2
Aplicação da Codificação Wyner-Ziv para Vídeo

2.1.
Cálculo de Taxa-Distorção para Codificação com Informação Lateral no Receptor

Wyner & Ziv [2] desenvolveram uma nova abordagem a partir do teorema de Slepian e Wolf, a fim de estabelecer limites da teoria da informação para compressão com perdas e com informação lateral presente no decodificador. Wyner e Ziv encaminharam da seguinte forma a análise de taxa-distorção para esta situação: sejam *X* e *Y* amostras de duas sequências aleatórias dependentes e igualmente distribuídas, de alfabetos possivelmente finitos, modelando os dados da fonte e a informação lateral, respectivamente. Os valores dos dados da fonte *X* são codificados sem acesso à informação lateral *Y* (figura 2.1). O decodificador, porém, tem acesso a *Y*, e obtém uma reconstrução $\hat{X}$ dos valores da fonte (no alfabeto χ de possíveis valores de *X*), onde $\hat{X} = E[X / \hat{Y}]$. Considera-se também que uma pequena distorção $D = E[d(X, \hat{X})]$ é admissível.

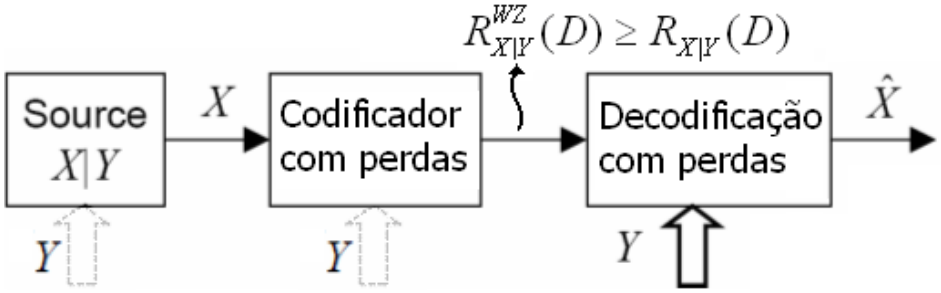

Figura 2.1: Compressão com perdas de uma sequência X utilizando a informação lateral Y (relacionada estatisticamente a X) no decodificador.

Neste contexto, a função taxa-distorção Wyner-Ziv $R_{X|Y}^{WZ}(D)$ representa o limite inferior alcançável da taxa de bits, dada uma distorção máxima *D*. Denotamos por $R_{X|Y}(D)$ a taxa requerida se a informação lateral estivesse disponível também no codificador. Wyner e Ziv provaram que, não

PUC-Rio - Certificação Digital Nº 0610789/CA

surpreendentemente, uma perda de taxa $R_{X|Y}^{WZ}(D) - R_{X|Y}(D) \geq 0$ é incorrida quando o codificador não tem acesso à informação lateral. Porém, eles também mostraram que $R_{X|Y}^{W|Z}(D) - R_{X|Y}(D) = 0$ no caso de fontes gaussianas sem memória, adotando o critério de distorção do erro médio quadrático (MSE) [2]. A codificação distribuída é o dual do teorema apresentado por Costa et al. sobre codificação de canal com a informação lateral presente somente no codificador [32-34]. A diferença $R_{X|Y}^{WZ}(D) - R_{X|Y}(D) = 0$ também se mantém para sequências *X* que são formadas pela soma de uma informação lateral arbitrariamente distribuída *Y* com ruído gaussiano independente [34]. Para estatísticas gerais e medidas de distorção utilizando o erro médio quadrático, Zamir provou que a perda de taxa é menor que 0.5 bit/amostra [35].

Em geral, um codificador Wyner-Ziv pode ser pensado como consistindo de um quantizador seguido por um codificador Slepian-Wolf, conforme ilustra a figura 2.2, onde o quantizador divide o espaço de sinais em células (ou grupos), as quais podem ser formadas por subcélulas mapeadas com o mesmo índice de quantização Q.

