

Codificadores de Vídeo Usando Ridgelets e Planos de Bits Generalizados

Leonardo Hidd Fonteles, Rogério Caetano, Eduardo A. B. da Silva

Resumo—Recentemente foi proposto um novo modelo de decomposição de funções usando planos de bits generalizados onde uma função é mapeada diretamente em um conjunto de índices. Esta decomposição foi também utilizada com sucesso na substituição das operações de decomposição/quantização do codificador de vídeo proposto por Neff e Zakhor, baseado em matching pursuits. Neste artigo nós investigamos bons dicionários para tais decomposições. Estes dicionários são baseados nas funções ridgelets, propostas recentemente por Candès e Donoho. Mostramos que melhoras na relação taxa-distorção podem ser alcançadas através de combinações de produtos entre ridgelets escaladas e rotacionadas apropriadamente.

Palavras-Chave—Matching Pursuits, Ridgelets, Planos de bits Generalizados, Codificação de Vídeo.

Abstract—It has recently been proposed a novel decomposition of functions using generalized bit-planes, where a function is mapped into a set of indexes. It also has been successfully used for replacing the decomposition/quantization operations in Neff and Zakhor's matching pursuits video encoder. In this paper we investigate good codebooks for such decompositions whose construction is based on ridgelets, proposed by Candès and Donoho. We show that by combining products of properly scaled and rotated ridgelets one can obtain codebooks which provide improved rate-distortion performances.

Keywords—Matching Pursuits, Ridgelets, Generalized Bitplanes, Video Coding.

I. INTRODUÇÃO

Em [1] foi proposto um algoritmo, chamado Matching Pursuits (MP), que decompõe um dado sinal em um conjunto redundante de funções. Este algoritmo foi implementado com sucesso no contexto de codificação de vídeo em [2], tornando-se uma boa alternativa aos codificadores de vídeo baseados na Transformada Cosseno Discreta (DCT). Ele reduz sensivelmente os efeitos de blocos e produz melhoras concretas em relação aos valores de PSNR (Peak Signal-to-Noise Ratio).

O matching pursuits decompõe um sinal $\mathbf{x} \in \mathbb{R}^N$ fazendo uso de um dicionário supercompleto

Leonardo H. Fonteles, PEE/COPPE/Universidade Federal do Rio de Janeiro, leonardo@lps.ufrj.br. Rogério Caetano, Fundação Desembargador Paulo Feitoza e PEE/COPPE/Universidade Federal do Rio de Janeiro, rcaetano@fpf.br. Eduardo A. B. da Silva, PEE/COPPE e DEL/Poli/Universidade Federal do Rio de Janeiro, eduardo@lps.ufrj.br. Este trabalho foi parcialmente financiado pela CAPES.

$\mathcal{C} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M\}$, $\|\mathbf{v}_i\| = 1, \forall i$, como segue [2]:

$$\mathbf{x} \approx \mathbf{x}^{(Q)} = \sum_{n=1}^Q p_n \mathbf{v}_{\gamma_n} \quad (1)$$

Para que esta decomposição seja realizada com sucesso é utilizado um algoritmo voraz. Primeiro, é encontrado o vetor \mathbf{v}_{γ_n} mais próximo do sinal \mathbf{x} que é então projetado em \mathbf{v}_{γ_n} , gerando a projeção p_1 . Em seguida, é calculado o resíduo $\mathbf{r}_1 = \mathbf{x} - p_1 \mathbf{v}_{\gamma_1}$. Este processo é repetido inúmeras vezes até que seja atingido um número de passos Q pré-definido de acordo com o critério taxa-distorção utilizado [1]. Logo após, os coeficientes p_n são quantizados e codificados juntamente com os índices γ_n .

Recentemente foi proposta uma outra decomposição de sinais baseada em matching pursuits com planos de bits generalizados, chamada de *Matching Pursuits with Generalized Bit-Planes* (MPGBP) [3], [4], [5]. Ela aproxima, em P passos, um sinal $\mathbf{x} \in \mathbb{R}^N$ como a seguir [2]:

$$\mathbf{x} \approx \mathbf{x}^{(P)} = \sum_{m=1}^P \alpha^{k_m} \mathbf{v}_{r_m} \quad (2)$$

onde $0 < \alpha < 1$.

Tal decomposição mapeia o sinal \mathbf{x} em uma seqüência de índices (k_m, r_m) de modo a realizar a decomposição e a quantização ao mesmo tempo. O *plano de bits generalizado* j é definido como sendo um conjunto de vetores $\{\mathbf{v}_{r_m} | k_m = j\}$ [3], [4], [5].

