



DISSERTAÇÃO DE MESTRADO EM
ENGENHARIA ELÉTRICA

**SUPER-RESOLUÇÃO UTILIZANDO
QUADROS CHAVE EM SEQUÊNCIAS
DE VÍDEO DE RESOLUÇÃO MISTA**

Fernanda Brandi da Silva

Brasília, Janeiro de 2009

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO EM
ENGENHARIA ELÉTRICA

**SUPER-RESOLUÇÃO UTILIZANDO
QUADROS CHAVE EM SEQUÊNCIAS
DE VÍDEO DE RESOLUÇÃO MISTA**

Fernanda Brandi da Silva

Dissertação de Mestrado submetida ao Departamento de Engenharia Elétrica da Faculdade de Tecnologia da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do grau de Mestre.

Banca Examinadora

Prof. Ricardo Lopes de Queiroz, Ph.D., _____
UnB/ENE - Orientador

Prof. Juliana Camapum, Ph.D., UnB/ENE - _____
Examinador Interno

Prof. Joceli Mayer, Ph.D., UFSC - _____
Examinador Externo

FICHA CATALOGRÁFICA

DA SILVA, FERNANDA BRANDI

Super-Resolução Utilizando Quadros Chave
em Sequências de Vídeo de Resolução Mista. [Distrito Federal] 2009.

xiii, 90p., 297 mm (ENE/FT/UnB, Mestre, Telecomunicações

Processamento de Sinais, 2008). Dissertação de Mestrado.

Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Elétrica.

1. Super-Resolução

2. Quadros Chave

3. Codificação de Vídeo Distribuída

4. Codificação de Complexidade Reversa

5. H.264

6. Vídeo para Dispositivos Móveis

I. ENE/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

BRANDI, F. (2009). Super-Resolução Utilizando Quadros Chave em Sequências de Vídeo de Resolução Mista. Dissertação de Mestrado em Engenharia Elétrica com ênfase em Telecomunicações, Publicação PPGENE.DM - 362A/09, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 90p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Fernanda Brandi da Silva.

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Super-Resolução Utilizando Quadros Chave em Sequências de Vídeo de Resolução Mista.

GRAU / ANO: Mestre / 2009

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

Fernanda Brandi da Silva
SQSW 300 - Bloco O - Apto 411
70673-052 Sudoeste, Brasília - DF - Brasil.

Dedicatória

Dedico o presente trabalho a meus pais e irmãs que principalmente em momentos difíceis, dando broncas, conselhos, ou me deixando descobrir meus próprios rumos, sempre me deram suporte para seguir em frente nessa caminhada tortuosa.

Fernanda Brandi da Silva

Agradecimentos

À minha mãe, Valéria, que desde sempre respondia minhas milhares de perguntas com a maior paciência, exceto quando eu abusava da boa vontade dela na hora de dormir. Ela que nunca repreendeu seriamente meu espírito curioso e questionador que desde criança tomou conta de mim e que acredito, de algum modo, me levou para o caminho da engenharia.

Ao meu pai, Ruy César, que mesmo com seu jeito reservado e mais distante sempre me ajudou especialmente nas situações mais sérias e decisivas. Que ele continue sendo a pessoa responsável e justa que sempre foi, mas, torço eu, que ele possa curtir mais a vida e desista de ver aqueles filmes de gosto duvidoso para que possamos conversar também sobre outras amenidades.

Às minhas irmãs, Patrícia e Cristiane, que em todo momento sempre foram parte essencial da família e que sempre me apresentaram pontos de vista diferentes para as questões da vida. Que elas alcancem felicidades e realizações na vida acadêmica, profissional e pessoal tanto quanto eu acho que muito merecem.

Ao pessoal do GPDS e do grupo Image pelos ensinamentos sempre valiosos... Mintsu, Renan, Zaghetto, Grilo, Karen, Bruno, Capim, Alves, Buraco, por propiciarem um ambiente de estudo e pesquisa tão agradável, eclético e enriquecedor. Uma menção especial ao meu orientador, prof. Queiroz, por sempre ter me incentivado a trabalhar com ele e me levou ao caminho da pesquisa. Ao Mintsu um agradecimento especial por muitas conversas, ensinamentos, idéias e madrugadas de companheirismo e trabalho, uma real fonte zen de inspiração.

Aos professores que serviram de inspiração por mostrarem gosto e talento na transmissão de conhecimentos e que de fato poderiam ser chamados de mestres. A eles minha sincera admiração por seguirem uma profissão por vezes tão desgastante e sem o devido reconhecimento que tanto lhes é de direito.

Aos amigos distantes que mesmo com a distância geográfica não são menos amigos. E, por último mas nunca menos especial, Maíra Lioi, pela grande amizade que espero que perdure por toda a vida. Mesmo.

Fernanda Brandi da Silva

RESUMO

É desejável gerar imagens que possuam alta-resolução espacial. Uma alta-resolução espacial pode contribuir para uma melhor classificação e/ou segmentação de regiões ou uma melhor análise em imagens multi-espectrais; e também pode gerar uma imagem mais agradável em sistemas de alta-definição. A resolução de uma imagem depende da resolução do dispositivo de aquisição. Super-resolução é o nome das técnicas que aumentam a resolução espacial de um sistema de imageamento. Então, uma imagem super-resolvida pode alcançar resolução mais alta comparada ao que o dispositivo pode gerar. Neste trabalho é apresentada uma técnica de super-resolução em um esquema de resolução mista. Alguns quadros, chamados quadro chave, estão disponíveis em resolução normal enquanto os outros quadros estão em resolução reduzida e precisam então ser super-resolvidos para alcançarem resolução normal. No esquema proposto, a informação disponível provinda dos quadros chave é processada para super-resolver os quadros não-chave que foram previamente sub-amostrados. Diferentemente de outras arquiteturas de super-resolução, no esquema proposto é assumido que o sistema de aquisição pode alcançar a resolução desejada porém por limitações do sistema de armazenamento ou transmissão é melhor que somente alguns quadros sejam transmitidos em resolução normal. Tais limitações podem incluir restrições de banda ou redução de complexidade na codificação. A redução da complexidade no processo de codificação é importante em dispositivos móveis com limitações de energia e processamento tais como telefones celulares e câmeras digitais.

ABSTRACT

It is desirable to generate an image that has a high spatial resolution. A high spatial resolution can contribute to a better classification and/or segmentation of regions or a better analysis in multi-spectral images; and it can also generate a more pleasant view in high definition systems. The resolution of an image depends on the resolution of the acquisition device. Super-resolution is the name of the techniques that enhance the spatial resolution of an imaging system. Therefore, a super-resolved image could achieve a higher resolution than the acquisition device can generate. In this work it is presented a super-resolution technique in a mixed resolution framework. Some frames, called key frames, are available at regular resolution while the middle frames are at low-resolution and thus they need to be super-resolved to the regular resolution. In the proposed scheme, the available information from the key frames is processed to super-resolve the non-key frames which were previously down-sampled. Different from other super-resolution architectures, in the proposed framework it is assumed that the acquisition device can achieve the desirable resolution, but due to storage or transmission system limitations it is better if only some frames are transmitted at regular resolution. Such limitations can include bandwidth restriction or encoding complexity reduction. The reduction of encoding complexity is important in power constrained portable devices, such as popular mobile phones and digital cameras.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	VÍDEO E COMPRESSÃO DE VÍDEO	1
1.2	CONTEXTUALIZAÇÃO	2
1.2.1	CODIFICAÇÃO DE VÍDEO	3
1.2.2	ESTIMAÇÃO DE MOVIMENTO	4
1.2.3	EFEITOS DA FILTRAGEM NO REDIMENSIONAMENTO DE IMAGENS	5
1.2.4	MÉTRICAS DE AVALIAÇÃO DE QUALIDADE DE IMAGEM.....	8
1.3	DEFINIÇÃO DO PROBLEMA E OBJETIVOS.....	9
1.4	APRESENTAÇÃO DO MANUSCRITO	10
2	SUPER-RESOLUÇÃO	11
2.1	SUPER-RESOLUÇÃO BAYESIANA.....	17
2.1.1	MODELAMENTO BAYESIANO	19
2.1.2	INFERÊNCIA BAYESIANA	19
2.1.3	MODELANDO IMAGENS DE BAIXA-RESOLUÇÃO.....	20
2.1.4	ESTIMAÇÃO DE IMAGENS DE ALTA-RESOLUÇÃO.....	23
2.2	SUPER-RESOLUÇÃO BAYESIANA APLICADA À COMPRESSÃO DE IMAGENS	25
3	SUPER-RESOLUÇÃO PROPOSTA UTILIZANDO QUADROS CHAVE	29
3.1	VISÃO GERAL DO SISTEMA DE CODIFICAÇÃO E DECODIFICAÇÃO	29
3.1.1	SISTEMAS DE CODIFICAÇÃO E DECODIFICAÇÃO TRADICIONAIS	30
3.1.2	SISTEMAS DE CODIFICAÇÃO E DECODIFICAÇÃO DE COMPLEXIDADE REVERSA	31
3.1.3	SUPER-RESOLUÇÃO UTILIZANDO QUADROS CHAVE SOB A ÓTICA DO CODIFICADOR	33
3.1.4	SUPER-RESOLUÇÃO UTILIZANDO QUADROS CHAVE SOB A ÓTICA DO DECODIFICADOR	35
3.2	MÉTODO DE SUPER-RESOLUÇÃO PROPOSTO: ABORDAGEM BÁSICA....	37
3.2.1	EXPLORANDO A CORRELAÇÃO ENTRE OS QUADROS.....	38
3.2.2	RECUPERAÇÃO DA ALTA-FREQUÊNCIA	39
3.2.3	REALIZANDO A ESTIMAÇÃO DE MOVIMENTO.....	43
3.2.4	SUPER-RESOLUÇÃO DOS QUADROS NÃO-CHAVE	44
3.3	MÉTODO DE SUPER-RESOLUÇÃO PROPOSTO: REFINAMENTOS	47
3.3.1	ESTIMAÇÃO DE MOVIMENTO NA ALTA-FREQUÊNCIA	47
3.3.2	SELEÇÃO DE MACROBLOCOS DE INTERESSE.....	49
3.3.3	ANALISANDO AS MÉTRICAS UTILIZADAS NA ESTIMAÇÃO DE MOVIMENTO	53
3.3.4	UTILIZANDO A FILTRAGEM DE TEXTURAS	57
3.4	APLICANDO OS REFINAMENTOS AO MÉTODO BÁSICO DA SUPER- RESOLUÇÃO PROPOSTA	61
4	EXPERIMENTOS E RESULTADOS PRÁTICOS	65
4.1	ALTERANDO A DENSIDADE DE QUADROS CHAVE	67

4.1.1	ANÁLISE DE RESULTADOS.....	67
4.2	ALTERANDO A COMPRESSÃO DOS QUADROS.....	72
4.2.1	ANÁLISE DE RESULTADOS.....	72
4.3	RESULTADOS SUBJETIVOS.....	77
4.3.1	ANÁLISE DE RESULTADOS.....	77
5	CONCLUSÃO.....	83
5.1	TRABALHOS PUBLICADOS	84
5.2	TRABALHOS FUTUROS	85
	REFERÊNCIAS BIBLIOGRÁFICAS.....	87

LISTA DE FIGURAS

1.1	Amostragem temporal e espacial de uma sequência de vídeo.	1
1.2	Diagrama de bloco representando a decimação por um fator M	6
1.3	Operação geral de decimação.	6
1.4	Interpolação por um fator L	7
1.5	Operação geral de interpolação.	7
2.1	Vários sensores capturam imagens de baixa resolução de uma mesma cena.	11
2.2	Uma única câmera captura uma sequência de imagens de uma cena dinâmica.	12
2.3	Exemplo de grades e seus deslocamentos. (a) Imagem original, (b) grade de 8×8 pixels, (c) grade de 16×16 pixels, (d) deslocamento da grade em nível de pixel e (e) deslocamento da grade em nível de subpixel.	14
2.4	Exemplo de sequências de alta-resolução (f) e as observações em baixa-resolução (g).	18
2.5	Exemplo do uso de uma janela na super-resolução de um quadro em uma cena dinâmica.	21
3.1	Exemplo de uso para o sistema tradicional de codificação onde a máquina responsável pela codificação possui disponibilidade de processamento e sem restrições de energia.	30
3.2	Exemplo de uso para o sistema de codificação de complexidade reversa onde a máquina responsável pela codificação possui baixa capacidade de processamento e restrições de energia.	31
3.3	Ilustração de uma sequência de vídeo de (a) resolução original e (b) resolução mista. As imagens que mantêm sua resolução original são chamados quadros chave, enquanto as imagens de resolução reduzida são chamados quadros não-chave.	32
3.4	Diagrama simplificado do sistema de super-resolução proposto mostrando o lado do codificador de complexidade reduzida para o cenário considerado. Dada uma sequência de vídeo, os quadros não-chave são decimados antes da codificação. Este processo pode ocorrer separadamente para os quadros chave e não-chaves.	33
3.5	Diagrama simplificado do sistema de super-resolução proposto mostrando o lado do decodificador em um <i>codec</i> de complexidade reversa considerando um cenário de sequência de vídeo de resolução mista e incluindo o bloco da técnica de super-resolução proposta.	35
3.6	Exemplo ilustrativo dos quadros no decodificador: (a) quadro chave em resolução original (presente no decodificador), (b) quadro não-chave em resolução reduzida (decimado no codificador e enviado ao decodificador) e (c) quadro não-chave com dimensão ampliada (interpolado no decodificador).	36
3.7	Diagrama simplificado no lado do codificador mostrando o cenário considerado com uma sequência de resolução mista formada por quadros não-chave, indicados como $G'[n]$, e por quadros chave, $G[n]$	37
3.8	Diagrama simplificado no lado do decodificador mostrando o cenário considerado com uma sequência de resolução mista, $F'[n]$ para quadros não-chave e $F[n]$ para quadros chave. Os blocos tracejados indicam as operações realizadas em nível de macrobloco.	38

3.9	Informação disponível no lado do decodificador: todas as componentes dos quadros chave e somente as componentes de baixa e média frequência dos quadros não-chave. A estimação de movimento é realizada nas componentes de frequências mais baixas.	41
3.10	Exemplo de quadro chave redimensionado para simular a perda das altas-frequências onde se vê (a) quadro original, (b) quadro decimado e (c) quadro interpolado nitidamente borrado.	42
3.11	Exemplo de alta, média e baixa frequência de um dado quadro. A componente de alta-frequência é extraída do quadro original subtraído do quadro interpolado.	42
3.12	Etapas iniciais no lado do decodificador: (a) quadros não-chave são interpolados e retornam a sua resolução original porém com qualidade inferior, (b) quadros chave sofrem uma degradação simulada sendo sucessivamente redimensionados (decimados e interpolados) pelo mesmo fator M utilizado nos quadros não-chave e (c) quadros chave seguem sem modificações à próxima etapa do processo de super-resolução.....	42
3.13	Exemplo de distribuição de quadros chave e não-chave dentro de uma dada sequência. A frequência de quadros chave é definida por um fator h pré-estabelecido no codificador. A predição de movimento é feita bidirecionalmente sendo realizada uma procura entre um macrobloco do quadro não-chave e os dois quadros chave mais próximos.	44
3.14	Detalhe do diagrama mostrando as etapas de estimação de movimento e posterior super-resolução dos quadros não-chaves (adição de alta-frequência). A saída do sistema ($\{\tilde{X}[n]\}$) retrata os quadros chaves ($\{F[n]\}$) e os quadros não-chaves super-resolvidos ($\{F'_{SR}[n]\}$). Vale lembrar que os blocos tracejados indicam as operações realizadas em nível de macrobloco.	44
3.15	Após a estimação de movimento, a próxima etapa é incluir nos quadros não-chave a alta-frequência dos quadros chave. Em (a) é possível ver a alta-frequência dos quadros chave sendo inserida no quadro não-chave e em (b) o estabelecimento do quadro não-chave super-resolvido recompondo as componentes de alta, média e baixa frequências novamente em um único quadro. É importante observar que o processo (a) ocorre em nível de macrobloco, portanto, o quadro de altas-frequências estimadas possui somente os macroblocos de alta-frequência "vencedores"na estimação de movimento, ou seja, que tinham menor SAD/SSD....	46
3.16	Exemplo da alta-frequência de um quadro chave (a) original e (b) borrado. É possível ver que mesmo as altas-frequências do quadro borrado serem consideravelmente mais suaves, elas ainda se apresentam visíveis.	48
3.17	Detalhe do diagrama destacando o novo estágio de filtragem antes da estimação de movimento. Os quadros chave e não-chave borrados são processados por um filtro passa-altas resultando, respectivamente, nos quadros $F_{HF}[n]$ e $F'_{HF}[n]$	49
3.18	Detalhe ampliado da alta-frequência presente em um quadro chave. É possível ver que somente as altas-frequências mais expressivas do quadro original (a) permanecem perceptíveis no quadro borrado (b). Observe que é possível ver as bordas, texturas e ruídos em (a) porém as texturas e ruídos não são percebidos em (b).	51
3.19	Inclusão do bloco com o limiar T_f que analisa a quantidade de alta-frequência dos quadros não-chave borrados ($F'_{HF}[n]$) resultando na decisão de quais macroblocos devem ou não ser super-resolvidos.	52

3.20	Exemplo do comportamento do limiar T_f . A linha do eixo horizontal retrata o limite inferior de qualidade estabelecido pela qualidade do quadro borrado. O ponto de máximo observado nesse caso é aproximadamente $T_f = 0,02$	53
3.21	Histograma mostrando os valores normalizados da SSD para as melhores combinações entre blocos dos quadros chave e não-chave para as sequências (a) <i>Foreman</i> , (b) <i>Coastguard</i> , (c) <i>Mobile</i> e (d) <i>Mother and Daughter</i>	55
3.22	Histograma mostrando os valores normalizados da SSD para as melhores combinações entre blocos dos quadros chave e não-chave para as sequências (a) <i>Akyio</i> e (b) <i>Football</i>	56
3.23	Detalhe do diagrama destacando o filtro sigma. A alta-frequência dos quadros chave passam por uma filtragem antes de serem adicionadas ao quadro não-chave dependendo da decisão tomada baseada no valor da SSD normalizada e no limiar T_σ	58
3.24	Exemplo de quadro de alta-frequência $F'_{HC}[n]$	60
3.25	Exemplo de quadro de alta-frequência processada por um filtro sigma $F'_{TC}[n]$	60
3.26	Diagrama de blocos completo no lado do codificador.	61
3.27	Diagrama de blocos completo no lado do decodificador incluindo as etapas de refinamento do método de super-resolução proposto. Os blocos tracejados do diagrama representam as operações realizadas em nível de macrobloco.	62
4.1	Sequências de vídeo: (a) <i>Foreman</i> , (b) <i>Coastguard</i> , (c) <i>Mobile</i> , (d) <i>Mother and Daughter</i> , (e) <i>Akyio</i> e (f) <i>Football</i>	66
4.2	Gráfico com os resultados variando a densidade de quadros chave para a sequência <i>Foreman</i>	69
4.3	Gráfico com os resultados variando a densidade de quadros chave para a sequência <i>Coastguard</i>	69
4.4	Gráfico com os resultados variando a densidade de quadros chave para a sequência <i>Mobile Calendar</i>	70
4.5	Gráfico com os resultados variando a densidade de quadros chave para a sequência <i>Mother and Daughter</i>	70
4.6	Gráfico com os resultados variando a densidade de quadros chave para a sequência <i>Akyio</i>	71
4.7	Gráfico com os resultados variando a densidade de quadros chave para a sequência <i>Football</i>	71
4.8	Gráfico com os resultados variando o QP na codificação dos quadros para a sequência <i>Foreman</i>	74
4.9	Gráfico com os resultados variando o QP na codificação dos quadros para a sequência <i>Coastguard</i>	74
4.10	Gráfico com os resultados variando o QP na codificação dos quadros para a sequência <i>Mobile Calendar</i>	75
4.11	Gráfico com os resultados variando o QP na codificação dos quadros para a sequência <i>Mother and Daughter</i>	75
4.12	Gráfico com os resultados variando o QP na codificação dos quadros para a sequência <i>Akyio</i>	76
4.13	Gráfico com os resultados variando o QP na codificação dos quadros para a sequência <i>Football</i>	76
4.14	Quadro 6 da sequência <i>Foreman</i> : (a) original, (b) interpolado e (c) super-resolvido.	78