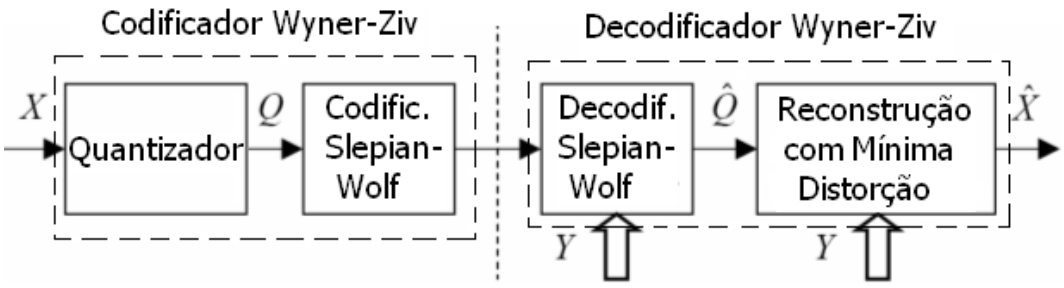

Figura 2.2: Um codificador Wyner-Ziv prático é obtido cascateando um quantizador e um codificador Slepian-Wolf.

Assim, este trabalho fundamenta-se na aplicação do teorema de Wyner-Ziv para codificação com perdas (residuais) e com a informação lateral presente apenas no decodificador, já que o contexto em que se aplica o codec proposto refere-se a um ambiente onde há necessidade de se utilizar um codificador de baixa complexidade, deslocando a grande carga computacional para o decodificador, como será mostrado na seção 4.2.

PUC-Rio - Certificação Digital Nº 0610789/CA

2.2.
Revisão de literatura sobre codificação Wyner-Ziv

Da mesma forma que acontece com a codificação Slepian-Wolf, somente recentemente, esforços em direção a esquemas de codificação Wyner-Ziv práticos têm sido realmente explorados. As primeiras tentativas para projetar decodificadores para correta reconstrução dos valores da fonte, com o auxílio da informação lateral, foram inspiradas em provas da teoria da informação. Zamir & Shamai [36] provaram que, sob certas circunstâncias, códigos lineares reticulados aninhados podem aproximar-se da função taxa-distorção de Wyner-Ziv, em particular se os dados da fonte e a informação lateral são conjuntamente gaussianos. Esta idéia foi também desenvolvida e aplicada por Pradhan *et al*. [3], e Servetto [38], o qual publicou diferentes projetos para análise de desempenho, ambos focados no caso gaussiano. Xiong *et al*. [39, 40] implementaram o codificador Wyner-Ziv como um quantizador reticulado aninhado seguido por um codificador Slepian-Wolf e, em [41], utilizaram um quantizador com codificação por treliça.

Esta configuração (apresentada na figura 2.2) foi considerada, por exemplo, por Fleming *et al*. [42], o qual generalizou o algoritmo de Lloyd [43] para o projeto de quantização vetorial Wyner-Ziv ótima para taxa fixa. Mais tarde, Fleming & Effros [44] incluíram quantizadores vetoriais otimizados no sentido taxa-distorção, nos quais a medida da taxa é uma função do índice de quantização, ou da palavra-código. Infelizmente, a dimensionalidade do quantizador vetorial e o comprimento do bloco de código entrópico são idênticos em sua formulação, e por isso, estes quantizadores ou carecem de bom desempenho ou são exageradamente complexos.