O algoritmo a ser utilizado para realizar a decomposição mostrada na Equação (2) adiciona um \mathbf{v}_{r_m} por vez até que um critério taxa-distorção seja atingido. Dado um dicionário $\mathcal{C} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q\}$, $\|\mathbf{v}_i\| = 1, \forall i$, este algoritmo é dado como segue:

- 1) Inicialmente $\mathbf{w} = \mathbf{x}$ e $m = 1$.
- 2) Repetir até atingir o critério de parada:
 - a) Escolher $r_m \in \{1, \dots, q\}$ tal que
$$\mathbf{w} \cdot \mathbf{v}_{r_m} = \max_{1 \leq j \leq q} \{\mathbf{w} \cdot \mathbf{v}_j\}.$$
 - b) Escolher $k_m = \left\lceil \frac{\ln(\mathbf{w} \cdot \mathbf{v}_{r_m})}{\ln(\alpha)} \right\rceil$, onde $\lceil \cdot \rceil$ é o menor inteiro maior que ou igual ao argumento.
 - c) Substituir \mathbf{w} por $\mathbf{w} - \alpha^{k_m} \mathbf{v}_{r_m}$.
 - d) Incrementar m .
- 3) Parar.

As condições para que $\mathbf{x}^{(P)}$ na Equação (2) convirja a \mathbf{x} enquanto o número de passos P tende ao infinito são dadas pelo seguinte teorema [3], [4]:

Teorema 1: Seja $\mathbf{x} \in \mathbb{R}^N$, $\|\mathbf{x}\| \leq 1$, um sinal aproximado pelo algoritmo MPGBP usando um dicionário \mathcal{C} com P passos, gerando $\mathbf{x}^{(P)}$ como na Equação (2), e seja $\Theta(\mathcal{C})$ o maior ângulo entre qualquer sinal $\mathbf{y} \in \mathbb{R}^N$ e o átomo mais próximo do dicionário \mathcal{C} . Temos que $\|\mathbf{r}^{(P)}\| = \|\mathbf{x} - \mathbf{x}^{(P)}\| \leq \beta_c^P$, onde $\beta_c = \sqrt{1 - (2\alpha - \alpha^2) \cos^2(\Theta(\mathcal{C}))} < 1$, $\forall 0 < \alpha < 1$ e $0 \leq \Theta(\mathcal{C}) < \frac{\pi}{2}$.

Em [3], [4] foi realizada uma comparação entre os desempenhos dos algoritmos MPGBP e MP no contexto de codificação de vídeo. O quadro diferença (comumente chamado de *dfd*, displaced frame difference), gerado pelo processo de estimação/compensação de movimento, é codificado por estes algoritmos usando a mesma estrutura e o mesmo dicionário de Neff e Zakhor [2]. Os resultados mostram que o desempenho do algoritmo MPGBP é consistentemente melhor que o do algoritmo de Neff e Zakhor [2]. Entretanto, uma importante questão ficou sem resposta: existem outros dicionários cuja utilização resulta em uma melhor performance do codificador MPGBP?

A motivação deste trabalho parte, principalmente, da tentativa de responder esta pergunta. Para tanto é necessária uma análise criteriosa da relação taxa-distorção. Do Teorema 1, vemos que no passo P o erro de aproximação é limitado por β_c^P , assim como a distorção D , como mostra a seguinte equação [3], [6]:

$$D \leq \frac{\beta_c^{2P}}{N(\mathcal{C})} \quad (3)$$

onde $N(\mathcal{C})$ representa a dimensão do dicionário. Logo, quanto menor for o valor de β_c , menor será a distorção após P passos. Como β_c decresce juntamente com $\Theta(\mathcal{C})$, então, para se ter uma distorção pequena, $\Theta(\mathcal{C})$ também deve ser pequeno. Como a cardinalidade precisa, em geral, aumentar para diminuir $\Theta(\mathcal{C})$, e a taxa R após P passos aumenta à medida que a cardinalidade $q(\mathcal{C})$ do dicionário cresce [3], [6]:

$$R \leq \frac{1}{N(\mathcal{C})} [P \log_2 q(\mathcal{C}) + P b_e] \quad (4)$$

com b_e representando o número de bits gastos na codificação de cada índice k_m , então um bom compromisso entre taxa e distorção é atingido para dicionários com valores de $\Theta(\mathcal{C})$ pequenos, mas que não possuam $q(\mathcal{C})$ muito grande.

Como $\Theta(\mathcal{C})$ é o maior ângulo entre qualquer vetor do \mathbb{R}^N e o vetor de \mathcal{C} mais próximo, ele é obtido através da região de \mathbb{R}^N que possui o maior “buraco” (ver a Figura 1a). Do exemplo em 2-D mostrado na Figura 1, podemos estipular que, dado um dicionário \mathcal{C} (Figura 1a), $\Theta(\mathcal{C})$ pode ser reduzido através: **1)** da inserção de vetores extras de modo a preencher apropriadamente as regiões “vazias” do espaço (Figura 1b) ou **2)** de uma melhor distribuição dos vetores (Figura 1c). Observe que com **2)** tanto a cardinalidade $q(\mathcal{C})$ quanto a dimensão $N(\mathcal{C})$ do dicionário \mathcal{C} são preservadas. Por outro lado, com **1)** a cardinalidade $q(\mathcal{C})$ aumenta e há um compromisso entre o decréscimo de $\Theta(\mathcal{C})$ e o aumento de $q(\mathcal{C})$.