4.15	Quadro 6 da sequência <i>Coastguard</i> : (a) original, (b) interpolado e (c) super-resolvido.	79
4.16	Quadro 4 da sequência <i>Mobile Calendar</i> : (a) original, (b) interpolado e (c) super-resolvido.	80
4.17	Quadro 4 da sequência <i>Football</i> : (a) original, (b) interpolado e (c) super-resolvido.	81

LISTA DE TABELAS

3.1	Tabela de índices de correlação entre o primeiro e os outros quadros (até o sétimo quadro) dentro de cada uma das sequências de vídeo. Correlações consideradas sem qualquer estimacão de movimento.	40
3.2	Tabela de índices de correlação entre o primeiro e os outros quadros (do oitavo ao décimo quinto quadro) dentro de cada uma das sequências de vídeo. Correlações consideradas sem qualquer estimacão de movimento.	40
3.3	Tabela exemplificando o ganho de qualidade objetiva (PSNR) nos quadros super-resolvidos em diversas sequências quando é feita a estimacão de movimento na alta-frequência dos quadros chave e não-chave ao invés de realizar a estimacão na baixa-frequência.	49
3.4	Tabela exemplificando os resultados de qualidade objetiva ao fazer super-resoluçã utilizando cada uma das métricas para estimacão de movimento (SAD ou SSD). Perceba que as duas técnicas produzem resultados muito próximos e, portanto, não é concluído um melhor método para esta etapa.	56
3.5	Tabela ilustrando o efeito do uso de diversos valores para o limiar T_{ssd}	57
3.6	Tabela exemplificando os resultados de qualidade objetiva ao fazer super-resoluçã sem e com filtro sigma.	59

LISTA DE SIGLAS, ABREVIACOES E ACRONIMOS

Abreviações, Acrônimos e Siglas

1D	Uma dimensão
2D	Duas dimensões
3D	Três dimensões
AVC	<i>Advanced Video Coding</i>
bps	<i>bits</i> por segundo
CIF	<i>Common Intermediate Format</i>
CODEC	Codificador e Decodificador
decod	decodificador
DSC	<i>Distributed Source Coding</i>
DVC	<i>Distributed Video Coding</i>
DVD	<i>Digital Versatile Disc</i>
EM	Estimação de Movimento
encod	codificador
FPS	<i>Frames per Second</i>
H.261	Padrão de Compressão de vídeo
H.263	Padrão de Compressão de vídeo
H.263+	Padrão de Compressão de vídeo (versão melhorada do H.263)
H.264/AVC	Padrão de Compressão de vídeo (Advanced Video Coding)
H.264/SVC	Padrão de Compressão de vídeo (Scalable Video Coding)
H.26L	Codificador implementado entre os padrões H.263+ e H.264/AVC
H.26X	Padrões de compressão de vídeo da <i>International Telecommunication Union</i>
ISO	<i>International Organisation for Standardisation</i>
ITU	<i>International Telecommunication Union</i>
JM	<i>Joint Model</i>
JSMV	<i>Joint Scalable Video Model</i>
Kbps	Kilo <i>bits</i> por segundo
KF	<i>Key Frame</i>
Mbps	Mega <i>bits</i> por segundo
MC	<i>Motion Compensation</i>
ME	<i>Motion Estimation</i>
MPEG	<i>Motion Picture Experts Group</i>
MPEG-1	Padrão de Compressão de vídeo
MPEG-2	Padrão de Compressão de vídeo
MPEG-4	Padrão de Compressão de vídeo
MSE	<i>Mean Square Error</i>
MV	<i>Motion Vector</i>
NKF	<i>Non-Key Frame</i>
pixel	<i>picture element</i>
PSNR	<i>Peak Signal to Noise Ratio</i>
QCIF	<i>Quarter Common Intermediate Format</i>
RD	<i>Rate-Distortion</i>
SAD	<i>Sum of Absolute Differences</i>
SI	<i>Side Information</i>
SNR	<i>Signal to Noise Ratio</i>
SSD	<i>Sum of Squared Differences</i>
SVC	<i>Scalable Video Coder</i>
TV	Televisão

1 INTRODUÇÃO

1.1 VÍDEO E COMPRESSÃO DE VÍDEO

Uma cena de vídeo natural é espacial e temporalmente contínua. A digitalização de um vídeo requer a amostragem temporal, que divide o vídeo em uma sequência de imagens estáticas (quadros ou *frames*), e a amostragem espacial (que divide cada imagem em pontos, denominados *picture elements* ou simplesmente *pixels*). A digitalização é exemplificada na Figura 1.1. Cada *pixel* é representado como um número ou conjunto de números que descreve o brilho (luminância) e a cor (crominância) dessa amostra. A quantidade de *pixels* em cada imagem define a resolução espacial do vídeo, enquanto a quantidade de quadros por segundo define a resolução temporal.

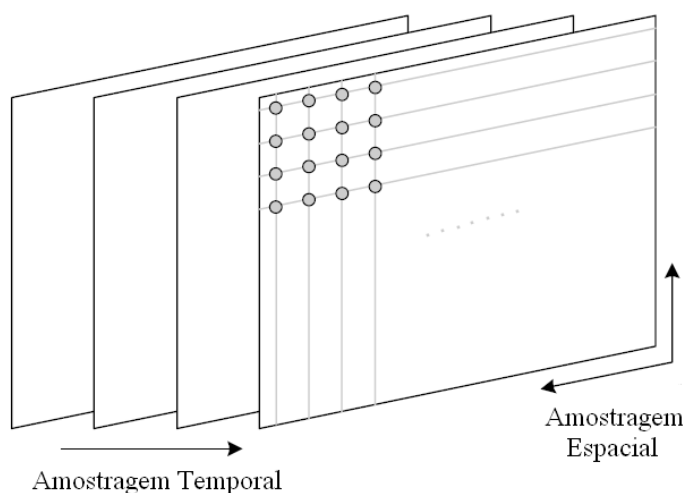


Figura 1.1: Amostragem temporal e espacial de uma sequência de vídeo.

A amostragem típica de um vídeo em DVD, por exemplo, é de 720×576 *pixels* para cada imagem e 25 quadros registrados/exibidos por segundo. Nessas condições, um filme de 2 horas ocuparia aproximadamente 224 *gigabytes* considerando que cada *pixel* possui uma profundidade de 24 *bits*. Como a mídia de DVD suporta, no máximo, 17 *gigabytes* (duplo lado e dupla camada de gravação), seriam necessárias 14 mídias para armazenar um vídeo de 2 horas nesse formato, sem qualquer tipo de compressão.

Esse fato exposto eliminaria o uso prático de vídeo digital. Para utilizar vídeos em formato

digital, foi necessário utilizar técnicas de compressão de sinais, o que significa representar uma informação em um menor número de bits. Os padrões de compressão de vídeo atuais mais conhecidos são H.26x [1, 2] e MPEG-x [3, 4] por conseguirem taxas de compressão bem elevadas. O padrão H.264/AVC [2] especificamente é, no momento, o estado-da-arte em codificação de vídeo.

Na compressão de um vídeo, explora-se a correlação espacial (a semelhança entre *pixels* vizinhos em cada imagem) e a correlação temporal (a semelhança entre quadros adjacentes) a fim de se obter um melhor desempenho de compressão, ou seja, a melhor relação taxa/distorção.

Atualmente, com a imensa capacidade de compressão de dados e com o aumento significativo das bandas de transmissão e dos espaços de armazenamento, pode-se imaginar que os obstáculos dentro da área de compressão de imagens venham se esvaindo. Afirmação essa que de fato não é verdade pois, apesar dos algoritmos de compressão estarem cada dia mais eficientes, a quantidade de dados a ser comprimida é maior a cada dia. Este fato é agravado se for considerado o grande avanço das tecnologias de captura de vídeo e sua convergência para dispositivos portáteis cada vez mais acessíveis e populares resultando no crescimento exponencial da produção de material multimídia. Juntamente com o avanço dessas tecnologias, surgiu uma nova necessidade dentro do campo na compressão direcionada a esses equipamentos com menor capacidade de armazenamento, processamento e limitações no que diz respeito a recursos energéticos.

1.2 CONTEXTUALIZAÇÃO

A resolução de uma imagem, como mencionado anteriormente, é a quantidade de *pixels* que representa uma dada cena. Quanto maior a resolução espacial, melhor a qualidade da imagem no que concerne a representação dos detalhes capturados da cena. Portanto, quanto maior a resolução espacial de um vídeo, mais detalhamento e acurácia em relação à cena original são exibidos. Uma problemática quanto ao aumento da resolução engloba a elevação significativa da quantidade de dados a serem armazenados e/ou enviados na transmissão, esbarrando, assim, em limitações de disponibilidade de espaço e de banda. Uma alternativa para contornar essas limitações é utilizar todos os quadros com resolução mais baixa (consequentemente perdendo em detalhamento) ou

utilizar somente alguns quadros com grande resolução espacial e os outros quadros da sequência de vídeo com resolução espacial reduzida, desse modo, preservando o máximo detalhamento em alguns dos quadros.

1.2.1 Codificação de Vídeo

Embora os sistemas de codificação (*codecs*) atuais tenham desempenhos diferentes, todos tem uma estrutura muito similar: utilizam transformadas para explorar a redundância espacial, técnicas de estimação e compensação de movimento para explorar a redundância temporal, e por fim técnicas de codificação de entropia para codificar os resíduos (diferença entre o quadro estimado e o real) da melhor maneira possível. Todos esses procedimentos são aplicados no codificador, restando ao decodificador apenas a operação inversa de cada um desses procedimentos. De maneira geral, o codificador é de 5 a 10 vezes mais complexo que o decodificador [5, 6, 7, 8].

Estes sistemas tradicionais de codificação se apóiam neste paradigma onde o codificador é um sistema consideravelmente mais complexo que o decodificador. Esse comportamento retrata justamente a realidade onde vídeos são codificados uma primeira vez em um equipamento com grande capacidade de recursos e, posteriormente, transmitidos ou disponibilizados para futuras decodificações em outros dispositivos.

Considerando que dispositivos móveis não possuem recursos ilimitados para codificar vídeos, deve-se considerar suas restrições para escolher um sistema de codificação mais apropriado. Um novo paradigma chamado Codificação Distribuída de Vídeo (DVC, do inglês *Distributed Video Coding*) [9] permite a reversão da complexidade do codificador de vídeo, ou seja, transferir parte da complexidade do codificador para o decodificador sem perda de desempenho com relação às técnicas tradicionais.

Codificadores de complexidade reversa [10, 11] se mostram então mais adequados ao uso nos dispositivos móveis pois acompanham as limitações mencionadas anteriormente. Porém, uma das questões levantadas é como diminuir ainda mais a complexidade no codificador. Uma das respostas é reduzir ou até eliminar etapas que mais demandem recursos.

Diferentemente dos codificadores tradicionais, que utilizam quadros obtidos através de

estimação e compensação de movimento como informação adicional para codificar o quadro atual, os codificadores DVC usam a informação temporal apenas no decodificador. Assim, em muitas implementações, o processo de estimação de movimento é deslocado do codificador para o decodificador, levando consigo grande parte da complexidade do codificador, uma vez que este processo é responsável por grande parte do esforço computacional [6].

1.2.2 Estimação de Movimento

O processo de estimação e compensação de movimento parte do princípio de que um quadro pode ser modelado como a translação de quadros anteriores [12]. O processo de estimação de movimento consiste basicamente na procura em quadros já codificados de um macrobloco que melhor aproxime o macrobloco que está sendo codificado. A informação de movimento é representada por um vetor de deslocamento bidimensional chamado vetor de movimento.

Devido a representação do quadro em macroblocos, o processo de estimação de movimento utiliza técnicas de casamento de blocos, onde o vetor de movimento é encontrado minimizando uma função que calcula o custo da escolha de um determinado vetor de movimento. Muitas medidas de custo foram introduzidas na literatura [13] contudo, uma das métricas mais difundidas até mesmo pela simplicidade de cálculo é a Soma das Diferenças Absolutas (SAD, do inglês *Sum of Absolute Differences*). A SAD é baseada na norma L_p que é calculada da seguinte forma:

$$\|x\|_p = \left(\sum_i \sum_j |(x_{ij} - x'_{ij})|^p \right)^{1/p} \quad (1.1)$$

Para o caso da SAD especificamente (norma L_1), define-se:

$$SAD(B_{i,j}) = \sum_k \sum_l |B_{i,j}(k, l) - B_{i-u, j-v}(k, l)| \quad (1.2)$$

Onde $B_{i,j}(k, l)$ representa o *pixel* (k, l) de um macrobloco $B_{i,j}$ cujo primeiro pixel se encontra na posição (i, j) , e $B_{i-u, j-v}(k, l)$ representa o *pixel* (k, l) de um macrobloco candidato de um quadro de referência na posição (k, l) transladado por um vetor de movimento (u, v) .

Para encontrar o vetor de movimento que produz menor erro é necessário calcular a SAD para macroblocos em várias posições. O modo ótimo de se encontrar o melhor vetor de movimento é

procurar todas as posições dentro do quadro e até mesmo estendendo a busca para mais de um quadro, porém este processo de procura é extremamente custoso. Uma alternativa para reduzir o esforço computacional consiste em utilizar buscas sub-ótimas [14, 15] e limitar a procura a uma janela pré-definida. Geralmente, as buscas são feitas nos quadros de referência mais próximos (*a priori*, anterior e posterior) aumentando a probabilidade de encontrar um macrobloco mais semelhante àquele em questão.

1.2.3 Efeitos da Filtragem no Redimensionamento de Imagens

Ao codificar um quadro, quanto maior sua resolução espacial, certamente mais operações computacionais serão envolvidas para comprimi-lo e maior taxa de transmissão será requerida. Assim, codificar um quadro de menor dimensão pode diminuir a taxa e reduzir a quantidade de operações computacionais no processo de codificação.

Dependendo da aplicação, mesmo que se reduza previamente a dimensão da imagem para economizar recursos, deve-se exibir a imagem final preferivelmente em sua dimensão e resolução originais exigindo então que haja pós-processamentos para redimensionar e recuperar a resolução original da imagem.

Muitas vezes, os termos *dimensão* e *resolução* da imagem são utilizados sem distinção mas vale lembrar que a dimensão de uma imagem é o número de *pixels* presentes na imagem e a resolução de uma imagem é a quantidade de informação contida na cena. Portanto, aumentar a dimensão de uma imagem não necessariamente aumenta a resolução da mesma.

Para diminuir ou aumentar a amostragem de um dado sinal são utilizados processos de sub- e super-amostragem, respectivamente. As diversas técnicas de decimação (sub-amostragem) e interpolação (super-amostragem) de imagens são objeto de estudo na área de processamento de sinais e são apresentadas amplamente na literatura especializada.

A decimação [16, 17] de um sinal digital $x(m)$ por um fator M é o mesmo que reduzir sua taxa de amostragem em M vezes equivalendo a manter somente uma a cada M amostras do sinal. Essa operação é representada na Figura 1.2 a seguir.

A relação entre o sinal decimado e o sinal original é portanto:

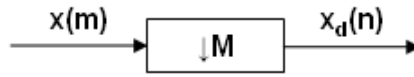


Figura 1.2: Diagrama de bloco representando a decimação por um fator M .

$$x_d(n) = x(nM) \quad (1.3)$$

Analisando o sinal no domínio da frequência, observa-se que o sinal $x_d(n)$ é composto de cópias do espectro de $x(m)$ expandidas por M e depois repetidas com período 2π . Isso também equivale a cópias do espectro de $x(m)$ repetidas com período $2\pi/M$ e depois expandidas por M . Isso implica que, a fim de evitar *aliasing* [18] após a decimação, a largura de faixa do sinal $x(m)$ tem que ser limitada ao intervalo $[-\pi/M, \pi/M]$. Portanto, a operação de decimação é geralmente precedida por um filtro passa-baixas (veja Figura 1.3) que aproxima a seguinte resposta em frequência:

$$H_d(e^{jw}) = \begin{cases} 1, & \text{se } w \in [-\frac{\pi}{M}, \frac{\pi}{M}]; \\ 0, & \text{caso contrário.} \end{cases}$$

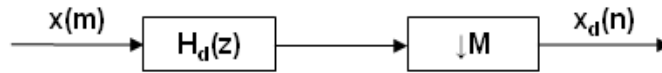


Figura 1.3: Operação geral de decimação.

Ao incluir a operação de filtragem, o sinal decimado é obtido como uma a cada M amostras da convolução do sinal $x(m)$ com resposta ao impulso $h_d(m)$ do filtro, ou seja:

$$x_d(n) = \sum_{m=-\infty}^{\infty} x(m)h_d(nM - m) \quad (1.4)$$

O processo de interpolação [16, 17] ocorre de modo semelhante ao descrito para a decimação. Ao interpolar um sinal, é aumentada a taxa de amostragem de um sinal $x(m)$ por um fator L , incluindo $L - 1$ zeros entre cada duas amostras. Essa operação é representada na Figura 1.4 e na relação abaixo:

$$\hat{x}_i(n) = \begin{cases} x(n/L), & \text{se } n = kL, \text{ com } k \text{ inteiro;} \\ 0, & \text{caso contrário.} \end{cases}$$

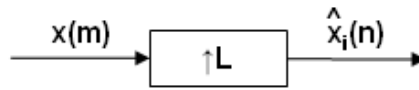


Figura 1.4: Interpolação por um fator L .