Uma extensão mais geral do algoritmo de Lloyd aparece no trabalho de Monedero *et al*. [48]. Um quantizador é projetado assumindo que um codificador Slepian-Wolf ideal é utilizado para codificar o índice de quantização. A introdução de uma medida de taxa que dependa de ambos, do índice de quantização e da informação lateral, separa a dimensionalidade do quantizador do comprimento do bloco do codificador Slepian-Wolf, um requisito fundamental para o projeto de um sistema prático. Girod [49] mostrou que em altas taxas de transmissão, sob certas condições, os quantizadores mais eficientes são os

PUC-Rio - Certificação Digital Nº 0610789/CA

quantizadores reticulados, pois células de quantização não conectadas não precisam ser mapeadas para o mesmo índice, e assintoticamente, não existe perda de desempenho por não se ter acesso à informação lateral no codificador. Isto é confirmado pelos resultados experimentais de Zhang *et al*. [48].

2.3. Funcionamento da Codificação de Vídeo de Baixa Complexidade

As implementações dos padrões de codificação de vídeo atuais, tal como os métodos de codificação da recomendação ITU-T H.264/AVC [84], impõem uma carga computacional muito maior no codificador do que no decodificador, ou seja, o codificador é bem mais complexo que o decodificador. Esta assimetria é bem adequada para transmissão em *broadcasting* ou para sistemas de *streaming* de vídeo-sob-demanda onde o vídeo é comprimido uma vez e decodificado muitas vezes.

Entretanto, algumas aplicações podem requerer o sistema dual, isto é, codificadores de baixa complexidade, à custa de decodificadores de alta complexidade. Exemplos de tais sistemas incluem sensores de vídeo *wireless* para vídeo-vigilância, câmeras de PC *wireless*, câmeras móveis e filmadoras em rede. Em todos estes casos, a compressão deve ser implementada na câmera, onde a memória e a capacidade computacional, em geral, são escassas.

A teoria de Wyner-Ziv apresentada em [2, 34, 35] é o fundamento de um sistema de codificação de vídeo não-convencional, no qual os frames individuais são codificados independentemente, mas decodificados condicionalmente, ou melhor, conjuntamente. De fato, tal sistema pode alcançar um desempenho que é mais próximo da codificação *interframe* convencional do que de uma codificação *intraframe* convencional. Em contraste com a codificação preditiva convencional de vídeo, onde os frames anteriores compensados em movimentos (preditores) são usados como informação lateral, no sistema proposto para DVC, estes frames são utilizados como informação lateral somente no decodificador, isto é, visando atender à exigência de baixa complexidade no codificador, o processo de estimação de movimento e predição inexistem no mesmo.

PUC-Rio - Certificação Digital Nº 0610789/CA

O codificador de vídeo Wyner-Ziv tem grande vantagem em relação ao custo computacional, visto que ele comprime cada frame de vídeo independentemente, requerendo somente processamentos do tipo *intraframe* no codificador. E o decodificador, na parte fixa da rede, exploraria a dependência estatística entre os frames, aplicando um processamento mais complexo. Além de deslocar computacionalmente a custosa estimação de movimento e a correspondente compensação do codificador para o decodificador, a desejada assimetria é também consistente com os algoritmos de codificação de Slepian-Wolf e Wyner-Ziv, os quais tendem a utilizar codificadores mais simples, mas decodificadores com maiores demandas.

Mesmo se o receptor também for um dispositivo com restrição de complexidade, como seria o caso de codecs de vídeo em terminais móveis em ambas extremidades, ainda assim é vantajoso empregar a codificação Wyner-Ziv em conjunto com uma arquitetura de transcodificação, conforme ilustrada na figura 2.3 [68].