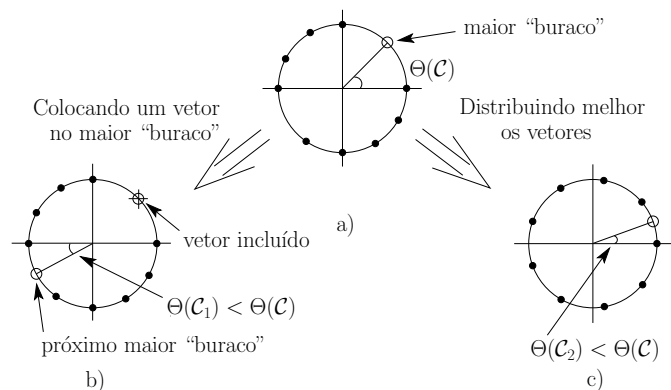


Fig. 1. $\Theta(\mathcal{C})$ em a) pode ser reduzido: b) adicionando um vetor extra no maior “buraco” do dicionário \mathcal{C} no \mathbb{R}^2 ; c) distribuindo melhor os vetores em \mathbb{R}^2 .

Assim, este artigo é orientado à busca de bons dicionários para o algoritmo MPGBP, os quais devem apresentar um bom compromisso entre $q(\mathcal{C})$ e $\Theta(\mathcal{C})$. Um levantamento bibliográfico realizado em busca de candidatos para desempenhar o papel de átomos do dicionário na representação MPGBP apontou para um novo conjunto de funções que têm despertado o interesse da comunidade de processamento de imagem/vídeo, as *Ridgelets* [7] (assim como *curvelets* [8] e *contourlets* [9]). Estas funções representam de maneira bastante eficiente sinais cujas descontinuidades estão dispostas ao longo de linhas. Como os contornos de uma imagem podem ser vistos como descontinuidades ao longo de linhas, elas representam uma alternativa efetiva para a decomposição de imagens [7]. Isto é especialmente relevante no caso de vídeo, pois as *dfds* são compostas essencialmente de contornos. Neste trabalho nós investigamos o desempenho do codificador de vídeo MPGBP usando dicionários supercompletos baseados em *ridgelets*.

A exposição deste artigo é feita da seguinte forma. Na Seção II discutimos os princípios da elaboração de bons dicionários. Na Seção III descrevemos o codificador utilizado, o qual é dotado de um mecanismo de controle preciso sobre a taxa para o algoritmo MPGBP. Este controle possibilita uma comparação mais justa entre os diferentes dicionários utilizados. Na Seção IV nós especificamos os átomos usados, apresentando e analisando os resultados das simulações. As conclusões são apresentadas na Seção V.

II. DICIONÁRIOS

Como visto anteriormente, procuramos por dicionários cujos valores de $\Theta(\mathcal{C})$ e de $q(\mathcal{C})$ sejam os menores possíveis. Qualquer dicionário \mathcal{C}' obtido por rotação dos elementos de \mathcal{C} , encarados como vetores em \mathbb{R}^N , possui $\Theta(\mathcal{C}') = \Theta(\mathcal{C})$, o que, através do Teorema 1, equivaleria dizer que ele tem o mesmo desempenho do dicionário \mathcal{C} quando utilizado no codificador MPGBP. Contudo, durante a primeira iteração do algoritmo MPGBP, a redução do erro de aproximação só depende das formas de onda dos átomos do dicionário. Quanto mais elas forem parecidas com trechos do sinal \mathbf{x} ,

maior será o valor do produto interno entre eles e, por conseguinte, menor será o valor de $\|\mathbf{x} - \alpha^{k_m} \mathbf{v}_{r_m}\|$. Desta forma, dentre todos os dicionários com o mesmo $\Theta(\mathcal{C})$, deve-se obter aquele cujas características sejam as mais próximas daquelas da imagem alvo. Com isso, vemos que um bom dicionário deve ter:

- (i) pequenos valores de $\Theta(\mathcal{C})$ e $q(\mathcal{C})$;
- (ii) átomos os mais parecidos possível com as regiões da imagem.

No contexto de codificação de vídeo, a imagem alvo é representada pelo quadro diferença originado, como dito anteriormente, por um processo de estimação/compensação de movimento. Esta imagem (dfd), é essencialmente composta por contornos de diversas larguras e orientados nas mais variadas direções. Assim, átomos baseados nas funções ridgelets, conhecidas por sua eficiência em codificar descontinuidades 1-D de direções aleatórias, são bons candidatos para satisfazer a condição (ii) enunciada acima.

No entanto, para que a condição (i) seja atendida devemos desconsiderar o uso da Transformada Ridgelet Ortogonal. Isto porque uma transformada ortogonal é uma mera rotação de qualquer outra transformada ortogonal. Logo, considerando-se a discussão levantada no início desta seção, todas as transformadas ortogonais de mesma dimensão têm um mesmo valor de $\Theta(\mathcal{C})$. Por isso, se faz necessário o uso de dicionários ridgelets supercompletos para a redução dos valores de $\Theta(\mathcal{C})$. Observe que este procedimento é retratado através da Figura 1, sendo o equivalente à adição de novos átomos visando o preenchimento dos “buracos” no dicionário.