Analisando no domínio da frequência, como o espectro do sinal digital original é periódico com período 2π , o espectro do sinal interpolado tem período $2\pi/L$. Portanto, a fim de se obter uma versão interpolada suave de $x(m)$, o espectro do sinal interpolado precisa ter a mesma forma do espectro de $x(m)$. Isso pode ser obtido pela eliminação das repetições do espectro de $\hat{x}_i(n)$ situadas fora do intervalo $[-\pi/M, \pi/M]$. Logo, a operação de interpolação é geralmente seguida de um filtro passa-baixas (veja Figura 1.5) que aproxima a seguinte resposta na frequência:

$$H_i(e^{jw}) = \begin{cases} L, & \text{se } w \in [-\frac{\pi}{M}, \frac{\pi}{M}]; \\ 0, & \text{caso contrário.} \end{cases}$$



Figura 1.5: Operação geral de interpolação.

A operação de interpolação é equivalente à convolução da resposta ao impulso do filtro de interpolação, $h_i(n)$, com o sinal $\hat{x}_i(n)$. Considerando que as únicas amostras não-nulas de $\hat{x}_i(n)$ são aquelas cujo índice é múltiplo de L , pode-se reescrever a relação como:

$$\hat{x}_i(kL) = \begin{cases} x(k), & \text{se } k \text{ for inteiro;} \\ 0, & \text{caso contrário.} \end{cases}$$

Com auxílio da equação anterior, é possível perceber que no domínio do tempo, o sinal interpolado filtrado se torna:

$$x_i(n) = \sum_{m=-\infty}^{\infty} \hat{x}_i(m)h(n-m) = \sum_{m=-\infty}^{\infty} x(k)h(n-kL) \quad (1.5)$$

Como se pode perceber, ambos processos de decimação e de interpolação impõe ao sinal de entrada filtragens passa-baixas. Desse modo, mesmo que se queira diminuir a complexidade de um sistema de codificação simplesmente reduzindo a resolução dos quadros, deve-se observar

que ao final dos processos, as imagens terão perdido as componentes de frequência mais alta prejudicando a qualidade das mesmas. A perda da alta-frequência em uma imagem, causa a suavização das bordas e dos detalhes causando um fenômeno aqui referenciado como borramento (*blurring* em inglês).

1.2.4 Métricas de Avaliação de Qualidade de Imagem

A codificação de um vídeo é usualmente realizada aceitando-se perdas entre o vídeo codificado e o vídeo original a fim de se obter maiores taxas de compressão. Dessa forma, faz-se necessária a definição de uma métrica para precisar essa distorção.

Em sistemas de vídeo digital utiliza-se comumente uma medida chamada PSNR (do inglês, *peak signal to noise ratio* [8] que mede a relação entre o sinal e o ruído de pico). A formulação da PSNR em decibéis é dada por:

$$PSNR_{dB} = 10 \cdot \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (1.6)$$

$$MSE = \frac{1}{N} \cdot \sum_{i=1}^N (x(i) - y(i))^2 \quad (1.7)$$

onde MSE indica o erro médio quadrático entre a imagem comprimida (y) e a imagem original (x), e o numerador da Equação 1.6 é o valor máximo que um *pixel* pode assumir nessa imagem, dependente da profundidade dos *pixel*, ou seja, quantos *bits* são gastos para representar um *pixel*.

Outra medida utilizada para avaliar uma imagem é a percepção subjetiva de qualidade. Esta avaliação se refere basicamente à como uma pessoa percebe a diferença entre as diferentes imagens expostas a ela. A percepção pode variar muito de acordo com a aplicação, com a atenção que se dá à imagem, com as experiências pessoais do observador, e muitos outros fatores, fazendo com que os resultados de uma medida subjetiva não possam ser repetidos nem medidos com precisão. De todo modo, é ainda uma medida a se considerar dado que o resultado final deve ser agradável aos olhos do expectador, independente da PSNR alcançada.

1.3 DEFINIÇÃO DO PROBLEMA E OBJETIVOS

O objetivo deste projeto é elaborar um sistema de super-resolução que consiga eficientemente recuperar a informação de alta-frequência nos quadros de baixa-resolução. Para tanto, é proposto utilizar quadros chave em resolução normal. Um cenário onde isto pode ser aplicado é em esquemas de DVC [9].

Dado que um dos propósitos dos sistemas de codificação de complexidade reversa é reduzir o esforço computacional no lado do codificador, alguns dos quadros em uma dada sequência de vídeo são sub-amostrados previamente à codificação. Isso é possível pois esse sistema em questão possui escalabilidade espacial. Porém, como visto na Seção 1.2.3, existe um efeito colateral desse redimensionamento ocasionando a perda da banda de alta-frequência dos quadros. Isso de fato gera perda de qualidade tanto objetiva como também, certamente, subjetiva pois as imagens decimadas/interpoladas resultantes ficam consideravelmente borradas.

Uma das alternativas para tentar solucionar este problema se baseia em técnicas de recuperação do sinal visual com objetivo de obter imagens com resolução mais alta com boa qualidade a partir de imagens de mais baixa resolução. Esse conjunto de técnicas é chamado de super-resolução. O Capítulo 2 se dedica a discutir essas técnicas mais a fundo portanto não serão fornecidos detalhes técnicos neste momento.

A base para este projeto é um cenário existente onde o *codec* é distribuído, escalável e de complexidade reversa. Este *codec* trabalha com uma dada sequência de vídeo de forma que alguns dos quadros, chamados *quadros chave*, são codificados normalmente em sua resolução original. Por outro lado, os outros quadros, designados como *quadros não-chave*, são redimensionados (decimados) para uma resolução reduzida antes da codificação. No decodificador, quadros chave e não-chave são decodificados porém os quadros não-chave são novamente redimensionados (interpolados) para retornarem novamente à sua dimensão original.

Considerando que quadros não-chave perdem suas componentes de alta-frequência durante os processos de decimação e interpolação (quando utilizadas as filtragens da Seção 1.2.3), o objetivo principal, a partir deste momento, é tentar recuperá-las após a decodificação. Como é possível contar com a presença dos quadros chave, é de grande valia explorar a informação

contida neles para ajudar na estimação da alta-frequência dos quadros borrados (não-chave) dado que quadros chave a princípio possuem ainda todas as componentes de frequência. A utilização dos quadros chave como fonte de referência intuitivamente é válida pela alta redundância de informação temporal contida entre quadros vizinhos.

O objetivo final deste projeto, então, é prover um método de super-resolução em um cenário onde se faça presente uma sequência de vídeo de resolução mista [19], ou seja, quadros com diferentes resoluções, tentando recuperar a qualidade da imagem dos quadros de menor resolução com o auxílio dos quadros de maior resolução.

1.4 APRESENTAÇÃO DO MANUSCRITO

Este documento está dividido em cinco capítulos. Este primeiro Capítulo tenta ambientar o leitor dentro do assunto a ser abordado e faz uma sucinta introdução a alguns dos temas na área de processamento de imagens que são mencionados posteriormente durante a apresentação do trabalho. No Capítulo 2 é feita uma revisão bibliográfica sobre alguns métodos de super-resolução conhecidos atualmente. Já no Capítulo 3 é exposto em mais detalhes o cenário motivacional, o método proposto de super-resolução e apresenta também alguns refinamentos aplicados à técnica desenvolvida. O Capítulo 4 é reservado para a exposição dos resultados objetivos e subjetivos, e apresenta também algumas análises pertinentes a eles. E, finalmente, no Capítulo 5 são feitas algumas considerações finais, exposição dos trabalhos realizados e sugestões para trabalhos futuros.

2 SUPER-RESOLUÇÃO

É usado o termo super-resolução para descrever processos de obtenção de imagens de alta-resolução a partir de um conjunto de observações de baixa-resolução [20]. Este conjunto de imagens de baixa-resolução pode ser formado por imagens decimadas ou adquiridas por múltiplos sensores capturando uma mesma cena durante um certo período de tempo. Para cenas estáticas (veja Figura 2.1) as observações são relacionadas por deslocamentos globais em nível de *subpixel* (geralmente ocorrendo devido a posições relativas das câmeras ou movimento do próprio sensor), enquanto para cenas dinâmicas (veja Figura 2.2) elas são relacionadas a deslocamentos de *subpixel* devido a movimento local dos próprios objetos juntamente com deslocamentos globais. Em ambos casos, o objetivo da super-resolução é utilizar o conjunto de imagens de baixa-resolução ou de quadros numa sequência de vídeo para gerar uma imagem com maior resolução espacial.

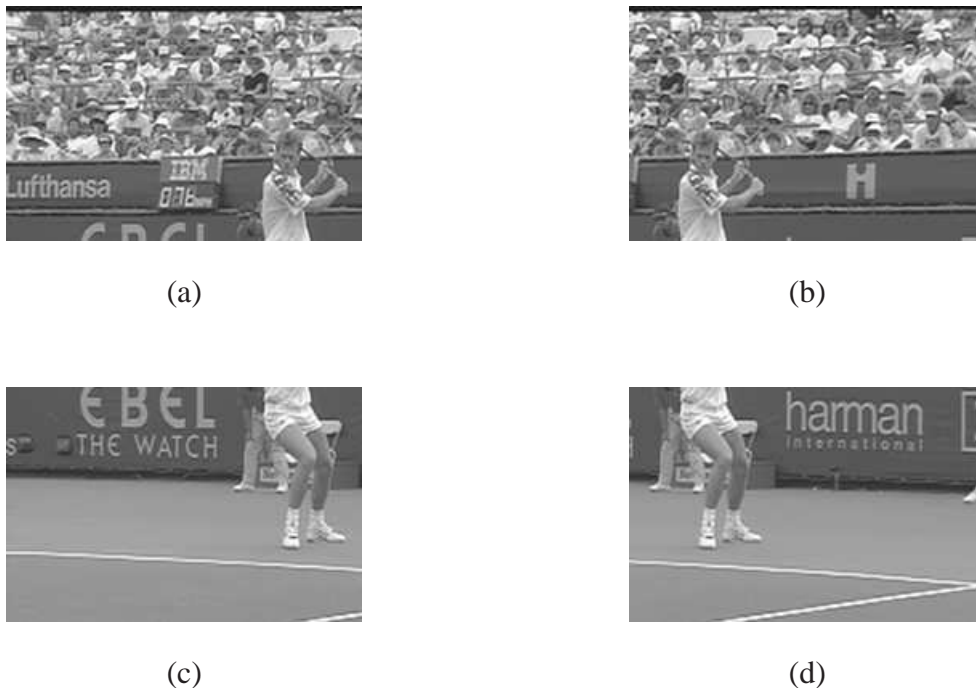


Figura 2.1: Vários sensores capturam imagens de baixa resolução de uma mesma cena.

Aumentar a resolução do sensor de imagem é claramente um modo de aumentar a resolução



(a)



(b)



(c)

Figura 2.2: Uma única câmera captura uma sequência de imagens de uma cena dinâmica.

da imagem adquirida. No entanto, essa solução pode não ser factível pois além de aumentar o custo associado do dispositivo, há o fato do ruído de aquisição ser inversamente proporcional ao "tamanho" do *pixel* [21]. Ainda, aumentar o tamanho do *chip* para acomodar a grande quantidade de *pixels* aumenta também a capacitância o que reduz a taxa de transferência do dispositivo. Como visto, alterar o *hardware* envolvido na captura provoca inevitavelmente dificuldades técnicas e alto custo material.

Uma alternativa para aumentar a resolução de imagens é utilizar super-resolução. O que torna a super-resolução viável são justamente os deslocamentos em nível de *subpixel*. Esses deslocamentos são ilustrados na Figura 2.3 para melhor entendimento da questão.

A Figura 2.3(a) mostra a imagem sem a grade. As Figuras 2.3(b) e (c) mostram as grades de 8x8 e 16x16 *pixels*, respectivamente, sobrepondo a mesma região exposta em (a). É possível perceber que uma grade com mais *pixels* englobando uma mesma área física (como visto na Figura 2.3(c)) permite produzir imagens com maior resolução espacial. Agora considere o caso de possuir duas grades de 8x8 *pixels* cobrindo uma mesma imagem. Dois arranjos possíveis são mostrados nas Figuras 2.3(d) e (e). O primeiro arranjo ilustrado na Figura 2.3(d) mostra a grade deslocada por um número inteiro de *pixels*. Já no segundo arranjo mostrado na Figura 2.3(e), o deslocamento se dá no nível de *subpixel*, ou seja, um deslocamento por número não-inteiro. Este último tipo de deslocamento promove informação adicional sobre a cena permitindo o aumento da resolução.

Intuitivamente, cada imagem de baixa-resolução representa uma versão sub-amostrada da cena original. Considerando que na obtenção das imagens de baixa-resolução é gerado *aliasing* [18] pelo não tratamento prévio por meio de filtragem (como visto no Seção 1.2.3), não seria possível remover tal fenômeno caso houvesse somente uma imagem disponível. No entanto, dado que diversas imagens de baixa-resolução estão disponíveis e as mesmas possuem deslocamentos relativos de *subpixel* então é possível eliminar o *aliasing* (se existente) através da informação complementar obtida dessas imagens. Com o exato conhecimento dos deslocamentos, as imagens observadas podem ser combinadas para remover o *aliasing* e gerar uma imagem com resolução mais alta. Se for assumido que a resolução da imagem de alta-resolução é tal que atenda ao critério de amostragem de *Nyquist* [18], esta imagem então retrata uma representação acurada da

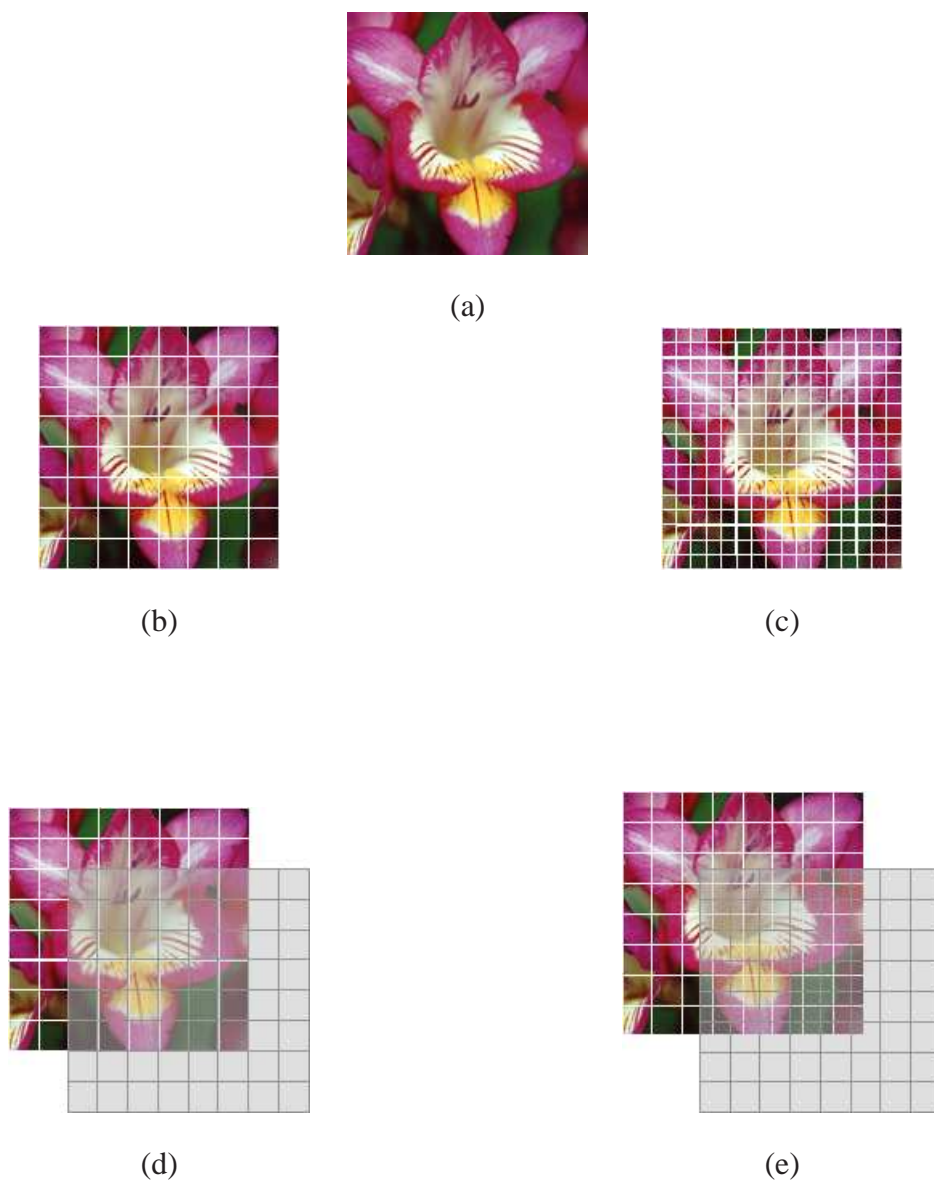


Figura 2.3: Exemplo de grades e seus deslocamentos. (a) Imagem original, (b) grade de 8x8 *pixels*, (c) grade de 16x16 *pixels*, (d) deslocamento da grade em nível de *pixel* e (e) deslocamento da grade em nível de *subpixel*.

cena original. Em outras palavras, informações com *aliasing* podem ser combinadas para obter uma imagem livre de *aliasing*.

A seguir são apresentados os motivos da super-resolução ser um método possível e plausível [22] na recuperação de sinais. Por simplicidade será utilizado como exemplo um sinal unidimensional (1D) sem qualquer tipo de ruído incluso. Considere o sinal 1D contínuo $f(x)$ e p versões dele, cada um deslocado por um valor δ_k , $k = 0, \dots, p - 1$, ou seja,

$$f_k(x) = f(x + \delta_k), \quad (2.1)$$

com $\delta_0 = 0$.

Considere agora a versão amostrada de f_k , isto é,

$$f_k[n] = f(nT + \delta_k), \quad (2.2)$$

onde $n = 0, \dots, N - 1$, e T representa o período de amostragem. Ou seja, tem-se p sinais de baixa-resolução e cada um de tamanho N .

No domínio da transformada contínua de Fourier (FT, do inglês *Fourier Transform*) [18], a Equação 2.1 resulta em:

$$F_k(w) = e^{j2\pi\delta_k w} F(w), \quad (2.3)$$

onde $F_k(w)$ e $F(w)$ são os sinais transformados de $f_k(x)$ e $f(x)$, respectivamente.

Considerando agora a transformada discreta de Fourier (DFT, do inglês *Discrete Fourier Transform*) [18], denota-se por $F_k[l]$ a DFT de N pontos da k -ésima imagem amostrada $f_k[n]$, isto é,

$$F_k[l] = \sum_{n=0}^{N-1} f_k[n] \exp \left[-j2\pi \frac{ln}{N} \right], \quad l = 0, \dots, N - 1. \quad (2.4)$$

Então, a FT de $f_k(x)$ e a DFT de sua versão amostrada são relacionadas pelo teorema da amostragem [23], isto é,

$$\begin{aligned}
F_k[l] &= \frac{1}{T} \sum_{m=-\infty}^{m=\infty} F_k \left(\frac{l}{NT} + m \frac{1}{T} \right) \\
&= \frac{1}{T} \sum_{m=-\infty}^{m=\infty} e^{j2\pi\delta_k(\frac{l}{NT} + m\frac{1}{T})} F \left(\frac{l}{T} \frac{l + mN}{N} \right)
\end{aligned} \tag{2.5}$$

para $l = 0, \dots, N - 1$.