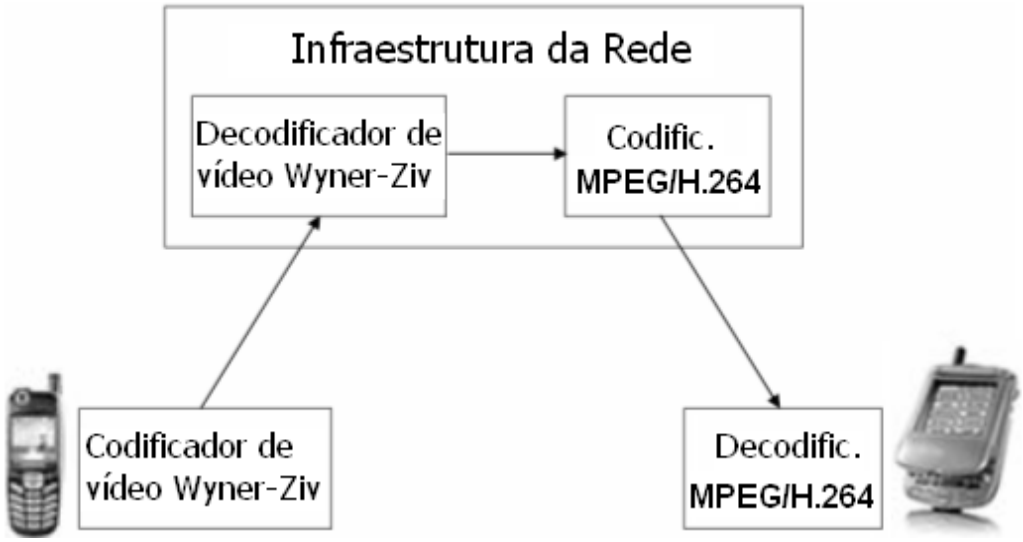

Figura 2.3: Arquitetura de transcodificação de vídeo para transmissão *wireless*, utilizada na rede móvel celular [68].

Na arquitetura apresentada na figura 2.3, a câmera do aparelho móvel captura e comprime o vídeo usando a técnica de codificação Wyner-Ziv e transmite os dados para a parte fixa da rede, onde se pode ter um codec altamente complexo. Lá, o *bitstream* é decodificado através do decodificador Wyner-Ziv e depois recodificado utilizando padrões convencionais de codificação de vídeo, tal

PUC-Rio - Certificação Digital Nº 0610789/CA

como ITU-T H.264/AVC [84]. Esta arquitetura não somente joga a carga computacional para a parte fixa da rede, como também compartilha o transcodificador entre vários usuários, provendo uma economia de custos adicionais.

2.4. Arquitetura-padrão para Codec DVC

A arquitetura mais utilizada para DVC e considerada como *estado da arte* no assunto provém da combinação de um codificador *intraframe* e um decodificador *interframe* para codificação de vídeo, conforme mostra a figura 2.4 [52-54], onde o módulo simbolizado pela letra Q representa a etapa de quantização dos coeficientes da transformada DCT (*Discrete Cosine Transform* - transformada discreta de cossenos). A transformada DCT será explicada com detalhes na seção 4.5.1.