Os dicionários são gerados de acordo com o procedimento descrito a seguir. Começamos definindo as *ridgelets discretas infinitas*, cuja *ridge* possui um ângulo ϖ com o eixo horizontal de acordo com [6]:

$$\rho_{\gamma,l,\varpi}^{(f)}(u,v) = f_{\gamma}(u \cdot \tan(\varpi) + v) \cdot W_l(u \cdot \tan(\varpi) + v) \quad (5)$$

onde $u, v \in \mathbb{Z}$, com $-\infty \leq u, v < \infty$ e γ representa um vetor de parâmetros de uma função f 1-D oscilatória e multiresolucional qualquer, muito embora tradicionalmente seja uma função wavelet. W_l é uma janela 1-D retangular e unitária de comprimento $l \in \mathbb{N}$ aplicada na direção da *ridge* e definida como:

$$W_l(n) = \begin{cases} 1, & n \in [-\frac{l-1}{2}, \frac{l-1}{2}] \\ 0, & n \notin [-\frac{l-1}{2}, \frac{l-1}{2}] \end{cases} \quad (6)$$

Note que para limitar a extensão desta ridgelet infinita necessitamos de um outro janelamento na direção ortogonal à *ridge*. De modo a realizar um janelamento suave optamos pela utilização de uma outra função ridgelet infinita, gerando, assim, um conjunto de ridgelets anisotrópicas e supercompletas dadas por [6]:

$$g_{\gamma_1, \gamma_2, l_1, l_2, \varpi}^{(f_1, f_2)}(u, v) = \rho_{\gamma_1, l_1, \varpi}^{(f_1)}(u, v) \cdot \rho_{\gamma_2, l_2, \frac{\pi}{2} - \varpi}^{(f_2)}(u, v) \quad (7)$$

A equação acima pode ser vista como uma ridgelet de comprimento infinito janelada ao longo de sua *ridge* por uma outra ridgelet de comprimento infinito. Observe que

os átomos gerados de acordo com a Equação (7) têm uma região de suporte de área retangular finita. Isto se deve ao fato de cada ridgelet infinita ter comprimento finito na direção ortogonal à sua *ridge* devido ao janelamento por W_c (ver Equações (5) e (6)).

Apesar das funções ridgelets terem sido criadas a partir de rotações das funções wavelets [7] nada impede que outras funções 1-D sejam utilizadas no seu lugar. Assim, neste trabalho $f_{\gamma}(n)$ é representado tanto pelas wavelets e scaling functions de Meyer [10] quanto pelas funções 1-D de Gabor [2], [6], as quais são geradas a partir de expansões/contrações, translações e modulações de uma mesma função Gaussiana.

Utilizando as wavelets temos que $f_{\gamma}(n) = f_{j,k}(n) = 2^{\frac{j}{2}} \psi(2^j n - k)$, onde $n \in 0, 1, \dots, l-1$, $j, k \in \mathbb{R}$, com j representando a escala e k a translação. As wavelets e scaling functions de Meyer podem ser definidas no domínio da frequência como:

$$\Psi(\omega) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{i\omega/2} \sin\left(\frac{\pi}{2}\nu\left(\frac{3}{2\pi}|\omega| - 1\right)\right), & \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ \frac{1}{\sqrt{2\pi}} e^{i\omega/2} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{4\pi}|\omega| - 1\right)\right), & \frac{4\pi}{3} \leq |\omega| \leq \frac{8\pi}{3} \\ 0, & |\omega| \notin \left[\frac{2\pi}{3}, \frac{8\pi}{3}\right] \end{cases} \quad (8)$$

$$\Phi(\omega) = \begin{cases} \frac{1}{\sqrt{2\pi}}, & |\omega| \leq \frac{2\pi}{3} \\ \frac{1}{\sqrt{2\pi}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2\pi}|\omega| - 1\right)\right), & \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ 0, & |\omega| \geq \frac{4\pi}{3} \end{cases} \quad (9)$$

onde $\nu(a) = a^4(35 - 84a + 70a^2 - 20a^3)$, $a \in [0, 1]$.

Contudo, utilizando as funções 1-D de Gabor temos que:

$$\begin{aligned} f_{\gamma}(n) &= f_{s,\xi,\phi}(n) \\ &= K_{s,\xi,\phi} \cdot g\left(\frac{n - \frac{l-1}{2}}{s}\right) \cdot \cos\left(\frac{2\pi\xi(n - \frac{l-1}{2})}{l} + \phi\right) \end{aligned} \quad (10)$$

onde $n \in 0, 1, \dots, l-1$; s , ξ e ϕ representam a escala, a frequência de modulação e o deslocamento de fase, respectivamente; a constante $K_{s,\xi,\phi}$ é escolhida de modo a normalizar $f_{s,\xi,\phi}$; e g é uma função gaussiana definida no domínio contínuo como $g(t) = \sqrt{2}e^{-\pi t^2}$.

É importante notar que os átomos gerados pela Equação (7) são, em geral, não-separáveis. O uso de dicionários não-separáveis pode acarretar em um grande crescimento da complexidade computacional do codificador. Entretanto, vários métodos de aceleração do processo de busca dos átomos foram propostos na literatura. Um bom exemplo de tais métodos é demonstrado em [11].