Analisa-se agora quantos sinais amostrados diferentes com período de amostragem T se deve ter para recuperar o sinal original f de banda limitada. Desde que f seja limitado em banda e dado o período de amostragem T , pode-se encontrar L tal que,

$$F(w) = 0 \text{ para } |w| \geq L/T \tag{2.6}$$

Note que para evitar *aliasing*, o sinal deveria ter sido amostrado com $T' = T/2L$, então para $L = 1$ deveria ter sido amostrado a $T/2$ ou para $L = 2$ o período de amostragem deveria ter sido $T/4$ ao invés de T .

Usando a Equação 2.5, cada $F_k[n]$, $k = 0, \dots, p - 1$, $n = 0, \dots, N - 1$, pode ser reescrita como uma combinação de $2L$ amostras de $F(w)$. Note que as mesmas amostras de $F(w)$ são usadas para definir a combinação produzindo $F_k[n]$ com n fixo e $k = 0, \dots, p - 1$. Da discussão acima fica claro que se tem um sistema de pxN equações lineares com $2LxN$ variáveis. Então, o sistema linear pode ser resolvido caso $p = 2L$. As $2LxN$ amostras calculadas de $F(w)$ podem ser usadas agora para estimar $f(x)$ com $x = 0, \dots, (N - 1)T$, com espaçamento $T/2L$ usando $k = 0$ e substituindo N por $2LN$ e T por $T/2L$ na Equação 2.5. A resolução é então aumentada por um fator de $2L$.

O exemplo apresentado anteriormente e sua justificativa teórica foram aplicadas ao cenário mais simples e a uma situação ideal. Existem muitas questões que precisam ser resolvidas em uma situação prática. Por exemplo, os deslocamentos entre as imagens não são exatamente conhecidos então precisam ser estimados a partir dos dados disponíveis. Ainda, tais deslocamentos podem ser mais complicados que uma translação em nível de *subpixel* e o movimento pode ter tanto uma componente global (cena) como local (objetos). E mais, o borramento e os processos de sub-amostragem devem ser levados em consideração juntamente com o ruído envolvido no processo

de observação. No caso de dados comprimidos, informações providas pelo *bitstream* (sequência de *bits*) assim como vetores de movimento e passos de quantização também devem ser utilizados para realizar a super-resolução.

É válido mencionar que a conversão de dados amostrados arbitrariamente para dados espaçados uniformemente é um importante passo no algoritmo. Em outras palavras, um sistema que realize a interpolação dos dados. Aspectos teóricos críticos dos métodos de super-resolução são então relacionados ao resultado da interpolação [24].

A seguir será feita uma breve descrição de um dos métodos mais utilizados no processo de super-resolução: o método Bayesiano.

2.1 SUPER-RESOLUÇÃO BAYESIANA

Para iniciar a discussão sobre super-resolução Bayesiana será promovida uma explanação geral sobre as notações utilizadas a partir desta Seção. Primeiramente, será exposto o caso de imagens não comprimidas como mostrado na Figura 2.4. É denotado pelo vetor \mathbf{f}_k as intensidades ordenadas lexicograficamente do k -ésimo quadro de alta-resolução. Um conjunto de quadros pode ser similarmente ordenado para formar um novo vetor \mathbf{f} . Denota-se pelo vetor \mathbf{g}_k as intensidades ordenadas do k -ésimo quadro de baixa-resolução sub-amostrado a partir da respectiva imagem de alta-resolução, \mathbf{f}_k . Similarmente, o conjunto de quadros de baixa-resolução correspondentes aos quadros de alta frequência em \mathbf{f} pode ser ordenado para formar o vetor \mathbf{g} .

Os quadros do vetor \mathbf{g} são obtidos a partir de \mathbf{f} como já dito anteriormente. É assumido que nesse ponto, os parâmetros requeridos pelo modelo (filtragem, borramento e fator de sub-amostragem etc) para gerar as observações de baixa-resolução a partir das imagens de alta-resolução são conhecidos, exceto pelos parâmetros do modelo de mapeamento entre dois quadros tais como translação, rotação ou qualquer outra relação mais sofisticada. Denota-se pelo vetor $\mathbf{d}_{l,k}$ justamente estes parâmetros do modelo de mapeamento que relaciona dois quadros quaisquer \mathbf{f}_k e \mathbf{f}_l . No contexto de sequências de vídeo dinâmicas (como visto na Figura 2.2), as diferenças entre dois quadros é tipicamente devida ao movimento da câmera e dos objetos em cena. O vetor $\mathbf{d}_{l,k}$ então contém os parâmetros de movimento para compensar o quadro \mathbf{f}_l a partir do quadro \mathbf{f}_k ,

isso significa que cada valor de *pixel* no quadro \mathbf{f}_l pode ser predito através do movimento $\mathbf{d}_{l,k}$ a partir de um pixel no frame \mathbf{f}_k .

Os vetores de movimento podem ter acurácia em nível de subpixel, então os valores de pixel em \mathbf{f}_k precisam ser interpolados ao proceder com a compensação de movimento. O conjunto de vetores $\mathbf{d}_{l,k}$ pode ser ordenado para formar o vetor \mathbf{d} . Este mesmo modelo de notação pode ser aplicada para problemas de super-resolução em imagens estáticas (como visto na Figura 2.1). Neste caso, existe uma só imagem de alta-resolução \mathbf{f} e várias observações de baixa-resolução \mathbf{g}_k , as quais são representadas pelos parâmetros de torção de baixa-resolução ou deslocamentos \mathbf{d} . O problema em cada um dos casos é estimar \mathbf{f} e \mathbf{d} baseado nas observações \mathbf{g} que serão genericamente denotadas por \mathbf{o} .

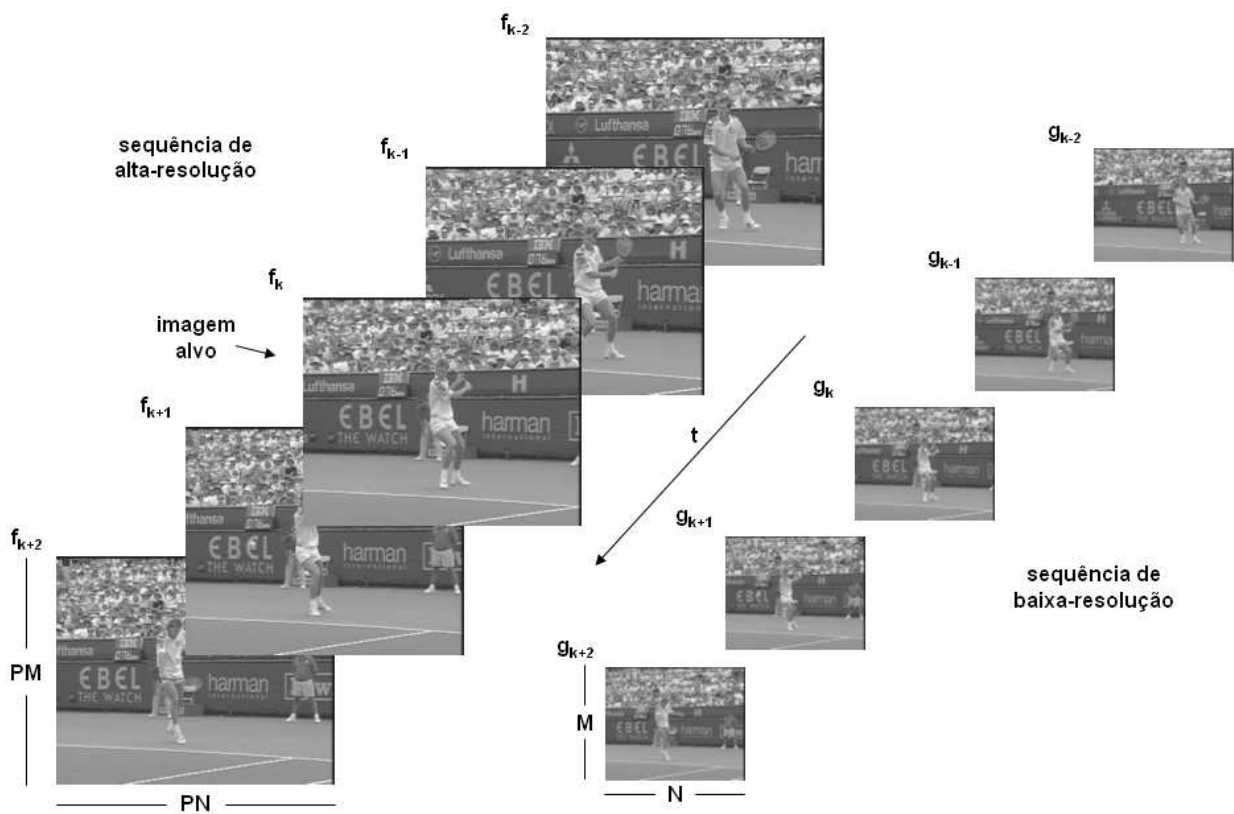


Figura 2.4: Exemplo de sequências de alta-resolução (\mathbf{f}) e as observações em baixa-resolução (\mathbf{g}).

No caso de imagens comprimidas, a situação é ligeiramente diferente da descrita acima. Neste caso, as observações \mathbf{g}_k passam por um processo de compressão gerando as observações comprimidas \mathbf{y}_k . Vetores de movimento são gerados pelo codificador durante a compressão relacionando pares de quadros de baixa-resolução. Tais vetores de movimento de baixa-resolução

são denotados por $\mathbf{v}_{l,k}$ (ou \mathbf{v} se forem ordenados) com a mesma interpretação dos vetores $\mathbf{d}_{l,k}$ mencionados anteriormente. Assim, a observação \mathbf{o} consiste de \mathbf{y} , \mathbf{v} e qualquer outra informação útil para o problema de super-resolução que pode ser encontrada no *bitstream*. O problema é, novamente, estimar \mathbf{f} e \mathbf{d} baseado na observação \mathbf{o} . Este cenário é também aplicável para o caso de imagens estáticas com a interpretação apropriada das várias quantidades mencionadas.

2.1.1 Modelamento Bayesiano

A partir deste momento, será exposto o problema da super-resolução dentro de um cenário Bayesiano. O princípio fundamental da filosofia Bayesiana evoca todos parâmetros e variáveis observáveis como quantidades estocásticas desconhecidas, atribuindo distribuições de probabilidade baseadas em crenças subjetivas. Dessa maneira, em super-resolução a imagem de alta-resolução original \mathbf{f}_k e os vetores de movimento \mathbf{d} são todos tratados como amostras de campos aleatórios com distribuições *a priori* que modelam o conhecimento sobre a natureza da imagem de alta-resolução original e vetores de movimento. A observação \mathbf{o} , que é uma função de \mathbf{f}_k e \mathbf{d} , é também tratada como uma amostra de um campo aleatório com correspondente distribuição condicional que modela o processo para obter \mathbf{o} a partir de \mathbf{f}_k e \mathbf{d} . Essas distribuições dependem dos parâmetros que serão denotados por Ω .

Assumindo que Ω é conhecida, a distribuição conjunta que modela o relacionamento entre os dados observados e as quantidades desconhecidas do problema de super-resolução se torna

$$\mathbf{P}(\mathbf{f}_k, \mathbf{d}, \mathbf{o}) = \mathbf{P}(\mathbf{f}_k, \mathbf{d})\mathbf{P}(\mathbf{o}|\mathbf{f}_k, \mathbf{d}), \quad (2.7)$$

onde $\mathbf{P}(\mathbf{f}_k, \mathbf{d}, \mathbf{o})$ é designado como vizinhança das observações.

2.1.2 Inferência Bayesiana

Uma vez completado o modelamento, a inferência Bayesiana é realizada utilizando a expressão *a posteriori*

$$\mathbf{P}(\mathbf{f}_k, \mathbf{d}|\mathbf{o}) = \frac{\mathbf{P}(\mathbf{f}_k, \mathbf{d})\mathbf{P}(\mathbf{o}|\mathbf{f}_k, \mathbf{d})}{\mathbf{P}(\mathbf{o})}. \quad (2.8)$$

Existem diversos modos de estimar a imagem de alta-resolução e os vetores de movimento utilizando a Equação 2.8. Em sua maneira mais simples, a inferência Bayesiana provê a solução *maximum a posteriori* (MAP), representada pelos valores $\hat{\mathbf{f}}_k$ e $\hat{\mathbf{d}}$, que maximiza a distribuição de probabilidade *a posteriori*, ou seja,

$$\begin{aligned}
\hat{\mathbf{f}}_k, \hat{\mathbf{d}} &= \arg \max_{\mathbf{f}_k, \mathbf{d}} \mathbf{P}(\mathbf{f}_k, \mathbf{d} | \mathbf{o}) \\
&= \arg \max_{\mathbf{f}_k, \mathbf{d}} \left\{ \frac{\mathbf{P}(\mathbf{f}_k, \mathbf{d}) \mathbf{P}(\mathbf{o} | \mathbf{f}_k, \mathbf{d})}{\mathbf{P}(\mathbf{o})} \right\} \\
&= \arg \max_{\mathbf{f}_k, \mathbf{d}} \{ \mathbf{P}(\mathbf{f}_k, \mathbf{d}) \mathbf{P}(\mathbf{o} | \mathbf{f}_k, \mathbf{d}) \} \\
&= \arg \min_{\mathbf{f}_k, \mathbf{d}} \{ -\log \mathbf{P}(\mathbf{f}_k, \mathbf{d}) - \log \mathbf{P}(\mathbf{o} | \mathbf{f}_k, \mathbf{d}) \}. \tag{2.9}
\end{aligned}$$

Dependendo das funções densidade de probabilidade *a priori* e condicional, pode ser difícil encontrar soluções analíticas para a Equação 2.9, logo soluções numéricas são necessárias. Ainda assim, mesmo com as soluções numéricas, um grande problema na otimização é a estimação simultânea de variáveis \mathbf{f}_k e \mathbf{d} . As soluções abordadas na literatura para a reconstrução da imagem super-resolvida e dos vetores de movimento recaem principalmente em metodologias de aproximação e minimização que serão mencionadas em mais detalhes na Seção 2.1.4 deste manuscrito.

2.1.3 Modelando Imagens de Baixa-Resolução

Serão discutidos nessa Subseção os modelos utilizados na literatura para obter as imagens de baixa-resolução observadas a partir da imagem fonte de alta-resolução alvo. A análise da obtenção das imagens de baixa-resolução resultará na determinação de $\mathbf{P}(\mathbf{o} | \mathbf{f}_k, \mathbf{d})$ requisitado pela formulação Bayesiana do problema da super-resolução (veja Equação 2.9) tanto para os casos de cenas estáticas quanto dinâmicas. Como mencionado anteriormente, a sequência de baixa-resolução pode estar comprimida portanto, por simplicidade, será somente apresentado o caso mais direto sem compressão.

Denota-se por $f(x, y, t)$ uma cena dinâmica contínua no tempo e no espaço que está sendo capturada. Se a cena for amostrada de acordo com o critério de Nyquist no tempo e no espaço, ela é representada por uma sequência de alta-resolução $f_l(m, n)$, onde $l = 1, \dots, L$, $m = 0, \dots, PM -$

1, e $n = 0, \dots, PN - 1$, e representam, respectivamente, as coordenadas discretas temporais e espaciais. O parâmetro P é referido como *fator de ampliação*.

Usando notação matricial, cada imagem $PM \times PN$ pode ser transformada em um vetor coluna $(PM \times PN) \times 1$, obtido da ordenação da imagens por linhas (ordenação lexicográfica). O vetor $(PM \times PN) \times 1$ que representa a l -ésima imagem em uma sequência de alta-resolução é designada por \mathbf{f}_l , com $l = 1, \dots, L$. Se ainda todos quadros \mathbf{f}_l , $l = 1, \dots, L$ são lexicograficamente ordenados, o vetor \mathbf{f} de dimensões $(L \times PM \times PN) \times 1$ é obtido.

A sequência de alta-resolução \mathbf{f} é imposta como entrada ao sistema de imageamento que gera as observações de baixa-resolução denotadas por \mathbf{g} como visto na Figura 2.4. O objetivo da super-resolução é obter uma estimativa de um quadro de alta-frequência, \mathbf{f}_k , a partir das observações de baixa-resolução disponíveis. Todas as técnicas baseadas nesse princípio podem ser aplicadas à super-resolução de vídeo usando, por exemplo, uma abordagem de janela móvel como o exemplo ilustrado na Figura 2.5.

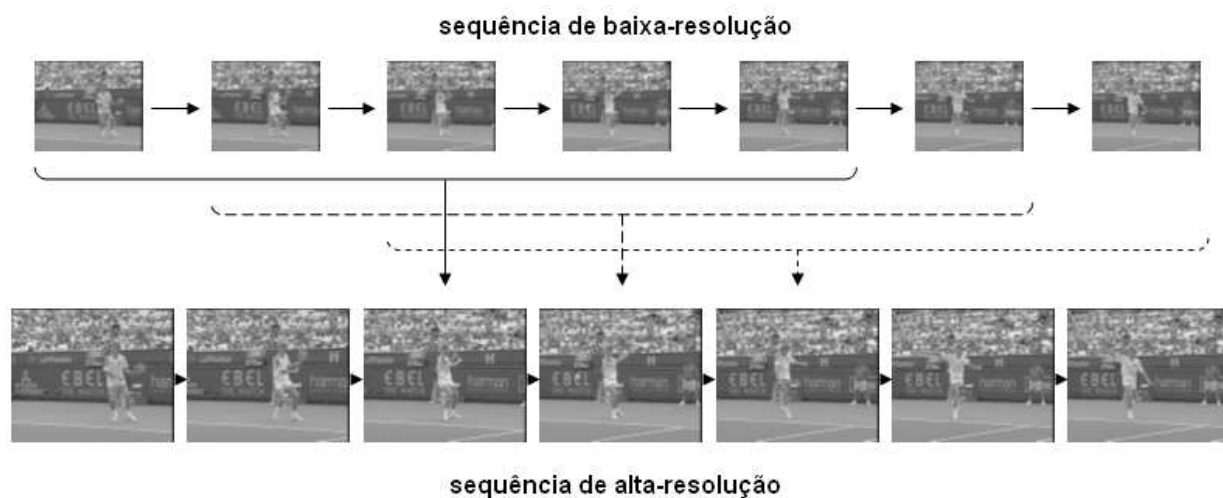


Figura 2.5: Exemplo do uso de uma janela na super-resolução de um quadro em uma cena dinâmica.

Para obter uma estimativa de \mathbf{f}_k , o sistema de imageamento e a relação temporal das sequências de alta- e baixa-resolução precisam ser modeladas e fundidas dentro do esquema Bayesiano.

