 3 têm-se, em diagramas de blocos, o funcionamento de um par de codificador e decodificador que implementa a nova técnica proposta. Considerando X como sendo um bloco de pixels, i como o índice ou palavra código (codeword), e D como a descrição das regiões circulares mais importantes em termos de atração visual, determinadas dinamicamente pelo próprio codificador a partir de toda a imagem I .

De posse da imagem original I , o gerador de Mapas de Importância (M.I.) estabelece as regiões de

maior importância da imagem, descrevendo-as através de seus centros e raios, sintetizados na descrição D a ser enviada ao decodificador uma única vez. Na seqüência, cada bloco da imagem original é avaliado quanto à sua localização dentro da imagem com relação à sua inclusão, ou não, em uma região de alta importância. Deste processamento tem-se como resultado a indicação da região R_n a que pertence o bloco D , que é codificado com base em um dos k codebooks gerados para cada região.

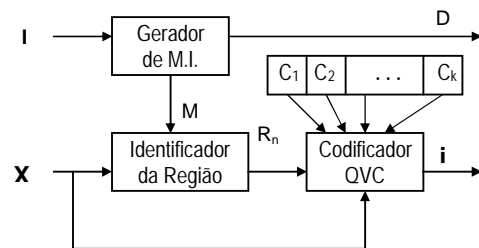


Figura 2 - Codificador QV-TVAP

A decodificação é mais direta. Com a recepção da descrição D das regiões de importância, a posição relativa de cada bloco X dentro da imagem original (a seqüência de codificação deve ser constante e

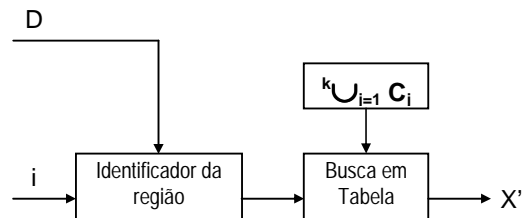


Figura 3 - Decodificador QVC

mantida inalterada, indicando uma ordem estabelecida no procedimento do código), simbolizado pelo seu índice (codeword). Como os índices recebidos são contínuos, pode-se individualizar cada uma das regiões e buscar coerentemente o codevector no codebook respectivo, a partir da codeword (índice) recebida. A *Figura 4* ilustra a identificação das regiões de maior importância em termos de atração da atenção visual para a imagem da *Lena*, utilizando-se regiões circulares concêntricas, admitindo como sendo o local de maior atração visual (foco) apenas um único ponto da imagem, e na *Figura 5*, tem-se a identificação de dois pontos de forte atração visual, gerando regiões circulares não-concêntricas. Enfatiza-se que a localização dos pontos de maior atração da atenção visual foram estipulados pelo modelo de Osberger [7] de forma automática e sem qualquer tipo de sintonia manual necessária.