III. CODIFICADOR UTILIZADO

O codificador aqui utilizado é, em grande parte, o mesmo descrito em [4], [5]. A única diferença entre ele e o usado neste trabalho é a inserção de um mecanismo de controle de taxa que, além de ser simples, provê controle de taxa preciso [3], [6]. Sua razão de ser está ligada ao fato de que uma comparação justa entre dicionários requer a utilização de rigorosamente as mesmas taxas.

Essencialmente este mecanismo calcula a taxa gasta por cada novo átomo codificado e interrompe a adição de átomos quando a taxa de bits desejada é alcançada. Note que adicionar um átomo é equivalente a adicionar um par de índices (k_m, r_m) . O átomo indexado é definido tanto por sua forma (dimensões, escala, rotação, etc) quanto por sua posição. Sua forma, assim como o expoente k_m , são codificados exatamente como em [4], [5]. Já a sua posição dentro de um macrobloco 16×16 é codificada diferencialmente (ver [2], [4]). Assim, para que haja um cálculo preciso da taxa de bits devemos recalculá-la, a cada átomo codificado adicionado, a taxa de bits gasta para codificar a posição diferencial de todos os átomos codificados no macrobloco.

IV. RESULTADOS EXPERIMENTAIS

Nesta seção nós descrevemos os resultados das simulações realizadas com o uso do codificador acima descrito combinado com os diferentes dicionários aqui especificados. Todas as seqüências de imagens utilizadas são do tipo QCIF (resolução 176×144) e compostas por 300 quadros amostrados a uma taxa original de 30fps (frames per second). Entretanto, nas simulações realizadas com taxas abaixo de 20kbps houve subamostragem para uma taxa de 7,5fps (equivalendo a codificação total de 75 quadros). Já nas simulações com taxas entre 20kbps e 100kbps, as seqüências foram subamostradas e codificadas a uma taxa de 10fps (equivalendo a codificação total de 100 quadros). Nas simulações acima de 100kbps não houve subamostragem. Em todos os casos foi utilizado um $\alpha = 0.56$ (ver Equação (2) e a referência [6]).

O critério para a avaliação do desempenho dos diferentes dicionários é baseado nos valores de PSNR obtidos com a codificação das seqüências de imagens. Valores de $\Theta(\mathcal{C})$ e $\bar{\Theta}(\mathcal{C})$ também são considerados para uma análise mais apurada dos resultados. $\Theta(\mathcal{C})$ é definido no Teorema 1 e, como mencionado na Seção I, quanto menor for o seu valor, melhor será o desempenho do dicionário. Se nós definirmos para cada passo i um valor de $\beta^{(i)}$ tal que o módulo do resíduo no passo $i+1$, $\|r^{(i+1)}\|$, seja igual a $\beta^{(i)}\|r^{(i)}\|$, então $\bar{\Theta}(\mathcal{C})$ é tal que $\beta_c(\bar{\Theta}(\mathcal{C})) = \mathbf{E}[\beta^{(i)}]$ (ver Teorema 1), onde $\mathbf{E}[\cdot]$ é o valor esperado obtido levando-se em consideração todos os passos e toda a seqüência de imagens. Como é desejado o máximo de redução dos resíduos a cada passo do algoritmo, é necessário que $\mathbf{E}[\beta^{(i)}]$ seja o menor possível. Tal fato implica que bons dicionários devem ter valores de $\bar{\Theta}(\mathcal{C})$ os menores possíveis.

Primeiramente nós usamos o dicionário empregado em [2] (que daqui por diante será referido como NZ – dicionário de Neff e Zakhor) que apresentou ótimos resultados no codificador de vídeo baseado em matching pursuits implementado por Neff e Zakhor, demonstrando, também, ser um bom dicionário para o algoritmo MPGBP [4], [5]. Os resultados obtidos com o uso deste dicionário no algoritmo MPGBP modificado (ver Seção III) são mostrados na Tabela I para maior conveniência. A cardinalidade do dicionário NZ é $q(\mathcal{C}) = 400$.

TABELA I
VALORES DE PSNR, $\Theta(\mathcal{C})$ E $\bar{\Theta}(\mathcal{C})$ OBTIDOS COM O DICIONÁRIO NZ
APLICADO AO CODIFICADOR MPGBP.

Seq+Taxa (kbps)	NZ		
	PSNR (dB)	$\Theta(\mathcal{C})$	$\bar{\Theta}(\mathcal{C})$
Mother24	36,19	86,17°	82,29°
Weather24	31,75	87,52°	82,61°
Silent24	32,59	87,08°	80,19°
Mother64	40,38	87,19°	83,51°
Weather64	37,56	88,26°	83,60°
Silent64	37,72	87,81°	82,25°
Foreman64	33,45	86,80°	80,77°
Foreman96	35,52	87,07°	81,60°
Weather96	40,38	88,48°	84,00°

Uma das razões para o seu bom desempenho é o fato de seus átomos possuírem características similares às características das imagens dfts, isto é, estes átomos cumprem satisfatoriamente a condição (ii) da Seção II. Assim, se nós desenvolvermos nossos dicionários a partir do dicionário NZ nós garantiremos que a condição (ii) será satisfeita pelo menos por alguns átomos. Logo, nós projetamos nossos dicionários adicionando outros átomos ao dicionário NZ para preencher seus “buracos” de forma a diminuir $\Theta(\mathcal{C})$ (ver Figura 1), e assim satisfazer a condição (i). Os átomos adicionados são baseados nas ridgelets uma vez que nós queremos continuar satisfazendo a condição (ii) (ver a discussão da Seção II).