A procura pelas imagens de alta-resolução f_1, \dots, f_L presumidamente satisfazem

$$f_l(m, n) = f_k(m + d_{l,k}^x(m, n), n + d_{l,k}^y(m, n)), \quad (2.10)$$

onde $d_{l,k}^x(m, n)$ e $d_{l,k}^y(m, n)$ denotam respectivamente as componentes horizontal e vertical do deslocamento, ou seja,

$$\mathbf{d}_{l,k}(m, n) = (d_{l,k}^x(m, n), d_{l,k}^y(m, n)). \quad (2.11)$$

O modelo da Equação 2.10 é razoável sob a pressuposta condição de iluminação constante da cena. Isto leva à estimação do fluxo óptico da cena. Note que este modelo exposto se aplica para ambos tipos de movimentos, global e local. Observe também que podem existir *pixels* em uma imagem sem nenhum vetor de movimento a eles atribuído (problema de oclusão), e *pixels* no qual o deslocamento pode não ser único. É válido lembrar que neste modelo não está sendo considerado o ruído.

Com a notação matricial a Equação 2.10 pode ser reescrita como

$$\mathbf{f}_l = \mathbf{C}(\mathbf{d}_{l,k})\mathbf{f}_k, \quad (2.12)$$

onde $\mathbf{C}(\mathbf{d}_{l,k})$ é a matriz $(PM \times PN) \times (PM \times PN)$ que mapeia o quadro \mathbf{f}_l para o quadro \mathbf{f}_k , e $\mathbf{d}_{l,k}$ é a matriz $(PM \times PN) \times 2$ definida pela ordenação lexicográfica das componentes vertical e horizontal dos deslocamentos entre dois quadros.

O problema da estimação de movimento consiste na estimação de $\mathbf{d}_{l,k}$ ou $\mathbf{C}(\mathbf{d}_{l,k})$ dados \mathbf{f}_l e \mathbf{f}_k . O que faz o problema ainda mais desafiador em super-resolução é o fato de que embora o campo de vetor de movimento de alta-resolução seja necessário, as imagens de alta-resolução não estão disponíveis e, portanto, este campo precisa ser estimado utilizando imagens de baixa-resolução (que também podem estar comprimidas). A acurácia do $\mathbf{d}_{l,k}$ é de suma importância na determinação da qualidade das imagens de alta-resolução procuradas.

Outro importante passo no intuito de prover uma solução para o problema da super-resolução é estabelecer modelos que descrevam a aquisição das imagens de baixa-resolução. Uma relação básica entre as imagens de alta-resolução, \mathbf{f}_l , e suas observações, \mathbf{g}_l , pode ser mostrada como

$$\mathbf{g}_l = \mathbf{A}_l \mathbf{H}_l \mathbf{f}_l + \eta_l, l = 1, \dots, L, \quad (2.13)$$

onde a matriz \mathbf{H}_l de dimensões $(PM \times PN) \times (PM \times PN)$ descreve a filtragem da imagem de alta-resolução, \mathbf{A}_l é a matriz de sub-amostragem de tamanho $M \times N \times (PM \times PN)$, e η_l denota o ruído de observação. Assume-se, geralmente, as matrizes \mathbf{A}_l e \mathbf{H}_l como conhecidas.

Existem alguns modelos na literatura que descrevem em detalhes a aquisição destas imagens porém, novamente, como este não é o foco principal deste trabalho, os mesmos não serão apresentados mas podem ser encontrados em diversas referências bibliográficas [25, 26, 27].

2.1.4 Estimação de Imagens de Alta-Resolução

Anteriormente foi mostrada a relação entre as observações de baixa-resolução e a imagem de alta-resolução e seus vetores de movimento com o uso de distribuições condicionais. Isto é, foi modelado $\mathbf{P}(\mathbf{o}|\mathbf{f}_k, \mathbf{d})$ onde \mathbf{o} denota o vetor de observação que pode ser formado tanto pelas observações de baixa-resolução não comprimidas \mathbf{g} , quanto pelas observações de baixa-resolução \mathbf{y} e seus vetores de movimento \mathbf{v} .

O próximo passo na aplicação do paradigma Bayesiano no problema de super-resolução é estudar os modelos *a priori* que foram propostos na literatura para as imagens de alta-resolução, $\mathbf{P}(\mathbf{f}_k)$. Com isso, serão estudados todos os elementos do modelamento Bayesiano para o problema da super-resolução, ou seja, as quantidades do lado direito da Equação 2.14.

$$\mathbf{P}(\mathbf{f}_k, \mathbf{d}, \mathbf{o}) = \mathbf{P}(\mathbf{f}_k) \mathbf{P}(\mathbf{d}|\mathbf{f}_k) \mathbf{P}(\mathbf{o}|\mathbf{f}_k, \mathbf{d}) \quad (2.14)$$

Como mencionado na Seção 2.1.2, os trabalhos na literatura tem lidado com essas questões a partir de duas metodologias que promovem aproximações para a solução mostrada na Equação 2.15 a seguir.

$$\hat{\mathbf{f}}_k, \hat{\mathbf{d}} = \arg \max_{\mathbf{f}_k, \mathbf{d}} \mathbf{P}(\mathbf{f}_k, \mathbf{d}|\mathbf{o}) = \arg \max_{\mathbf{f}_k, \mathbf{d}} \mathbf{P}(\mathbf{f}_k) \mathbf{P}(\mathbf{d}|\mathbf{f}_k) \mathbf{P}(\mathbf{o}|\mathbf{f}_k, \mathbf{d}) \quad (2.15)$$

Tais metodologias serão mais detalhadas e contextualizadas a seguir.

- Primeiro Algoritmo: metodologia de super-resolução alternada.

1. Obtenha $\hat{\mathbf{f}}^0$, uma estimativa inicial da sequência de alta-resolução \mathbf{f} (geralmente aqui é utilizada a sequência de baixa-resolução interpolada);
2. Ajuste $i = 0$;
3. Até o critério de convergência ser encontrado
 - (a) Usando $\hat{\mathbf{f}}^i$, encontre uma estimativa dos vetores de movimento de alta-resolução \mathbf{d} , $\hat{\mathbf{d}}^{i+1}$ (de acordo com a Equação 2.15);
 - (b) Utilizando $\hat{\mathbf{d}}^{i+1}$, ache uma estimativa para a imagem de alta-resolução \mathbf{f}_k , $\hat{\mathbf{f}}_k^{i+1}$ (também de acordo com a Equação 2.15);
 - (c) Atualize $\hat{\mathbf{f}}^{i+1}$, a nova sequência estimada de alta-resolução. Este passo pode consistir somente na substituição de $\hat{\mathbf{f}}_k^i$ por $\hat{\mathbf{f}}_k^{i+1}$ na sequência $\hat{\mathbf{f}}^i$;
 - (d) Ajuste $i = i + 1$;
4. Ajuste $\hat{\mathbf{f}}_k = \hat{\mathbf{f}}_k^i$.

- Segundo Algoritmo: metodologia de super-resolução sequencial.

1. Obtenha $\hat{\mathbf{f}}$, uma estimativa inicial do quadro de alta-resolução \mathbf{f} (geralmente aqui é utilizada a sequência de baixa-resolução interpolada);
2. Utilizando $\hat{\mathbf{f}}$, ache uma estimativa dos vetores de movimento de alta-resolução \mathbf{d} , $\hat{\mathbf{d}}$ por algum dos métodos descritos em [20] (não necessariamente utilizando a Equação 2.15);
3. Usando $\hat{\mathbf{d}}$, encontre a estimativa para a imagem de alta-resolução \mathbf{f}_k , $\hat{\mathbf{f}}_k$ por algum dos métodos descritos em [20] (de acordo com a Equação 2.15).

Como descrito em [20], modelamento e inferência Bayesiana promovem um esquema mais forte e flexível do que aquele necessário para simplesmente encontrar estimativas aproximando a solução da Equação 2.15 através das metodologias supracitadas.

2.2 SUPER-RESOLUÇÃO BAYESIANA APLICADA À COMPRESSÃO DE IMAGENS

Até este momento, foram discutidos meios para aumentar a resolução espacial de imagens e sequências de vídeo com objetivo de obter maior nível de detalhamento de imagens devido às limitações nos sistemas de imageamento. Nesta Seção será discutido métodos de sub-amostragem intencional de uma sequência de vídeo como parte do pré-processamento antes da compressão, e a aplicação das técnicas de super-resolução como um pós-processamento da sequência comprimida. Este sistema pode ser utilizado como um mecanismo de controle de taxa para otimizar a qualidade do vídeo de alta-resolução reconstruído. A principal diferença entre os métodos de super-resolução descritos nas Seções anteriores e a super-resolução para compressão descrita aqui é que neste último caso a sequência original está disponível, portanto a reconstrução pode ser avaliada, criando um interessante conjunto de possibilidades.

Decimação e interpolação interpretam papéis chave no desenvolvimento de *codecs* de vídeo que suportam escalabilidade espacial. Os conceitos e técnicas de super-resolução podem ser aplicados a padrões de compressão de vídeo existentes sem alterar a compatibilidade do padrão. No entanto, ainda mais importante, podem ser aplicados no desenvolvimento de novos esquemas de compressão de imagem.

Quando um codificador híbrido baseado em transformadas e compensação de movimento é usado para comprimir a sequência original $\mathbf{f} = \mathbf{f}_1, \dots, \mathbf{f}_L$, o l -ésimo quadro reconstruído é dado por

$$\hat{\mathbf{f}}_l = \mathbf{T}^{-1}Q \left[\mathbf{T}(\mathbf{f}_l - MC_l(\hat{\mathbf{f}}_l^P, \mathbf{v}_l)) \right] + MC_l(\hat{\mathbf{f}}_l^P, \mathbf{v}_l), \quad l = 1, \dots, L, \quad (2.16)$$

onde $Q[\cdot]$ representa o quantizador, \mathbf{T} e \mathbf{T}^{-1} as transformadas direta e inversa respectivamente, e $MC_l(\hat{\mathbf{f}}_l^P, \mathbf{v}_l)$ indica a estimacão de movimento predita de \mathbf{f}_l formada pela compensação de movimento dos quadros previamente decodificados $\hat{\mathbf{f}}_l^P$ com o uso dos vetores de movimento \mathbf{v}_l .

Para controlar e alcançar uma melhor solução de compromisso entre a taxa de transmissão e a distorção, a sequência de alta-resolução original pode ser pré-processada para obter

$$\mathbf{b}_l = \mathbf{H}_l \mathbf{f}_l, \quad l = 1, \dots, L, \quad (2.17)$$

onde \mathbf{H}_l é uma matriz pré-processada $(PM \times PN) \times (PM \times PN)$ que não inclui a sub-amostragem. A sequência $\mathbf{b} = \{\mathbf{b}_1, \dots, \mathbf{b}_L\}$ é então comprimida usando o esquema de compressão da Equação 2.16 para obter $\mathbf{b}^c = \{\mathbf{b}_1^c, \dots, \mathbf{b}_L^c\}$, a sequência comprimida. Levando em consideração a relação entre quadros da Equação 2.12, isto é, substituindo \mathbf{f}_l por $\mathbf{C}(d_{l,k})\mathbf{f}_k$ tem-se que o l -ésimo quadro, \mathbf{b}_l^c , é dado por

$$\mathbf{b}_l^c = \mathbf{T}^{-1}Q [\mathbf{T}(\mathbf{H}_l \mathbf{C}(d_{l,k})\mathbf{f}_k - MC_l(\mathbf{b}_l^{cP}, \mathbf{v}_l))] + MC_l(\mathbf{b}_l^{cP}, \mathbf{v}_l), \quad (2.18)$$

onde $MC_l(\mathbf{b}_l^{cP}, \mathbf{v}_l)$ é a predição da compensação de movimento de \mathbf{b}_l formado pelo quadro compensado previamente decodificado \mathbf{b}_l^{cP} com os vetores \mathbf{v}_l .

Esta sequência pode agora ser pós-processada por um filtro O_l para obter outra estimativa do l -ésimo quadro original, isto é,

$$\hat{\mathbf{f}}_l = \mathbf{O}_l[\mathbf{b}^c], \quad l = 1, \dots, L. \quad (2.19)$$

Note que os filtros de pré- e pós-processamento podem ser dependentes da imagem. Toda a sequência ou subconjunto de \mathbf{b}^c (não só \mathbf{b}_l^c) pode ser usado para obter $\hat{\mathbf{f}}_l$, e \mathbf{O}_l pode ser um filtro linear ou não-linear em \mathbf{b}^c .

O problema recíproco, projetar um pré-processamento ótimo dado o conhecimento do pós-processamento (\mathbf{O}_l na Equação 2.19), pode também ser manejado. Neste caso, o problema é achar a melhor sequência reconstruída pelo projeto da melhor matriz de pré-processamento \mathbf{H}_l para cada l tal que

$$\hat{\mathbf{H}}_l = \arg \min_{\mathbf{H}_l} \|\hat{\mathbf{f}}_l - \mathbf{f}_l\|^2. \quad (2.20)$$

Claramente, $\hat{\mathbf{f}}_l$ depende de \mathbf{H}_l . Simplesmente modelando o processo de compressão como um processo de ruído Gaussiano para cada quadro se tem

$$\hat{\mathbf{H}}_l = \arg \min_{\mathbf{H}_l} \|\mathbf{O}_l \mathbf{H}_l \mathbf{f}_l - \mathbf{f}_l\|^2, \quad (2.21)$$

o que é novamente um problema mal-colocado. Um modelamento mais realista do processo de compressão que também leva em consideração o termo anteriormente estimado \mathbf{H}_n , $n < l$ é estudado em [28].

Resumidamente, o esquema apresentado indica uma sub-amostragem intencional em adição ao pré-processamento prévio à compressão da sequência de vídeo. Já no lado do decodificador, a sequência é pós-processada utilizando uma técnica de super-resolução de modo a reconstruir a sequência para sua resolução espacial original [29, 30]. Contando que a decimação é proposital e controlada, informações sobre este processo podem ser transmitidas do codificador ao decodificador para alcançar melhor reconstrução da sequência de vídeo, podendo até mesmo, em casos específicos, ser incluído um sistema de malha fechada para que ocorra ainda trocas mais intensas de informações sobre os processamentos envolvidos e a super-resolução realizada em cada quadro.

3 SUPER-RESOLUÇÃO PROPOSTA

UTILIZANDO QUADROS CHAVE

Este Capítulo visa apresentar o que foi realizado no projeto explicitando em cada Subseção os estágios percorridos. O resultado gerado neste projeto foi uma técnica não-Bayesiana de super-resolução para sequências de vídeo desconsiderando treinamentos, distribuições de probabilidade e envio de informações adicionais do codificador ao decodificador. Para o cenário em questão, foi considerada a utilização de quadros chave de alta resolução (tamanho padrão, sem amostragem espacial) que visam ajudar na melhoria da resolução de quadros não-chave de baixa resolução (com amostragem espacial). Este método foi inspirado por *Freeman* [31] e a motivação deste projeto consistiu em prover super-resolução de quadros no lado do decodificador em um esquema de codificação de vídeo de complexidade reversa e escalonável [32, 11, 33, 34]. Apesar da motivação ser o uso da super-resolução nesse esquema específico de codificação de vídeo, a técnica proposta pode ser também utilizada em outros casos onde exista limitação de largura de banda de transmissão não sendo possível enviar todos os quadros codificados em resolução normal.

3.1 VISÃO GERAL DO SISTEMA DE CODIFICAÇÃO E DECODIFICAÇÃO

A técnica de super-resolução proposta levou em consideração um cenário de codificação e decodificação (*codec*) de complexidade reversa e que utilizem sequências de vídeo de resolução mista [11, 33, 34]. Apesar do processo de codificação em si não ser o foco deste projeto, o cenário utilizado como ambientação para a técnica de super-resolução proposta considerou algumas características e restrições impostas por um paradigma de codificação. Tais particularidades e o desenvolvimento do projeto serão especificados nas seções adiante.

3.1.1 Sistemas de Codificação e Decodificação Tradicionais

Usualmente, os *codecs* tradicionais e presentes na maioria dos dispositivos e sistemas atuais possuem um codificador consideravelmente mais complexo que o decodificador.

A maior complexidade computacional presente na parte do codificador em *codecs* tradicionais provém da estimação de movimento [6, 7, 8]. Tal tarefa, além de altamente iterativa, geralmente também é realizada em grande parte dos quadros presentes em uma sequência de vídeo através da codificação inter-quadros. Como explicado no Capítulo 1, a estimação de movimento no codificador encontra - dentro de um intervalo espacial e temporal pré-estabelecido - a melhor combinação entre dois macroblocos podendo enviar ao decodificador somente os vetores de movimento e o resíduo desta estimação, minimizando assim o envio de informações redundantes.

Já no lado do decodificador em um *codec* tradicional, é evidenciada uma complexidade substancialmente mais baixa que no codificador pois nessa etapa não há necessidade de uma nova estimação de movimento. Basicamente é realizada a reconstrução baseada nas informações enviadas pelo codificador.



Figura 3.1: Exemplo de uso para o sistema tradicional de codificação onde a máquina responsável pela codificação possui disponibilidade de processamento e sem restrições de energia

Dessa forma, é natural imaginar que se a complexidade computacional de um *codec* tradicional se faz presente majoritariamente do lado do codificador, as máquinas ou dispositivos encarregados dessa etapa devem consumir consideravelmente mais recursos energéticos e de processamento sendo mais apropriado o uso de máquinas como servidores ou computadores

peçoais (Figura 3.1). O trabalho da decodificação, por outro lado, não sobrecarrega tanto quanto o da codificação podendo ser realizada sem tantos impecilhos por uma máquina com menos atributos.

3.1.2 Sistemas de Codificação e Decodificação de Complexidade Reversa

Com a evolução dos dispositivos móveis e da difusão em massa dos conteúdos multimídia, há uma corrente necessidade de geração e transmissão de vídeos a partir de equipamentos portáteis tais como telefones celulares. Para esse novo caso, um *codec* baseado em um esquema de codificação tradicional pode não ser o mais indicado pois as restrições de processamento, memória e alimentação de energia do dispositivo podem ser de fato impeditivas à realização da codificação e transmissão. Nesse novo cenário (Figura 3.2), o tratamento das etapas de codificação e decodificação deve preferencialmente ser diferenciado. Como até hoje não é conhecido nenhum sistema eficiente que abrigue baixa complexidade no codificador e no decodificador sem que seja comprometida a qualidade da saída, a alternativa é se apoiar no paradigma da codificação de complexidade reversa.

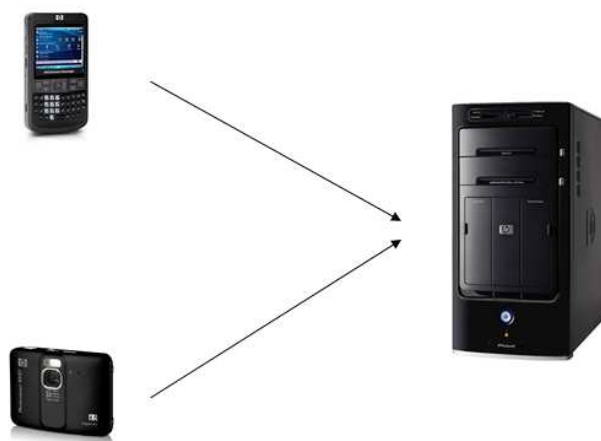


Figura 3.2: Exemplo de uso para o sistema de codificação de complexidade reversa onde a máquina responsável pela codificação possui baixa capacidade de processamento e restrições de energia.