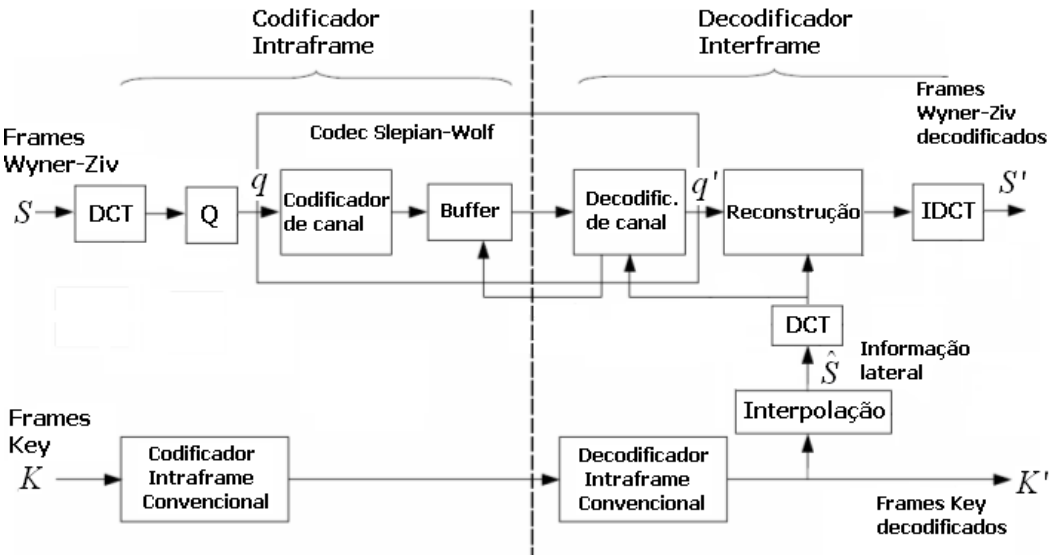

Figura 2.4: Codificador de vídeo de baixa complexidade e decodificador correspondente [52].

Nesse método de codificação ilustrado pela figura 2.4, um subconjunto de frames, igualmente espaçados na sequência, serve como *key frames*, *K*, os quais são codificados e decodificados utilizando um codec *intraframe* convencional (por exemplo, o ITU-T H.264/AVC [84]), com transformada discreta de cossenos (DCT) aplicada a blocos 8x8. Os frames entre os *key* frames são os Wyner-Ziv frames, os quais são intraframe-codificados, mas interframe-decodificados.

PUC-Rio - Certificação Digital Nº 0610789/CA

Para um frame Wyner-Ziv, chamado aqui de *S*, cada coeficiente DCT é uniformemente quantizado para um dos 2^M intervalos, onde *M* é o número de bits utilizados na codificação binária dos índices de quantização. O codificador Slepian-Wolf provê um bloco suficientemente grande de índices de quantização *q*. Rowitch & Milstein [59] implementaram o codificador Slepian-Wolf utilizando o código turbo perfurado compatível com a taxa (RCPT – *Rate-Compatible Punctured Turbo code*) para ser usado como codificador de fonte. O RCPT provê uma flexibilidade de taxa, a qual é essencial na adaptação às estatísticas variáveis entre a informação lateral $\hat{S}$ e o frame a ser decodificado S.

Para cada frame Wyner-Ziv, o decodificador gera a informação lateral $\hat{S}$ (uma estimativa do frame Wyner-Ziv *S*), através da interpolação dos *key* frames decodificados previamente, havendo também a possibilidade de utilizar os frames Wyner-Ziv decodificados previamente, caso o GOP (*Group of Pictures*) [50, 51] seja "KSSSK" em vez de "KSK", como é o padrão para DVC. Em codificação de vídeo, um *Group of Pictures*, ou uma estrutura de GOP, especifica a ordem em que os *intraframes* e os *interframes* estão arranjados. Cada stream de vídeo codificado consiste de sucessivos GOPs, onde um GOP sempre começa com um I-frame, seguido por frames P ou B [50, 51, 84].

Antes da informação lateral ser utilizada, é aplicada uma transformada DCT em cada bloco de NxN amostras da mesma, resultando em bandas de coeficientes DCT da informação lateral. Um banco de decodificadores (os quais podem ser compostos por códigos turbo, LDPC ou outro com desempenho satisfatório) reconstrói cada banda de coeficientes quantizados de forma independente, utilizando os coeficientes correspondentes da informação lateral. Então, cada banda de coeficientes quantizados é reconstruída (e escolhida) como sendo a melhor estimativa dos valores originais, dados os símbolos reconstruídos e a informação lateral.

Para explorar a informação lateral, o decodificador assume um modelo estatístico para o "canal de correlação". Especificamente, assume-se uma distribuição Laplaciana para a diferença entre os valores individuais dos pixels do frame Wyner-Ziv *S* e de sua estimativa $\hat{S}$, a qual é gerada a partir dos frames previamente reconstruídos. O decodificador estima o parâmetro laplaciano pela observação das estatísticas dos frames decodificados previamente, ou seja,

PUC-Rio - Certificação Digital Nº 0610789/CA

utilizam-se os intraframes no cálculo dos parâmetros necessários à construção da distribuição laplaciana. Vale ressaltar que esta distribuição de resíduos entre o frame Wyner-Ziv e sua estimativa foi adotada por todos os autores que desenvolveram projetos em DVC citados neste trabalho.