Na Tabela II nós vemos os resultados apresentados pelo dicionário formado com a adição de átomos compostos por ridgelets baseadas nas funções de Meyer (dicionário MEYER). Neste caso, f_i , $i = 1, 2$, são tanto as wavelets quanto as scaling functions de Meyer (ver as Equações (8) e (9)). As combinações de comprimento e escala usadas, (l_i, j_i) , $i = 1, 2$, são mostradas na Tabela III. Já o parâmetro de translação é $k = 2^{j-1}$ para as wavelets e $k = 0$ para as scaling functions [6]. Os ângulos de rotação ϖ de cada átomo de dimensões l_1 e l_2 foram escolhidos de acordo com o critério sugerido por Donoho e Flesia em [10], sendo dados por:

$$\varpi_b = \arctan\left(\frac{2b}{L-1}\right), \quad -\frac{L-1}{2} \leq b \leq \frac{L-1}{2} \quad (11)$$

onde $L = \max\{l_1, l_2\}$. A cardinalidade do dicionário MEYER resultante é de $q(\mathcal{C}) = 24617$ vetores, cerca de 60 vezes a cardinalidade do dicionário NZ.

Pode-se ver que atingimos uma redução nos valores de $\Theta(\mathcal{C})$ e $\bar{\Theta}(\mathcal{C})$ em todos os casos, acompanhado por um crescimento no valor de PSNR, quando comparamos com os resultados obtidos com o dicionário NZ. É importante observar que esta melhora foi atingida em quase todas as seqüências e taxas, mesmo tendo ocorrido um aumento extremamente significativo na cardinalidade $q(\mathcal{C})$. Isto confirma a conjectura, baseada no Teorema 1, a qual afirma que é possível se obter bons dicionários para o algoritmo MPGBP diminuindo-se $\Theta(\mathcal{C})$ a despeito do custo

TABELA II

VALORES DE PSNR, $\Theta(\mathcal{C})$ E $\bar{\Theta}(\mathcal{C})$ OBTIDOS COM O DICIONÁRIO MEYER APLICADO AO CODIFICADOR MPGBP.

Seq+Taxa (kbps)	MEYER			MEYER - NZ	
	PSNR (dB)	$\Theta(\mathcal{C})$	$\bar{\Theta}(\mathcal{C})$	Δ PSNR (dB)	$\Delta\bar{\Theta}(\mathcal{C})$
Mother24	36,46	85,36°	80,81°	0,27	-1,48°
Weather24	31,85	86,46°	81,23°	0,10	-1,38°
Silent24	32,63	85,86°	78,64°	0,04	-1,55°
Mother64	40,51	86,12°	82,25°	0,13	-1,26°
Weather64	37,80	86,91°	82,37°	0,24	-1,23°
Silent64	37,65	86,56°	80,98°	-0,07	-1,27°
Foreman64	33,83	85,84°	78,97°	0,38	-1,73°

TABELA III

COMPRIMENTOS (l) E ESCALAS (j) DO DICIONÁRIO MEYER.

l	1	3	5	7	7	9	9	11	13
j	0	1	2	1,5	2	2	2,5	3	3
l	15	21	21	23	27	29	35	35	-
j	3	3	3,5	4	4	4	4,5	5	-

do aumento da cardinalidade. Isto é, o aumento de taxa produzido pelo aumento em $q(\mathcal{C})$ foi compensado pela diminuição da distorção causada pela redução em $\Theta(\mathcal{C})$.

Já na Tabela V, vemos os resultados obtidos por um outro dicionário formado a partir do dicionário NZ adicionando-lhe átomos também baseados em ridgelets. Entretanto, desta vez não foram usadas as funções 1-D de Meyer, mas sim as funções 1-D de Gabor cujos parâmetros são mostrados na Tabela IV. Este novo dicionário é chamado de dicionário GR (Gabor Ridgelet). A sua cardinalidade é $q(\mathcal{C}) = 8090$ (cerca de 20 vezes a do NZ). Novamente nota-se que obtivemos um aumento substancial nos valores de PSNR para todas as seqüências e taxas juntamente com uma redução nos valores de $\Theta(\mathcal{C})$ e $\bar{\Theta}(\mathcal{C})$. De fato, o dicionário GR apresenta um desempenho melhor que o do dicionário MEYER. Entretanto tal fato não era inesperado, uma vez que as funções de Gabor em [2] foram geradas através de um processo de treinamento intencionalmente a obtenção de funções 2-D separáveis com características similares às características apresentadas pelas imagens dfds (isto é, elas foram projetadas especificamente para satisfazer a condição (ii)), enquanto que não houve qualquer treinamento na geração dos átomos adicionais que compõem o dicionário MEYER (por isso a diferença de cerca de 3 vezes entre suas cardinalidades).