Em um *codec* de complexidade reversa, o peso da maior complexidade do sistema é transferido para o decodificador. Para tanto, é preciso transferir ou eliminar as tarefas de maior processamento do codificador tal como a estimação de movimento. No cenário considerado, essa

etapa da codificação pode ser reduzida ou retirada para diminuir a complexidade computacional. No entanto, o intuito de qualquer sistema de codificação de vídeo é o decréscimo da quantidade de bits a ser transmitida ou armazenada prejudicando o menos possível a qualidade do vídeo. Se a estimação de movimento não for utilizada na codificação então um indesejável efeito é a não exploração e eliminação das redundâncias entre os quadros de uma dada sequência de vídeo gerando conseqüentemente o inevitável aumento da taxa de transmissão.

Contudo, uma alternativa a ser utilizada para o decréscimo da taxa de transmissão é a redução da resolução de quadros de modo que a quantidade total de dados seja menor. A Figura 3.3 a seguir ilustra um exemplo de sequência de vídeo de resolução mista.

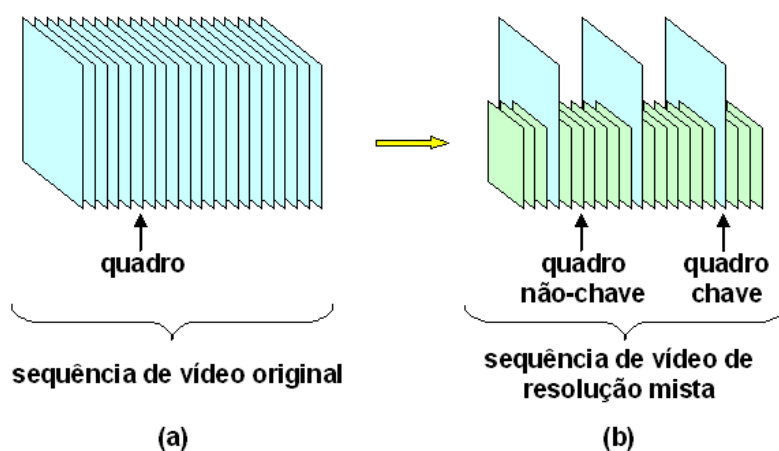


Figura 3.3: Ilustração de uma sequência de vídeo de (a) resolução original e (b) resolução mista. As imagens que mantêm sua resolução original são chamados quadros chave, enquanto as imagens de resolução reduzida são chamados quadros não-chave.

Alguns dos quadros da sequência original (Figura 3.3(a)) são sub-amostrados enquanto o restante dos quadros têm sua resolução preservada. A sequência de vídeo resultante (Figura 3.3(b)) engloba tanto imagens de resolução original quanto reduzida, sendo chamado portanto de sequência de vídeo de resolução mista. Para distinguir nominalmente os quadros, aqueles que permanecem inalterados são chamados *quadros chave* enquanto os de resolução diminuída são chamados *quadros não-chave*.

Neste trabalho especificamente, as sequências originais utilizadas possuem resolução CIF (352x288 *pixels*) e tiveram alguns de seus quadros reduzidos para resolução QCIF (176x144 *pixels*), portanto, uma redução direta de 3/4 do total de *pixels* relativos aos quadros originais.

Claramente não está sendo considerado nesse momento o uso de qualquer tipo de compressão de dados, a intenção é somente ilustrar a redução pura do número de *pixels* de cada imagem.

Nas Subseções 3.1.3 e 3.1.4 a seguir serão apresentadas uma visão geral de como a super-resolução proposta se encaixaria em um dado *codec*. Nenhum sistema de codificação e decodificação é nominalmente citado nessas subseções devido à flexibilidade de utilização do método proposto em muitos dos sistemas atualmente conhecidos.

3.1.3 Super-Resolução Utilizando Quadros Chave Sob a Ótica do Codificador

É importante lembrar que um dos intuitos desse esquema é reduzir a complexidade na etapa de codificação. Para tanto, caso o próprio codificador presente no *codec* não consiga codificar imagens de resoluções variadas numa mesma sequência de vídeo, não é necessária a substituição por um codificador especial. É possível utilizar um *codec* comum que codifique, mesmo que separadamente, os quadros chave e não-chave como mostrado na Figura 3.4.

As únicas diferenças para um codificador tradicional são a decimação de quadros não-chave e a conseqüente diminuição nos processos de estimação de movimento.

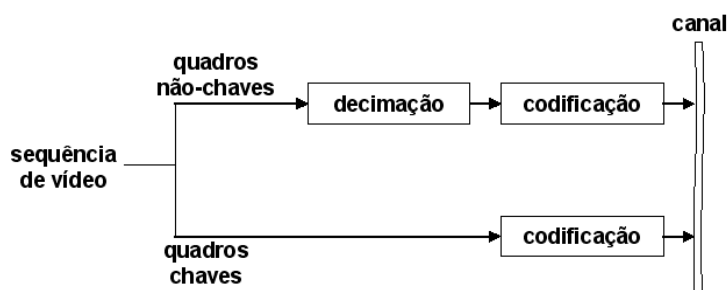


Figura 3.4: Diagrama simplificado do sistema de super-resolução proposto mostrando o lado do codificador de complexidade reduzida para o cenário considerado. Dada uma sequência de vídeo, os quadros não-chave são decimados antes da codificação. Este processo pode ocorrer separadamente para os quadros chave e não-chaves.

A redução de complexidade no codificador, induzida ou não, pode existir de diversas formas. Algumas das maneiras mais conhecidas e significativas são, por exemplo, a ocorrência de blocos *skip*, vetor de movimento nulo (em inglês, *zero motion vector*), o uso de técnicas sub-ótimas de estimação de movimento ou mesmo a completa retirada do sistema de estimação de movimento

inter-quadros (predição *intra*). Apesar de haver diminuição de complexidade quando ocorrem blocos *skip* (onde a informação de movimento do bloco é inferida, assim como seu resíduo) ou vetor de movimento nulo (onde não há deslocamento espacial entre o macrobloco em questão e o de referência, separados temporalmente), tal redução não pode ser forçada pois depende fortemente das características dos quadros da sequência de vídeo. Por outro lado, a escolha por um método sub-ótimo de buscas ou opção pela retirada da estimação de movimento é deixada a cargo do usuário do sistema.

Métodos de busca sub-ótimos são geralmente eficientes porém ainda possuem alto grau de atividade computacional. Já no caso de todos os quadros serem codificados no modo *intra*, eliminando a estimação de movimento inter-quadros, a redução da complexidade no codificador é drástica dada a diminuição do imenso gasto de processamento envolvido na extensa e sistemática procura por uma melhor combinação de macroblocos. Porém, sem a estimação de movimento, a taxa de transmissão normalmente aumentaria pois a grande redundância entre os quadros não é explorada.

Todavia, considerando este novo cenário para o qual a super-resolução proposta é considerada, existe o contraponto e a vantagem de que alguns ou até mesmo a maioria dos quadros presentes na sequência podem ter sua resolução diminuída, provocando grande redução na quantidade de dados a serem estimados, codificados e transmitidos. Independente das diversas oportunidades supracitadas de redução de complexidade (*skip*, vetor de movimento nulo, métodos de busca sub-ótimos etc), a diminuição da resolução (quantidade de *pixels*) induz invariavelmente a um menor esforço computacional.

Um ponto interessante a se considerar é o fato de que para o método de super-resolução proposto, não há nenhuma nova tarefa exaustiva incluída no codificador. O objetivo ideal e final do sistema seria reduzir a complexidade no lado do codificador e, se possível, também reduzir a taxa de transmissão sem comprometer a qualidade da imagem final. A questão da complexidade computacional foi drasticamente afetada pela retirada/diminuição da estimação de movimento. Já a taxa e a qualidade da saída dependerão intimamente do processamento envolvido no lado do decodificador e serão melhor discutidos na próxima Subseção.

3.1.4 Super-Resolução Utilizando Quadros Chave Sob a Ótica do Decodificador

No lado do decodificador, é recebido o sinal composto pela sequência de vídeo de resolução mista. Cada tipo de quadro (chave e não-chave) será processado de forma diferente portanto seguem caminhos distintos nessa etapa. Observe na Figura 3.5 uma visão geral do lado do decodificador no esquema considerado.

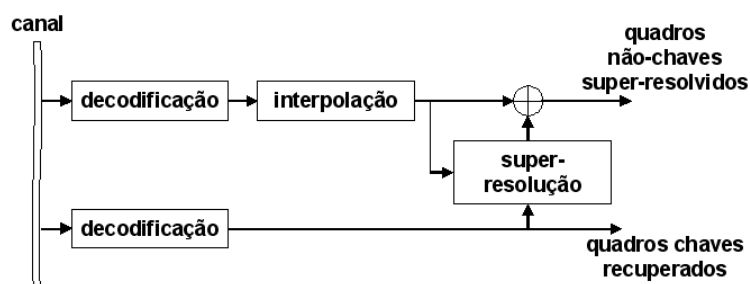


Figura 3.5: Diagrama simplificado do sistema de super-resolução proposto mostrando o lado do decodificador em um *codec* de complexidade reversa considerando um cenário de sequência de vídeo de resolução mista e incluindo o bloco da técnica de super-resolução proposta.

Primeiramente, os quadros chave e não-chaves são decodificados normalmente. No caso dos quadros chave, não há necessidade de nenhum outro tipo de tratamento após a decodificação. Esses quadros não possuem, *a priori*, qualquer diferença de qualidade em relação a codificação em um sistema tradicional.

Já os quadros não-chave, após a decodificação, são interpolados na proporção inversa da decimação no codificador para que voltem à sua dimensão original. Contudo, a qualidade dos quadros não-chave interpolados é visivelmente inferior comparativamente aos mesmos quadros em sua resolução original (que não estão presentes no decodificador) pois estão claramente borrados como mostrado na Figura 3.6(c).

Como explicado no Capítulo 1, a banda de alta-frequência desses quadros interpolados é perdida durante as etapas de redimensionamento. Já que a baixa-frequência dos quadros é a mais fiel possível em relação aos originais, a questão que permanece então é como estimar e recuperar a alta-frequência perdida nos quadros não-chave. O método de super-resolução proposto tenta diminuir os efeitos desse problema.

Perceba pelas Figuras 3.4 e 3.5 que a técnica de super-resolução proposta se concentra

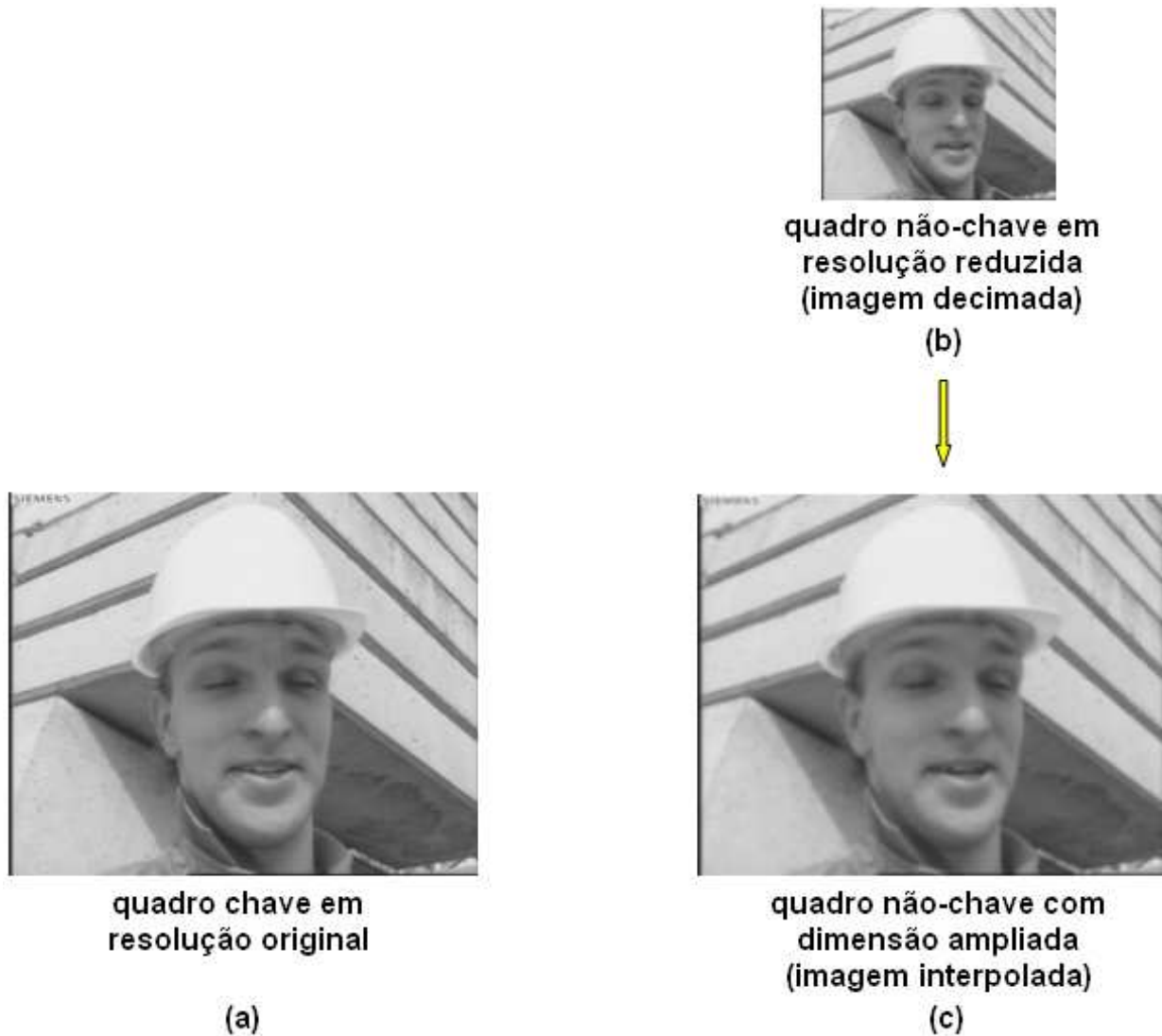


Figura 3.6: Exemplo ilustrativo dos quadros no decodificador: (a) quadro chave em resolução original (presente no decodificador), (b) quadro não-chave em resolução reduzida (decimado no codificador e enviado ao decodificador) e (c) quadro não-chave com dimensão ampliada (interpolado no decodificador).

inteiramente no lado do decodificador. Os detalhes do bloco de super-resolução e todo o processamento envolvido nessa etapa serão detalhados na Seção 3.2 adiante.

3.2 MÉTODO DE SUPER-RESOLUÇÃO PROPOSTO: ABORDAGEM BÁSICA

No lado do codificador, abreviado como *encod* nos diagramas, a sequência de vídeo ($\{X[n]$, para qualquer n }, abreviado nesse documento como $\{X[n]\}$ indicando toda uma sequência de quadros) é separada em quadros não-chave $\{G'[n]\}$ (em inglês, *non-key frames*) sub-amostrados por um fator M antes da codificação e quadros chave $\{G[n]\}$ (em inglês, *key frames*) que serão codificados em sua resolução original como exibido no diagrama da Figura 3.7. O operador $e(\cdot)$ no diagrama simboliza o quadro codificado com perdas (*lossy*, em inglês), ou seja, o quadro reconstruído na decodificação não é uma réplica do quadro antes da codificação. Vale lembrar que os processos de redimensionamento citados neste manuscrito (decimação e interpolação) sempre englobam a filtragem exposta na Seção 1.2.3 porém, por simplicidade, são sempre mencionados como operações únicas no texto e diagramas.

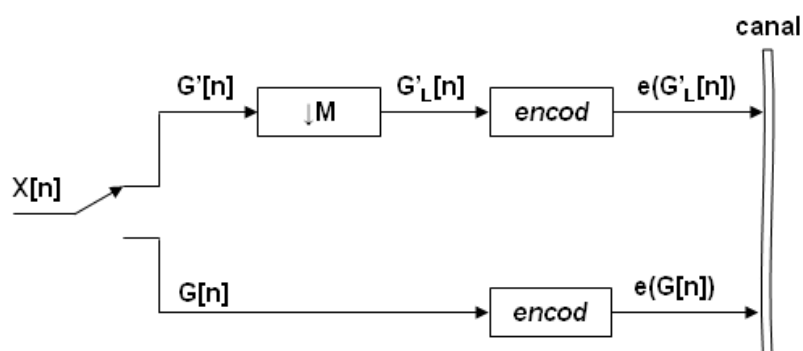


Figura 3.7: Diagrama simplificado no lado do codificador mostrando o cenário considerado com uma sequência de resolução mista formada por quadros não-chave, indicados como $G'[n]$, e por quadros chave, $G[n]$.

No lado do decodificador, abreviado nos diagramas como *decod*, esses quadros são recebidos e decodificados normalmente, cada um com sua devida resolução. Assim como no lado do codificador, quadros chave e não-chave são processados separadamente. Os detalhes da Figura

3.8 serão explicados nas seções seguintes.

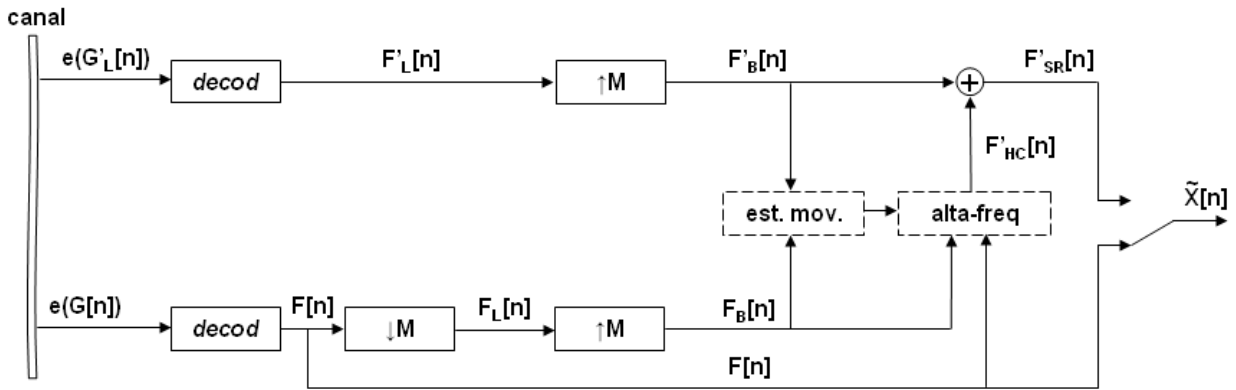


Figura 3.8: Diagrama simplificado no lado do decodificador mostrando o cenário considerado com uma sequência de resolução mista, $F'[n]$ para quadros não-chave e $F[n]$ para quadros chave. Os blocos tracejados indicam as operações realizadas em nível de macrobloco.

Os quadros não-chave são interpolados pelo mesmo fator M utilizado na decimação no lado do codificador gerando $\{F'_L[n]\}$ visto na Figura 3.8. Os processos de redimensionamento aplicados aos quadros não-chave provocam perdas em componentes de alta-frequência. Uma possível solução para minimizar esse efeito indesejável é o aproveitamento da alta-frequência presente nos quadro chave para recuperar parte da alta-frequência perdida pelos quadros não-chave.