O decodificador combina a informação lateral $\hat{S}$ (que é uma estimativa de *S*) e os bits de paridade recebidos para recuperar o *stream* de símbolos quantizados *q'* (estimativa dos índices de quantização). Se o decodificador não puder decodificar de maneira confiável os símbolos originais, ele pede bits de paridade adicionais para o *buffer* do codificador através de um canal de retorno (ou realimentação). O processo de pedir e decodificar é repetido até que uma probabilidade aceitável de erro de símbolo seja alcançada. Pela utilização da informação lateral, o decodificador freqüentemente prediz de forma correta o grupo ou índice de quantização *q* do símbolo original e assim, precisa requerer somente $k \leq M$ bits para estabelecer a qual dos 2^M grupos um símbolo pertence. Com isso, uma compressão é alcançada.

Depois do receptor decodificar o índice de quantização *q'*, ele reconstrói os blocos de coeficientes DCT, dados estes índices de quantização e os coeficientes DCT da informação lateral, ou seja, $S'_{DCT} = E[S_{DCT} \mid q', \hat{S}_{DCT}]$, baseado no método do mínimo erro médio quadrático ($MSE_{mín}$). O procedimento de reconstrução dos coeficientes DCT é o seguinte: se a informação lateral $\hat{S}$ estiver contida na célula (conjunto) de possíveis valores de *q*, o coeficiente DCT do frame atual reconstruído levará um valor próximo ao valor correspondente do coeficiente DCT da informação lateral. Entretanto, se a informação lateral e o índice de quantização *q'* discordam, isto é, se $\hat{S}$ está fora da célula de quantização, a função reconstrução força $\hat{S}$ a cair dentro da célula, assumindo no máximo, algum dos seus limites. Ele, portanto, limita a magnitude do erro de reconstrução a um valor máximo, determinado pela precisão do quantizador. Esta propriedade é perceptivelmente desejável já que elimina grandes erros, os quais seriam visualmente desconfortáveis ao telespectador, conforme o erro de predição ilustrado na figura 2.5(a) [68]. Essa figura apresenta um frame reconstruído, da sequência *Salesman*, no formato QCIF. Repare que na reconstrução do frame Wyner-Ziv (figura 2.5(b)), a interpolação compensada em movimento fornece bons resultados para a maior parte da imagem. Os grandes erros de interpolação

PUC-Rio - Certificação Digital Nº 0610789/CA

presentes no frame-informação lateral (figura 2.5(a)) são corrigidos pela decodificação conjunta do *bitstream* Wyner-Ziv.

(a) (b)

Figura 2.5: Frames da sequência *Salesman*, QCIF: (a) Informação lateral $\hat{S}$ no decodificador, gerada por interpolação compensada do movimento; (b) Frame reconstruído *S'* depois da decodificação conjunta Wyner-Ziv.

Neste exemplo, a informação lateral (figura 2.5(a)), gerada pela interpolação compensada em movimento, contém artefatos e distorções. O decodificador Wyner-Ziv consegue refazer os contornos da imagem e reconstrói as mãos e a face, ainda que a interpolação tenha falhado nestas partes da imagem (nas quais exige-se uma alta resolução/precisão).

Comparada à codificação preditiva compensada em movimento, a codificação Wyner-Ziv no domínio da transformada é ordens de magnitude menos complexa, pois a estimação de movimento, a compensação e a predição não são requeridas no codificador. No caso do codificador Slepian-Wolf considerado *estado da arte* e adotado como base para o desenvolvimento da arquitetura proposta, ele somente requer um canal de *feedback*.