Para uma melhor visualização do ganho de desempenho promovido pelo dicionário GR face ao dicionário NZ, expomos, por meio de gráficos comparativos, alguns resultados de simulações obtidos através do uso do codificador MPGBP com ambos os dicionários. Um conjunto maior de taxas foi utilizado, indo desde taxas muito baixas (da ordem de 10kbps) até taxas altas (da ordem de 1000kbps).

Nos gráficos das Figuras 2 e 3 comparamos os valores de PSNR obtidos com a codificação dos vídeos Mother e

TABELA IV

COMPOSIÇÃO DOS ÁTOMOS 1-D DE GABOR.

Escala (s)	Frequência (ξ)	Fase (ϕ)	Comprimento (l)
1	0	0	1
3	0	0	5
5	0	0	9
7	0	0	11
9	0	0	15
12	0	0	21
14	0	0	23
17	0	0	29
20	0	0	35
1.4	1	$\pi/2$	3
5	1	$\pi/2$	9
12	1	$\pi/2$	21
16	1	$\pi/2$	27
20	1	$\pi/2$	35
4	2	0	7
4	3	0	7
8	3	0	13
4	4	0	5
4	2	$\pi/4$	7
4	4	$\pi/4$	7

TABELA V

VALORES DE PSNR, $\Theta(\mathcal{C})$ E $\bar{\Theta}(\mathcal{C})$ OBTIDOS COM O DICIONÁRIO GR APLICADO AO CODIFICADOR MPGBP.

Seq+Taxa (kbps)	GR			GR - NZ	
	PSNR (dB)	$\Theta(\mathcal{C})$	$\bar{\Theta}(\mathcal{C})$	Δ PSNR (dB)	$\Delta\bar{\Theta}(\mathcal{C})$
Mother24	36,53	85,69°	81,03°	0,34	-1,26°
Weather24	32,25	86,62°	81,27°	0,50	-1,34°
Silent24	32,74	86,29°	78,90°	0,15	-1,29°
Mother64	40,72	86,51°	82,35°	0,34	-1,16°
Weather64	38,39	87,20°	82,34°	0,83	-1,26°
Silent64	37,96	86,92°	81,14°	0,24	-1,11°
Foreman64	33,93	85,94°	79,25°	0,48	-1,52°
Foreman96	36,01	86,34°	80,30°	0,49	-1,30°
Weather96	41,41	87,30°	82,72°	1,03	-1,28°

Weather, respectivamente. Note que o uso do dicionário GR garante ao algoritmo MPGBP uma melhora consistente de desempenho ao longo de toda a faixa compreendida entre 10kbps e 100kbps.

Através dos resultados da codificação do vídeo Foreman mostrados na Figura 4, observamos que mesmo em médias e altas taxas (100kbps - 1047kbps) o dicionário GR continua permitindo ao algoritmo MPGBP uma codificação mais eficiente dos quadros diferença, proporcionando um ganho de aproximadamente 1 dB de PSNR em toda a extensão desta banda.

A Figura 4 nos permite ainda observar que o codificador MPGBP possui desempenho comparável ao do estado da arte em codificação de vídeo utilizando o algoritmo MP [12].

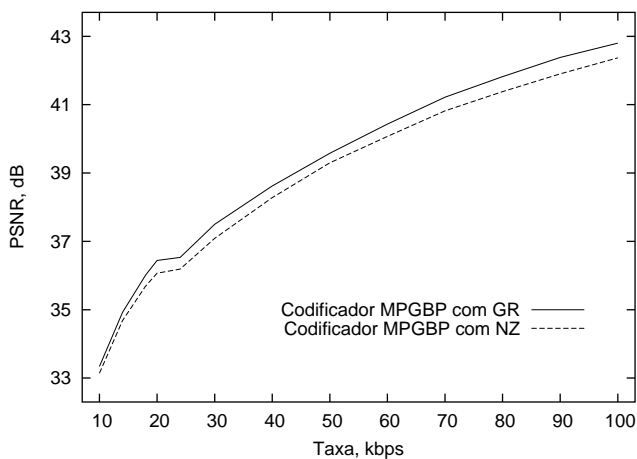


Fig. 2. Variação do PSNR em função da taxa utilizando o codificador MPGBP com os dicionários GR e NZ para a seqüência Mother.

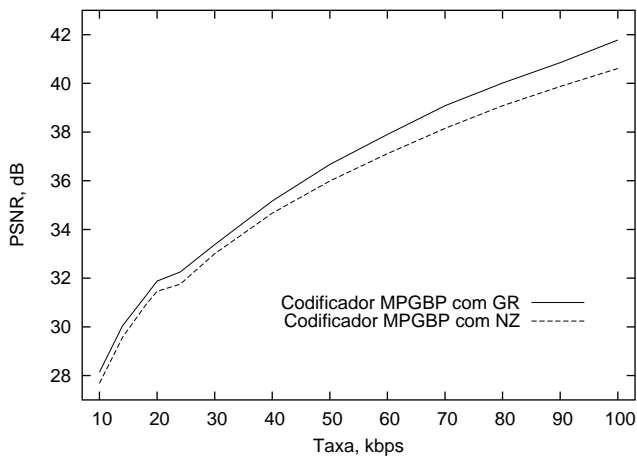


Fig. 3. Variação do PSNR em função da taxa utilizando o codificador MPGBP com os dicionários GR e NZ para a seqüência Weather.