3.2.1 Explorando a Correlação Entre os Quadros

O aproveitamento da informação de quadros vizinhos se deve ao fato de que normalmente quadros vizinhos são altamente correlacionados. Geralmente as sequências de vídeo são capturadas a uma taxa de 30 quadros por segundo (FPS, do inglês *frames per second*) e a maioria dessas sequências normalmente não possui um quantidade exagerada de movimento inter-quadros e variação de cenários e objetos. O coeficiente de correlação c em um sinal bidimensional (2D) é calculado da seguinte forma

$$c = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\sum_m \sum_n (A_{mn} - \bar{A})^2 (B_{mn} - \bar{B})^2}}, \quad (3.1)$$

onde,

$$\bar{Y} = \frac{\sum_m \sum_n (Y_{mn})}{MN} \quad (3.2)$$

é o valor médio de um dado sinal bidimensional Y de dimensões $M \times N$.

As Tabelas 3.1 e 3.2 a seguir mostram exemplos do coeficiente de correlação calculado entre o primeiro e todos os outros quadros subsequentes (até o décimo quinto quadro) para seis sequências diferentes de vídeo.

É possível observar que com exceção das sequências *Mobile* e *Football* - que apresentam muito movimento inter-quadros e novos objetos a cada cena - todas as outras sequências apresentadas possuem índice de correlação bastante elevado até mesmo para quadros mais distantes.

3.2.2 Recuperação da Alta-Frequência

Para melhor decidir sobre a alta-frequência a ser adicionada em um quadro não-chave, é preciso procurar em um quadro chave a alta-frequência que mais se assemelharia àquela que lá estava originalmente. Porém, no lado do decodificador as únicas fontes de informação para auxiliar nessa decisão seriam os quadros chave e os não-chave interpolados (consideravelmente borrados) como indicado na Figura 3.9.

Com o intuito de realizar a estimação de movimento entre quadros chave e não-chave de forma mais adequada é necessário processar os quadros chave antes da procura. Dado que os quadros não-chave perderam sua componente de alta-frequência devido aos processos de decimação e interpolação (Figura 3.11), é preciso degradar os quadros chave da mesma maneira (ver Figuras 3.9 e 3.10).

Como mostrado no diagrama da Figura 3.8 e em destaque na Figura 3.12, os quadros chave passam pelos mesmos processos de decimação e interpolação sofridos pelos quadros não-chave, considerando, obviamente, o mesmo fator de redimensionamento M . Veja na Figura 3.10 um exemplo desse redimensionamento sofrido por um quadro chave no decodificador onde o quadro resultante (Figura 3.10(c)) tem suas bordas nitidamente afetadas com a perda da alta-frequência.

Após serem decodificados, os quadros chave seguem por dois caminhos distintos (como visto

Seq./Quadro	Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7
Foreman	1,000	0,981	0,952	0,934	0,923	0,915	0,907
Coastguard	1,000	0,972	0,945	0,930	0,918	0,909	0,901
Mobile	1,000	0,917	0,794	0,712	0,674	0,658	0,643
Mother	1,000	0,997	0,994	0,994	0,996	0,998	0,996
Akyio	1,000	1,000	1,000	1,000	0,999	0,998	0,998
Football	1,000	0,582	0,432	0,370	0,299	0,266	0,205

Tabela 3.1: Tabela de índices de correlação entre o primeiro e os outros quadros (até o sétimo quadro) dentro de cada uma das sequências de vídeo. Correlações consideradas sem qualquer estimativa de movimento.

Seq./Quadro	Q. 8	Q. 9	Q. 10	Q. 11	Q. 12	Q. 13	Q. 14	Q. 15
Foreman	0,895	0,877	0,857	0,837	0,818	0,804	0,795	0,793
Coastguard	0,896	0,890	0,884	0,878	0,874	0,870	0,864	0,859
Mobile	0,625	0,602	0,580	0,560	0,546	0,531	0,517	0,503
Mother	0,995	0,995	0,997	0,997	0,997	0,995	0,994	0,994
Akyio	0,998	0,998	0,998	0,998	0,997	0,997	0,996	0,996
Football	0,170	0,122	0,095	0,111	0,129	0,132	0,144	0,164

Tabela 3.2: Tabela de índices de correlação entre o primeiro e os outros quadros (do oitavo ao décimo quinto quadro) dentro de cada uma das sequências de vídeo. Correlações consideradas sem qualquer estimativa de movimento.

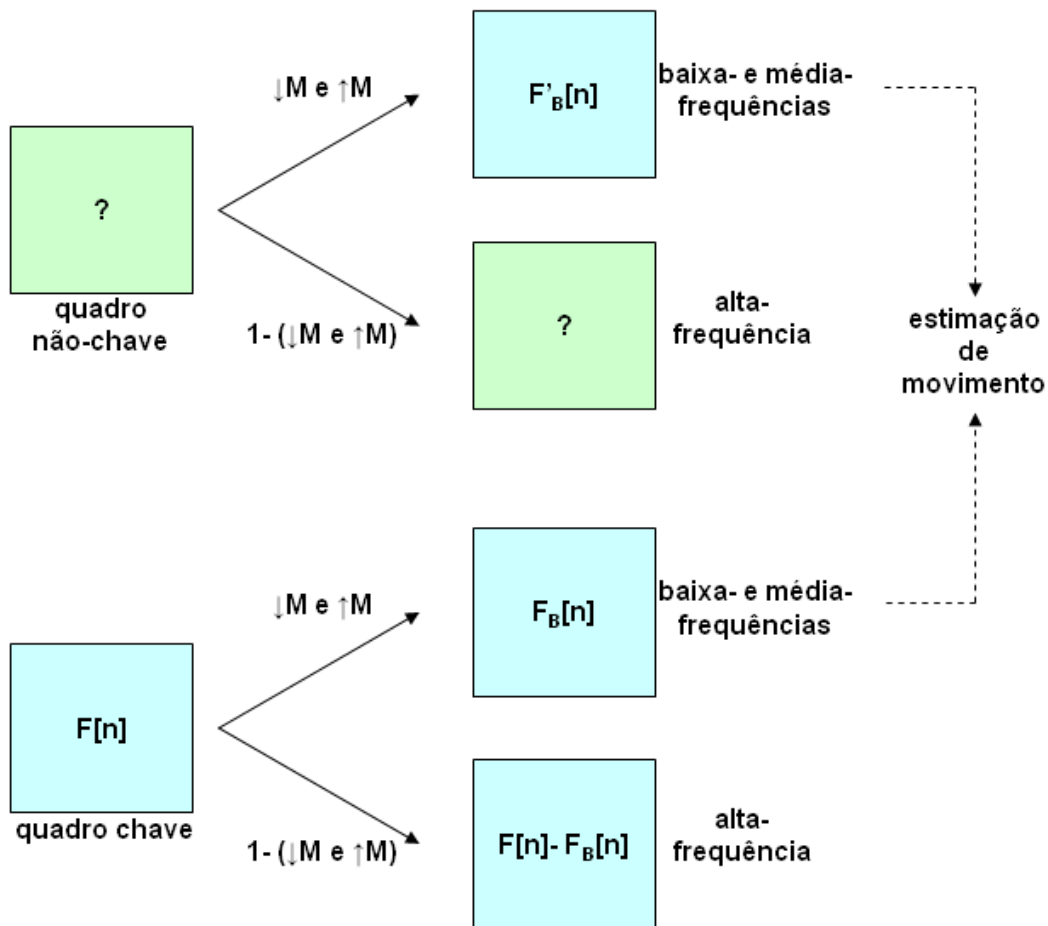


Figura 3.9: Informação disponível no lado do decodificador: todas as componentes dos quadros chave e somente as componentes de baixa e média frequência dos quadros não-chave. A estimação de movimento é realizada nas componentes de frequências mais baixas.

na Figura 3.12): o primeiro deles (Figura 3.12(b)) é a simulação da perda da componente de alta-frequência para melhor desempenho e confiabilidade da estimação de movimento, e o segundo caminho (Figura 3.12(c)) engloba os quadros chave originalmente decodificados que terão sua alta-frequência utilizada adiante para recuperar tal componente nos quadros não-chave. É válido ressaltar que qualquer processamento pós-decodificação aqui mencionado sofrido pelos quadros chave não é de modo algum definitivo, dado que tais quadros já possuem todas as componentes e não precisam ser super-resolvidos.



Figura 3.10: Exemplo de quadro chave redimensionado para simular a perda das altas-frequências onde se vê (a) quadro original, (b) quadro decimado e (c) quadro interpolado nitidamente borrado.

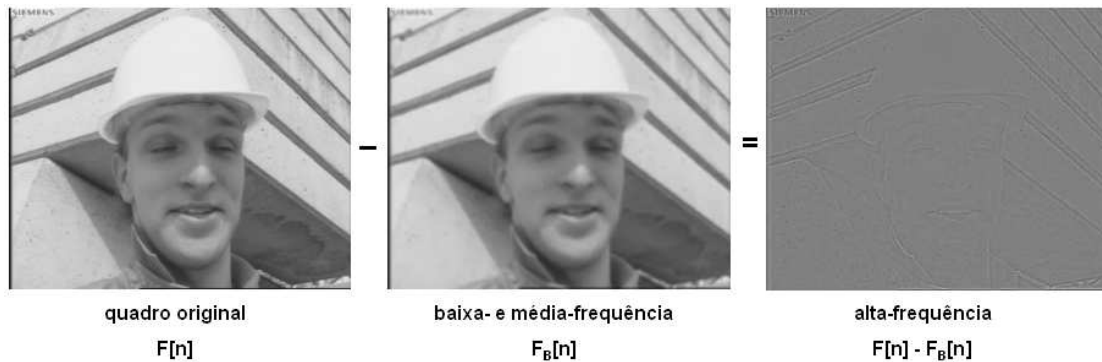


Figura 3.11: Exemplo de alta, média e baixa frequência de um dado quadro. A componente de alta-frequência é extraída do quadro original subtraído do quadro interpolado.

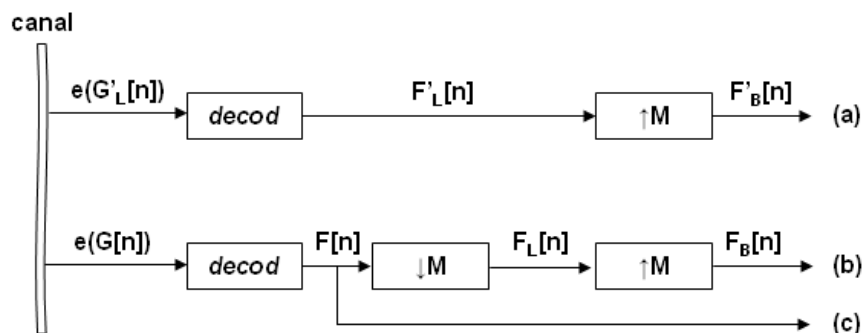


Figura 3.12: Etapas iniciais no lado do decodificador: (a) quadros não-chave são interpolados e retornam a sua resolução original porém com qualidade inferior, (b) quadros chave sofrem uma degradação simulada sendo sucessivamente redimensionados (decimados e interpolados) pelo mesmo fator M utilizado nos quadros não-chave e (c) quadros chave seguem sem modificações à próxima etapa do processo de super-resolução.

3.2.3 Realizando a Estimação de Movimento

O próximo passo é realizar a estimação de movimento no decodificador entre os quadros não-chave ($\{F'_B[n]\}$) e chave borrados ($\{F_B[n]\}$) como indicado na Figura 3.9. Neste projeto, o método de busca utilizado para determinar a melhor combinação entre macroblocos foi a busca completa dentro de uma janela previamente limitada de 32×32 *pixels*. Os tamanhos de macrobloco utilizados na procura foram diversos, variando entre 4×4 , 8×8 e 16×16 *pixels*. As métricas de comparação utilizadas foram Soma das Diferenças Absolutas (SAD, do inglês *Sum of Absolute Differences*) ou ainda Soma das Diferenças Quadradas (SSD, do inglês *Sum of Squared Differences*). Quanto menor o valor da SAD/SSD, melhor é o casamento entre dois sinais.

Tanto a SAD quanto a SSD são métricas conhecidas e muito utilizadas atualmente [12]. Ambas são baseadas na norma L_p que é calculada da forma

$$\|x\|_p = \left(\sum_i \sum_j |x_{ij} - x'_{ij}|^p \right)^{1/p}. \quad (3.3)$$

A SAD é calculada com $p = 1$ enquanto a SSD utiliza $p = 2$.

A estimação de movimento é feita bidirecionalmente, ou seja, entre um macrobloco do quadro não-chave e os macroblocos pertencentes à janela de busca de ambos quadros chave mais próximos (anterior e posterior). Como o critério utilizado de quais quadros serão chave ou não-chaves foi a distribuição temporalmente uniforme entre eles dentro de uma sequência, a distância média de um quadro não-chave para um quadro chave é sempre a metade do tamanho do intervalo h entre dois quadros chave como mostra a Figura 3.13. Tal proximidade tende a proporcionar coeficientes de correlação mais altos e melhores combinações entre macroblocos.

Após o término da busca, o macrobloco do quadro chave borrado que obteve menor SAD (ou SSD) em relação ao macrobloco do quadro não-chave determina o lugar de onde será retirada a componente de alta-frequência para super-resolver aquele macrobloco em questão do quadro não-chave.

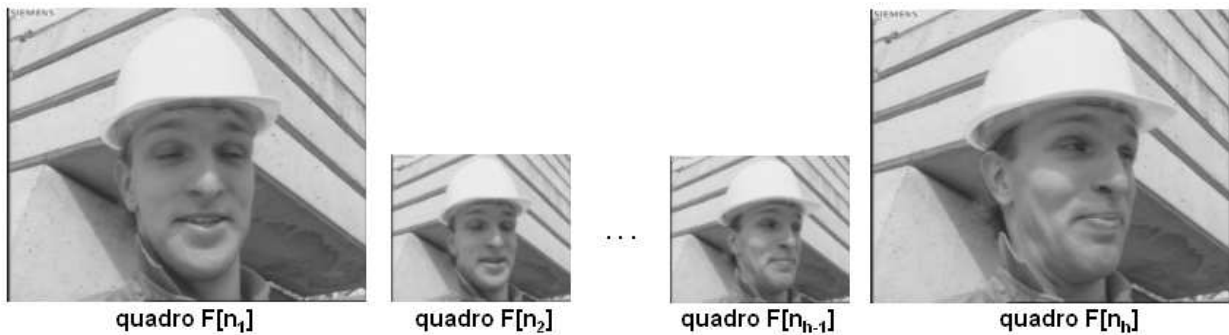


Figura 3.13: Exemplo de distribuição de quadros chave e não-chave dentro de uma dada sequência. A frequência de quadros chave é definida por um fator h pré-estabelecido no codificador. A predição de movimento é feita bidirecionalmente sendo realizada uma procura entre um macrobloco do quadro não-chave e os dois quadros chave mais próximos.

3.2.4 Super-Resolução dos Quadros Não-Chave

Definidas as posições dos macroblocos dos quadros chave de onde serão utilizadas as altas-frequências, a próxima etapa é super-resolver os macroblocos dos quadros não-chave. A Figura 3.14 mostra em destaque o diagrama das etapas de estimação de movimento e super-resolução.

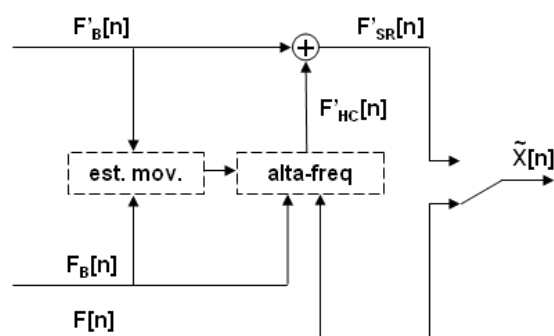


Figura 3.14: Detalhe do diagrama mostrando as etapas de estimação de movimento e posterior super-resolução dos quadros não-chaves (adição de alta-frequência). A saída do sistema ($\{\tilde{X}[n]\}$) retrata os quadros chaves ($\{F[n]\}$) e os quadros não-chaves super-resolvidos ($\{F'_{SR}[n]\}$). Vale lembrar que os blocos tracejados indicam as operações realizadas em nível de macrobloco.

Veja pela Figura 3.15 que um quadro com altas-frequências será criado e receberá a alta-frequência dos quadros chave de modo que ao final deste processo seja estimado todo um quadro de alta-frequência (Figura 3.15(a)). Então, o quadro não-chave super-resolvido é finalmente alcançado ao se adicionar o já existente quadro não-chave borrado $F'_B[n]$ ao recém-criado quadro

de altas-frequências estimadas (Figura 3.15(b)).

O quadro de altas-frequências estimadas ($F'_{HC}[n]$) utilizado para super-resolver seu respectivo quadro não-chave ($F'_B[n]$) é formado pelo conjunto de macroblocos de alta-frequência encontrado a partir do processo de estimação de movimento. Contudo, este quadro compensado possui somente as componentes de alta-frequência resultantes da subtração entre o macrobloco estimado retirado do quadro original $F[n]$ e sua respectiva versão observada a partir do quadro borrado $F_B[n]$ como visto na Figura 3.15. Ou seja, os blocos estimados são retirados do quadro obtido pela seguinte relação:

$$F_A[n] = F[n] - F_B[n]. \quad (3.4)$$

Observe que o termo $F_A[n]$ é formado somente por componentes de alta-frequência pois se trata do quadro original (que abriga todas as componentes) subtraído do quadro borrado que possui majoritariamente baixas e médias frequências. Como visto na Equação 3.4 e na Figura 3.15, as altas frequências ($F_A[n]$) e as baixas e médias frequências ($F_B[n]$) são totalmente complementares.

Em outras palavras, a estimação de movimento define os melhores casamentos de macroblocos a partir dos quadros borrados $F'_B[n]$ e $F_B[n]$ e, após a definição dos vetores de movimento, o quadro compensado de alta-frequência $F'_{HC}[n]$ é formado pelos blocos retirados do quadro $F_A[n]$ considerando os mesmos vetores de movimento encontrados na estimação. Resumidamente, o quadro é estimado através das baixas-frequências porém compensado com as altas-frequências.

Todo o processo explicado até agora retrata de maneira simplificada a idéia básica proposta para super-resolver quadros não-chave a partir de quadros chave em um cenário onde existem sequências de vídeo de resolução mista. Contudo, alguns refinamentos do projeto ainda serão expostos na próxima seção detalhando melhorias apoiadas na idéia básica da super-resolução proposta.