As figuras 2.6 e 2.7 [68] ilustram o experimento realizando pelo grupo de pesquisa em DVC da Universidade de Stanford, Estados Unidos, onde se buscou avaliar o desempenho taxa-distorção do codificador de vídeo Wyner-Ziv, para as sequências *Salesman* e *Hall Monitor*, na resolução QCIF com 10 fps (frames por segundo), no domínio da transformada, utilizando um bloco DCT de tamanho 4x4. A informação lateral é gerada pela simples interpolação do movimento compensado, utilizando frames adjacentes reconstruídos.

PUC-Rio - Certificação Digital Nº 0610789/CA

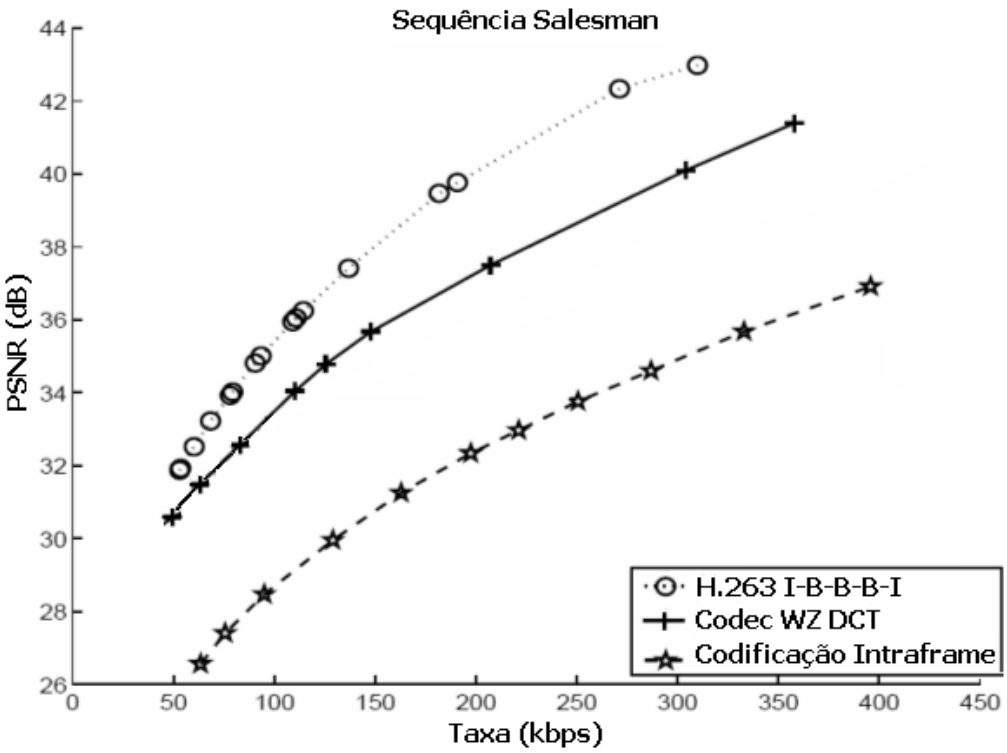

Figura 2.6: Desempenho taxa-distorção de um codec de vídeo Wyner-Ziv, comparado à codificação de vídeo intraframe e interframe convencional, para sequência *Salesman*.