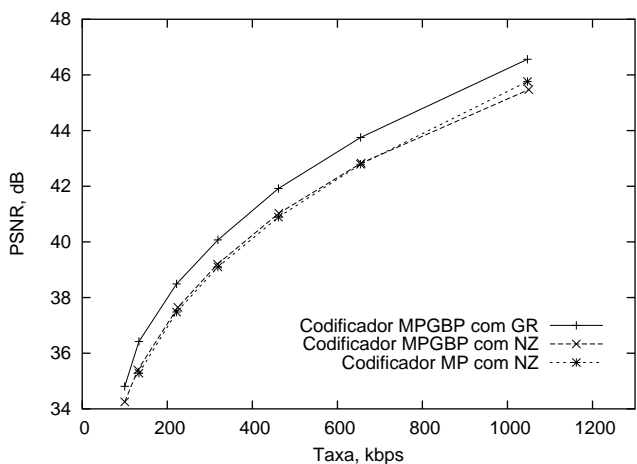


Fig. 4. Variação do PSNR em função da taxa utilizando o codificador MPGBP com os dicionários GR e NZ e utilizando a versão mais recente do codificador MP [12] com o dicionário NZ para a seqüência Foreman.

V. CONCLUSÕES

Neste trabalho nós investigamos bons dicionários para os codificadores de vídeo baseados no algoritmo MPGBP. Nós definimos condições a serem satisfeitas por estes por dicionários, as quais são: obter pequenos valores de $\Theta(\mathcal{C})$ e construir átomos com características as mais próximas possível das características da imagem. Então, propusemos um método de construção de dicionários baseados em ridgelets. Elaboramos dicionários através da adição de átomos baseados em ridgelets ao dicionário de Neff e Zakhor [2]. Estes novos átomos foram gerados tanto pelas funções de Meyer quanto pelas funções 1-D de Gabor apresentadas por Neff e Zakhor. Os resultados obtidos são motivantes, mostrando dois dicionários nos quais, apesar do grande crescimento na cardinalidade (e assim, da taxa para codificar cada átomo), a redução em $\Theta(\mathcal{C})$ compensa o crescimento da taxa, possibilitando um aumento significativo nos valores de PSNR para as seqüências e taxas aqui apresentadas. Desta forma, demonstramos que realmente há como projetar dicionários melhores, com um bom compromisso entre $\Theta(\mathcal{C})$ e a cardinalidade $q(\mathcal{C})$.

REFERÊNCIAS

- [1] Stéphane G. Mallat e Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, dezembro de 1993.
- [2] Ralph Neff e Avidesh Zakhor, "Very low bit rate video coding based in matching pursuits," *IEEE Transactions Circuits and Systems*, vol. 7, no. 1, pp. 158–171, fevereiro de 1997.
- [3] Rogério Caetano, *Codificação de Vídeo Usando Planos de Bits Generalizados*, D.Sc. thesis, COPPE - Universidade Federal do Rio de Janeiro, 2004, <http://www.lps.ufrj.br/profs/eduardo/>.
- [4] Rogério Caetano, Eduardo A. B. da Silva e Alexandre G. Ciancio, "Matching pursuits video coding using generalized bit-planes," *Proceedings 2002 International Conference on Image Processing*, vol. 3, no. 24–28, pp. 677–680, junho de 2002.
- [5] Rogério Caetano, Eduardo A. B. da Silva e Alexandre G. Ciancio, "Video coding using greedy decomposition on generalized bit-planes," *Electronics Letters*, vol. 38, no. 11, pp. 507–508, maio de 2002.
- [6] Leonardo Hidd Fonteles, "Codificadores de vídeo usando ridgelets e planos de bits generalizados," M.Sc. thesis, COPPE - Universidade Federal do Rio de Janeiro, 2004, <http://www.lps.ufrj.br/profs/eduardo/>.
- [7] Emmanuel J. Candès e David L. Donoho, "Ridgelets: a key to higher-dimensional intermittency?," *Phil. Trans. R. Soc. Lond. A.*, vol. 357, pp. 2495–2509, junho de 1999.
- [8] Emmanuel J. Candès e David L. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges," In A. Cohen, C. Rabut and L. L. Schumaker, editors, *Curve and Surface Fitting, Saint-Malo, Vanderbilt University Press*, 1999.
- [9] Minh N. Do e Martin Vetterli, "Contourlets: a directional multiresolution image representation," *Image Processing. 2002. Proceedings. 2002 International Conference on, Volume: 1*, 22–25, vol. 1, pp. I-357 – I-360, setembro de 2002.
- [10] David L. Donoho e Ana Georgina Flesia, *Beyond Wavelets*, chapter Digital Ridgelet Transform based on True Ridge Functions, Grant Welland (ed.), pp. 1–30, Academic Press, Setembro de 2003.
- [11] Ralph Neff e Avidesh Zakhor, "Matching pursuits video coding - part 1: Dictionary approximation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 1, pp. 13–26, janeiro de 2002.
- [12] C. D. Vleeschouwer e A. Zakhor, "In loop atom modulus quantization for matching pursuit and its applications to video coding," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1226–1242, outubro de 2003.