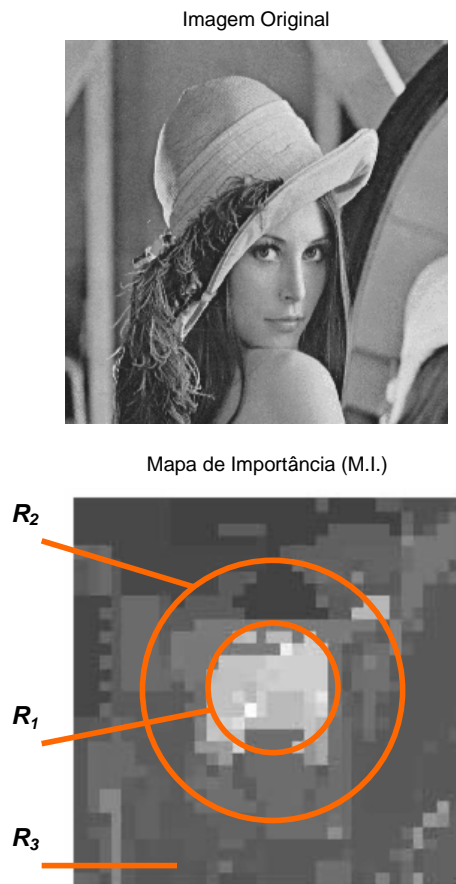


Figura 4 – Regiões Circulares Concêntricas p/ Lena

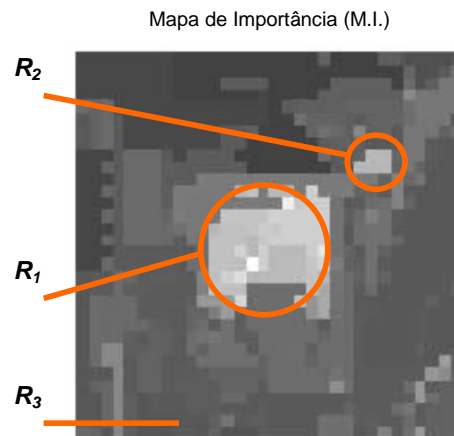


Figura 5 – Duas Regiões Circulares Não-Concêntricas para Lena

VI. CONCLUSÕES

A busca por técnicas que possibilitem a codificação fonte de imagens em baixas taxas de bits por pixel ($< 1,0$ bpp) tem impulsionado os estudos da percepção visual e a volta às técnicas que redundem em alta taxa de compressão, como a QV. Neste trabalho apresentou-se uma técnica que usa a atração visual como elemento de identificação da área de maior interesse em uma imagem e que usa a QV de taxa variável, com codebook de maior tamanho para estas áreas de interesse, e de menor tamanho para as demais. Subjetivamente pode-se afirmar, com base em alguns testes realizados, que pode-se manter a qualidade da imagem fazendo uso de menores taxas de codificação. Novos estudos devem medir mais precisamente e aprimorar a qualidade das imagens resultantes da aplicação da QV-TVAP.

REFERÊNCIAS

- [1] P. C. Cosman, K. O. Perlmutter, S. M. Perlmutter, R. A. Olshen, e R. M. Gray., "Training Sequence Size and Vector Quantization Performance", IEEE Trans., Doc. 1058-6393/91,1991, pp. 434-438.
- [2] P. M. Embree, e D. Danieli, "C++ Algorithms for Digital Signal Processing", Prentice Hall PTR, Inc., 2nd edition, 1999, p.579.
- [3] C. A. Fleury, "Quantização Vetorial de Taxa Variável Adaptada Perceptivamente Aplicada a Imagens Estáticas", Tese de Doutorado, Univ. Federal de Uberlândia, Brasil, Outubro-2003.
- [4] A. Gersho, e R. M. Gray, "Vector Quantization and Signal Compression", Kluwer Academic Publishers, 7th printing, USA, 1999.

- [5] R. C. Gonzalez, e R. E. Woods, “Digital Image Processing”, Addison-Wesley Publishing Company, Inc., 1993, p.716.
- [6] Y. Linde, A. Buzo, e R. Gray, “An Algorithm for Vector Quantizer Design”, IEEE Trans. on Communications, Vol. 28, Jan - 1980.
- [7] W. Osberger, “Perceptual Vision Models For Picture Quality Assessment And Compression Applications”, Tese de Doutorado, Queensland Univ. of Technology, Março-1999, Australia.
- [8] A. Oppenheim, e R. W. Schaffer, “Digital Signal Processing”, Prentice-Hall, Inc., 1975, p.585.
- [9] B. Ramamurthi, e A. Gersho, “Classified Vector Quantization of Images”, IEEE Trans. on Comm., Vol. 34, No. 11, Nov-1986, pp. 1105 – 1115.
- [10] R. J. Schalkoff, “Digital Image Processing and Computer Vision”, John Wiley & Sons, Inc., 1989, p.489.