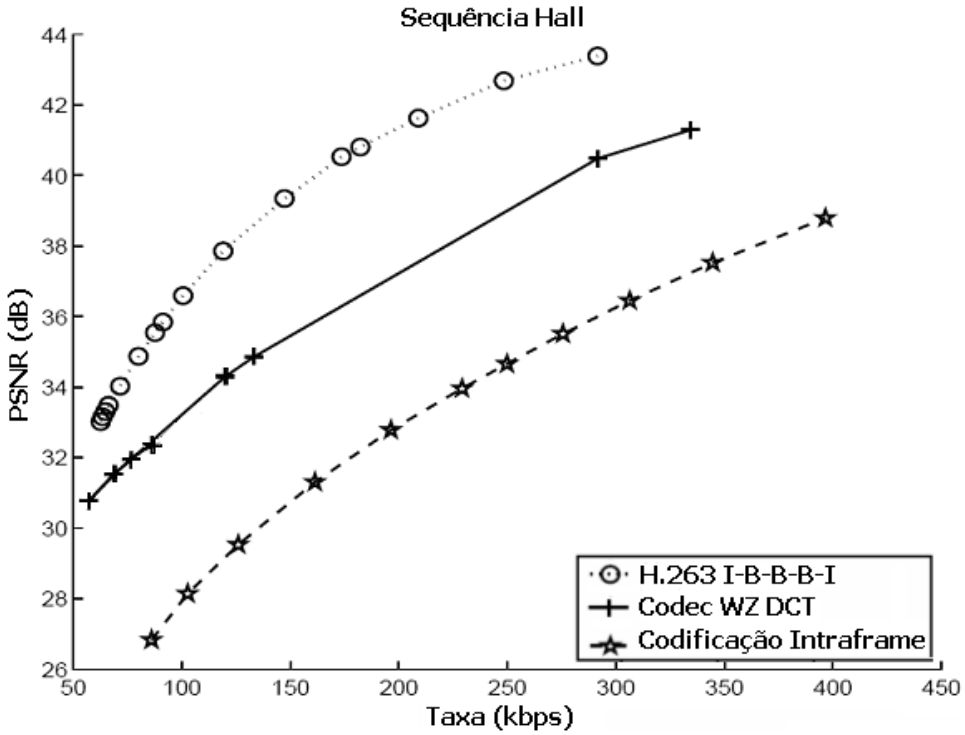

Figura 2.7: Desempenho taxa-distorção de um codec de vídeo Wyner-Ziv, comparado à codificação de vídeo intraframe e interframe convencional, para sequência *Hall Monitor*.

PUC-Rio - Certificação Digital Nº 0610789/CA

Os gráficos (figuras 2.6 e 2.7) mostram os valores da PSNR (*Peak Signal-to-Noise Ratio*) [84, 85, 87] em relação à taxa, os quais foram obtidos efetuando-se a média entre os frames *key* e os frames Wyner-Ziv para cada taxa de transmissão. Os resultados da técnica DVC empregada em Stanford [4] foram comparados à: (i) codificação *intraframe* baseada na transformada DCT (todos os frames foram codificados como frames I); (ii) codificação *interframe* H.263+ com uma estrutura de predição I-B-B-B-I. Observando-se os gráficos, nota-se que o codificador Wyner-Ziv no domínio da transformada (codec WZ DCT) tem um desempenho (qualidade) de 2 a 5 dB melhor que o da codificação *intraframe* convencional. Entretanto, ainda há uma significante diferença em direção à codificação *interframe* H.263+. Em seus experimentos preliminares realizados em um Pentium III, com 1.2 GHz, observaram um tempo médio para execução da codificação (sem operações de arquivo) de aproximadamente 2.1 mseg/frame para o método Wyner-Ziv, comparado ao tempo de 36.0 mseg/frame para codificação H.263+ I-frame e 227.0 mseg/frame para codificação H.263+ B-frame [68].

O método baseado na DCT possui maior complexidade em seu codificador do que o sistema no domínio do pixel, adotado em [69, 72]. Embora o codificador Wyner-Ziv tenha um bom desempenho, sua complexidade é similar ao método convencional de codificação de vídeo intraframe. No entanto, a codificação distribuída baseada na DCT permanece muito menos complexa do que os esquemas de codificação preditiva interframe, pois a estimação e a compensação do movimento não são necessárias no codificador.

PUC-Rio - Certificação Digital Nº 0610789/CA