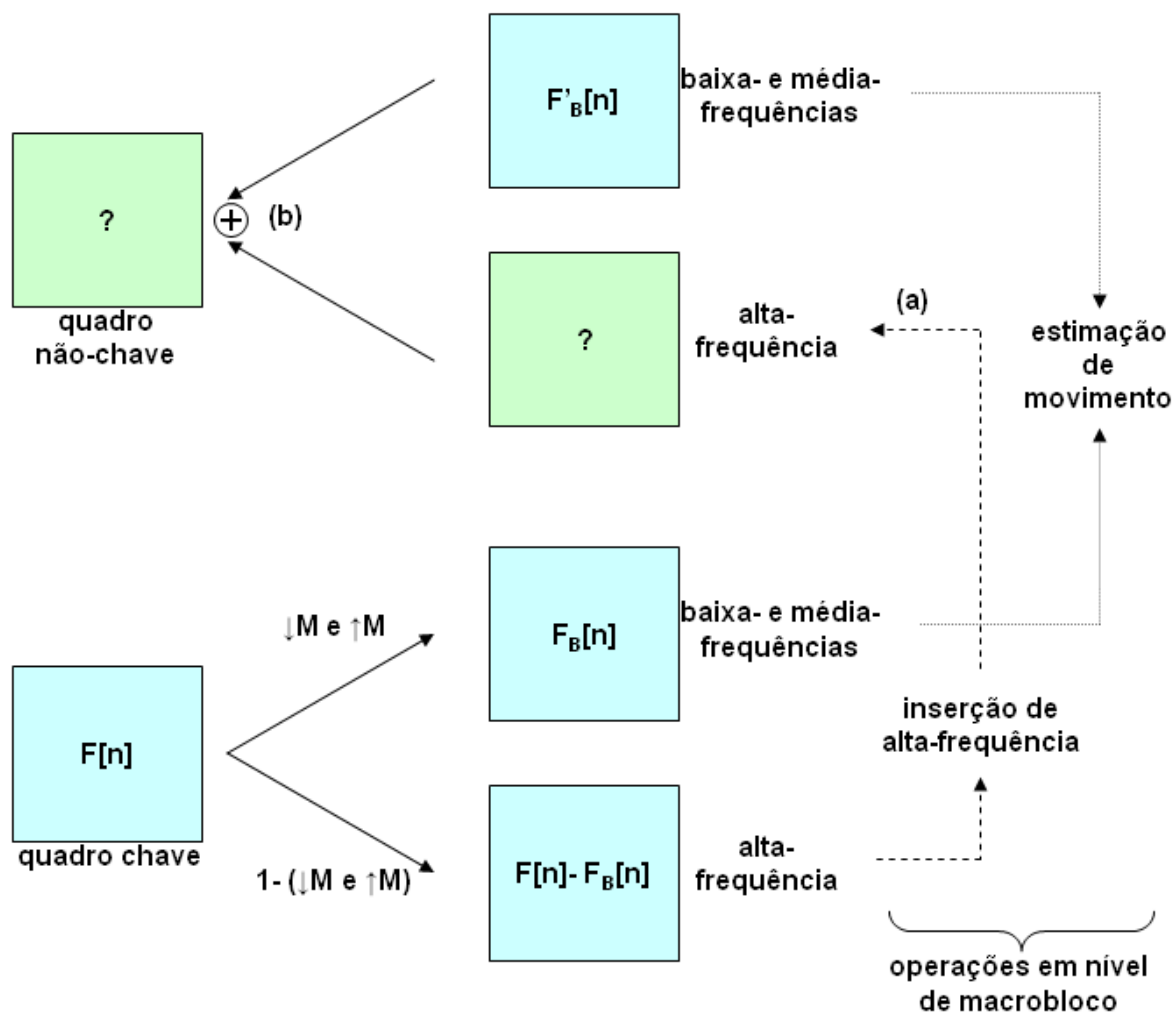


Figura 3.15: Após a estimação de movimento, a próxima etapa é incluir nos quadros não-chave a alta-frequência dos quadros chave. Em (a) é possível ver a alta-frequência dos quadros chave sendo inserida no quadro não-chave e em (b) o estabelecimento do quadro não-chave super-resolvido recompondo as componentes de alta, média e baixa frequências novamente em um único quadro. É importante observar que o processo (a) ocorre em nível de macrobloco, portanto, o quadro de altas-frequências estimadas possui somente os macroblocos de alta-frequência "vencedores" na estimação de movimento, ou seja, que tinham menor SAD/SSD.

3.3 MÉTODO DE SUPER-RESOLUÇÃO PROPOSTO: REFINAMENTOS

No lado do codificador não há refinamento a ser explorado além da necessidade de que os métodos de decimação sejam os menos degradantes possível. Especificamente neste projeto foram testadas duas técnicas já conhecidas de filtragem: o método bilinear e o bicúbico. Os princípios de tais filtrações podem ser conferidos em mais detalhes em [35].

Já no lado do decodificador, vários pontos serão abordados nas próximas subseções mostrando melhoramentos em relação à idéia original exposta na Seção 3.2. Os resultados dos testes validando o uso desses refinamentos serão mostrados no Capítulo 4, adiante.

3.3.1 Estimação de Movimento na Alta-Frequência

Testes realizados nesse projeto levaram a crer que dependendo em quais versões dos quadros chave e não-chave a estimação de movimento fosse realizada, o resultado final poderia ser alterado. A idéia original mostrada na Figura 3.8 era proceder a estimação de movimento entre os quadros borrados $F_B[n]$ e $F'_B[n]$. Porém, uma alternativa a essa primeira idéia foi testada fazendo as buscas entre os quadros chave e não-chave em versões ligeiramente diferentes.

Como o intuito maior deste projeto é recuperar a alta-frequência dos quadros não-chave, naturalmente foi levantada a questão de realizar a estimação de movimento entre as altas-frequências disponíveis nos quadros $F_B[n]$ e $F'_B[n]$. Tais quadros, mesmo perdendo suas componentes de frequência mais alta durante os processos de filtragem/redimensionamento, ainda possuem vestígios mais suaves das mesmas. A Figura 3.16 ilustra tal fenômeno.

Para proceder com a estimação de movimento entre componentes de alta-frequência foi necessário aguçar os detalhes das bordas dos quadros borrados. Para tal tarefa foi utilizado um filtro espacial simples. Este filtro passa-altas possui uma máscara chamada $M_{alta}[n]$ de tamanho 3×3 pixels com os seguintes valores:

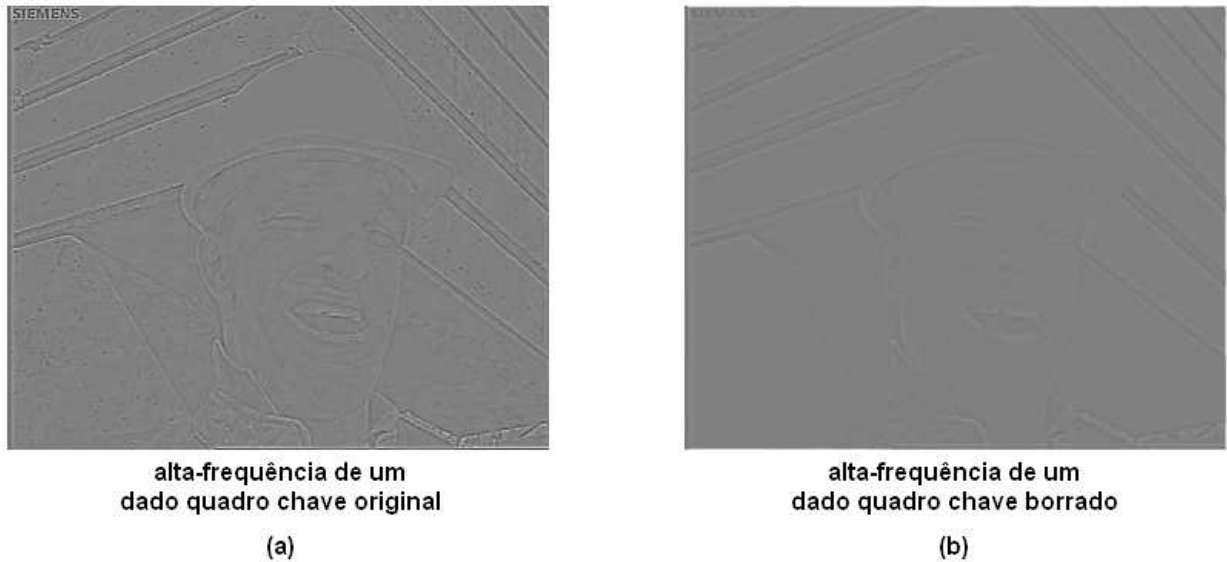


Figura 3.16: Exemplo da alta-frequência de um quadro chave (a) original e (b) borrado. É possível ver que mesmo as altas-frequências do quadro borrado serem consideravelmente mais suaves, elas ainda se apresentam visíveis.

$$M_{alta}[n] = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \quad (3.5)$$

A convolução entre um quadro borrado $F_B[n]$ (seja ele chave ou não-chave) e a máscara do filtro passa-altas $M_{alta}[n]$ resulta em uma nova versão deste quadro, aqui designado como $F_{HF}[n]$. A Equação a seguir mostra essa relação:

$$F_{HF}[n] = F_B[n] * M_{alta}[n]. \quad (3.6)$$

Portanto, após a devida filtragem dos quadros chave e não-chave, resulta-se em suas respectivas versões $F_{HF}[n]$ e $F'_{HF}[n]$. O diagrama ilustrando esta nova configuração é mostrado na Figura 3.17 a seguir.

Para que não sejam artificialmente introduzidas diferenças entre os quadros devida a essa etapa de filtragem, vale lembrar que o filtro passa-altas utilizado deve ser o mesmo tanto para quadros chave como para quadros não-chave.

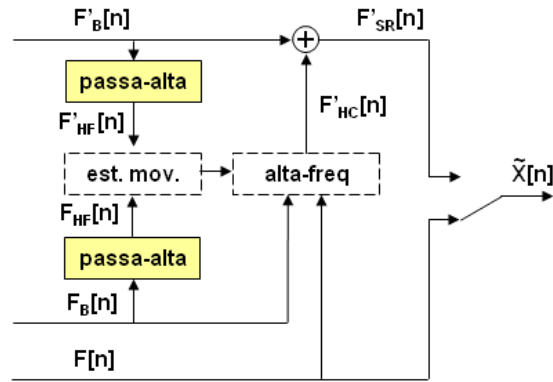


Figura 3.17: Detalhe do diagrama destacando o novo estágio de filtragem antes da estimação de movimento. Os quadros chave e não-chave borrados são processados por um filtro passa-altas resultando, respectivamente, nos quadros $F_{HF}[n]$ e $F'_{HF}[n]$.

O restante da estimação de movimento ocorre conforme já descrito anteriormente na Seção 3.2.3. A única distinção é o uso dos quadros $\{F_{HF}[n]\}$ e $\{F'_{HF}[n]\}$ ao invés dos quadros $\{F_B[n]\}$ e $\{F'_B[n]\}$. Um exemplo do efeito dessa troca de versões de quadros na estimação de movimento pode ser visto na Tabela 3.3.

Estimação/Sequência	Foreman	Coastguard	Mobile	Mother	Akyio	Football
não usando filtro passa-alta	34,5359	31,8552	23,2523	39,9319	40,7134	31,3319
usando filtro passa-alta	34,7690	32,0000	23,0999	40,1578	40,9791	31,7063

Tabela 3.3: Tabela exemplificando o ganho de qualidade objetiva (PSNR) nos quadros super-resolvidos em diversas sequências quando é feita a estimação de movimento na alta-frequência dos quadros chave e não-chave ao invés de realizar a estimação na baixa-frequência.

A Tabela 3.3 ilustra que para a maioria das sequências testadas, fazer a estimação de movimento na alta-frequência do quadro borrado produz um ganho de qualidade na imagem final em relação à estimação nos quadros borrados.

3.3.2 Seleção de Macroblocos de Interesse

Até o presente momento é suposto que todos os macroblocos do quadro não-chave passam pela estimação de movimento à procura de um melhor macrobloco do quadro chave que possa

super-resolvê-lo. Porém, através de testes foi observado que para certos macroblocos dos quadros não-chave é de maior interesse que os mesmos não sejam mais processados na tentativa de restauração da alta-frequência. Este fenômeno ocorre quando tais macroblocos têm componentes de alta-frequência em quantidades extremamente diminutas.

Como já dito anteriormente, após os processos de decimação e interpolação, as altas-frequências existentes originalmente num dado quadro são, de fato, fortemente suavizadas ou até mesmo eliminadas. Quando os quadros passam pela filtragem descrita na Subseção 3.3.1, ainda é possível constatar a presença de componentes de alta-frequência após o borramento porém, consideravelmente menos marcantes que na versão original do quadro como ilustrado na Figura 3.18.

A idéia aqui apresentada se apóia no fato de que bordas fortes (altas-frequências mais marcantes) se propagam através de escalas. O mesmo não ocorre com as texturas e ruídos que acabam sendo amenizadas e, conseqüentemente, perdidas nos processos de redimensionamento. Desse modo, a adição de alta-frequência do método de super-resolução proposto realiza um reforço de bordas (*edge enhancement*, em inglês) tentando minimizar a adição de elementos indesejáveis (ruídos e texturas).

Considerando os quadros não-chave $\{F'_{HF}[n]\}$ em especial, é observado que existem macroblocos que não possuem quantidade significativa de alta-frequência indicando que ali, originalmente, não havia uma componente expressiva desta natureza. Como dito, componentes de alta-frequência mais fortes, como as bordas, ainda podem ser observadas no quadro borrado porém, em contrapartida, em regiões do quadro original onde a alta-frequência se manifestava discreta (ruídos ou texturas), não é possível mais destacar nenhuma forte evidência de que ali havia esta componente e ainda, mesmo que se possa acusar ligeiramente sua presença, não há informações suficientes para precisar de fato suas reais características.

Os ruídos e texturas presentes em um quadro geralmente não são percebidas nas baixas-frequências do mesmo. Dessa forma, ainda que seja estabelecida uma boa combinação entre macroblocos borrados dos quadros chave e não-chave na estimação de movimento, não há garantias que a alta-frequência também pudesse fornecer uma combinação tão boa quanto na baixa-frequência. Intuitivamente, essa mesma problemática seria observada em qualquer

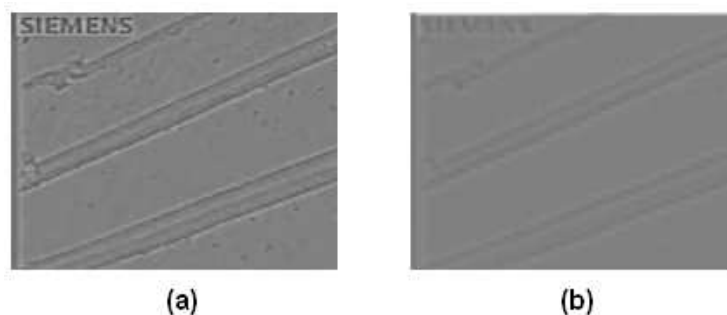


Figura 3.18: Detalhe ampliado da alta-frequência presente em um quadro chave. É possível ver que somente as altas-frequências mais expressivas do quadro original (a) permanecem perceptíveis no quadro borrado (b). Observe que é possível ver as bordas, texturas e ruídos em (a) porém as texturas e ruídos não são percebidos em (b).

outra circunstância e não somente naquela onde os macroblocos são enquadrados como regiões possuindo pequena quantidade de alta-frequência. Contudo, como pode ser visto na Figura 3.18, quanto mais alta a frequência observada no quadro original, mais evidente a mesma fica no quadro borrado indicando de fato sua presença e fornecendo informações suficientes para precisar melhor uma combinação durante a estimação de movimento. Perceba ainda pelo exemplo da Figura 3.18 que pequenos detalhes e texturas vistas na imagem de alta-frequência original não são mais observadas na alta-frequência do quadro borrado. A informação da componente de frequência disponível no quadro borrado pode ser considerada insuficiente para a decisão de super-resolver certas regiões de modo que o melhor a se fazer é manter o macrobloco inalterado.

Em outras palavras, o refinamento descrito nessa subseção tenta minimizar a inserção de alta-frequência em certos macroblocos que não apresentam indícios suficientes de que ali contivesse originalmente alguma componente dessa natureza. Esse refinamento mostrou que, por vezes, é menos nocivo ao macrobloco mantê-lo borrado do que adicionar alguma alta-frequência erroneamente pois o resultado pode ser a adição de ruído e/ou texturas inexistentes originalmente. Curiosamente, essa adição indevida de alta-frequência pode até fazer com que a qualidade subjetiva do quadro ser maior pois introduz elementos que deixam a imagem com detalhes mais aguçados (geralmente mais agradável ao expectador), porém quase que inevitavelmente a qualidade objetiva da imagem tende a cair.

Neste refinamento aqui proposto, foi incluído um limiar empírico estipulando que se a

quantidade de alta-frequência de um dado macrobloco pertencente a $F'_{HF}[n]$ não ultrapassar uma quantidade mínima T_f pré-estabelecida, tal bloco será mantido inalterado. Os macroblocos pertencentes a $F'_{HF}[n]$ serão designados neste texto como $B'_{HF}[i, j]$, onde B'_{HF} indica um macrobloco borrado de um quadro não-chave, e i e j retratam as posições vertical e horizontal do macrobloco dentro do quadro.

Após a análise desta condição, somente os macroblocos $B'_{HF}[i, j]$ que contenham quantidades de alta-frequência superiores ao limiar T_f serão super-resolvidos, caso contrário permanecerão somente com suas componentes já existentes (bloco borrado, $B'_B[i, j]$, que já é fiel ao macrobloco do quadro não-chave original). Matematicamente:

$$\tilde{B}'_{HF}[i, j] = \begin{cases} B'_{SR}[i, j], & \text{se } B'_{HF}[i, j] \geq T_f; \\ B'_B[i, j], & \text{caso contrário.} \end{cases} \quad (3.7)$$

Onde $\tilde{B}'_{HF}[i, j]$ retrata a saída deste macrobloco, sendo super-resolvida ou não.

O novo diagrama incluindo o bloco que retrata o refinamento aqui citado se encontra na Figura 3.19.

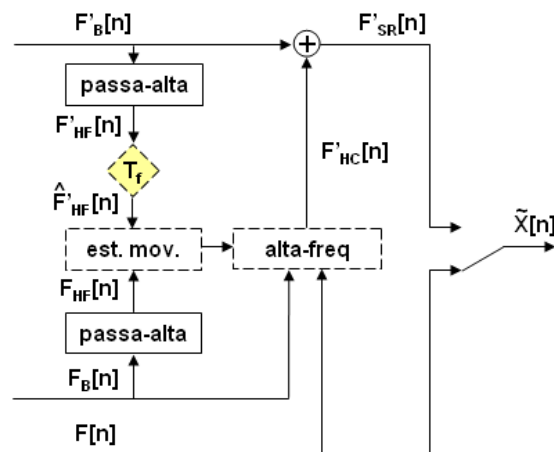


Figura 3.19: Inclusão do bloco com o limiar T_f que analisa a quantidade de alta-frequência dos quadros não-chave borrados ($F'_{HF}[n]$) resultando na decisão de quais macroblocos devem ou não ser super-resolvidos.

O limiar T_f pode variar de 0 a 1, de tal maneira que quanto mais baixo seu valor, mais macroblocos serão super-resolvidos. Nos casos extremos, sendo o valor de T_f igual a 0, todos os macroblocos serão super-resolvidos e sendo seu valor 1, nenhum macrobloco será alterado,

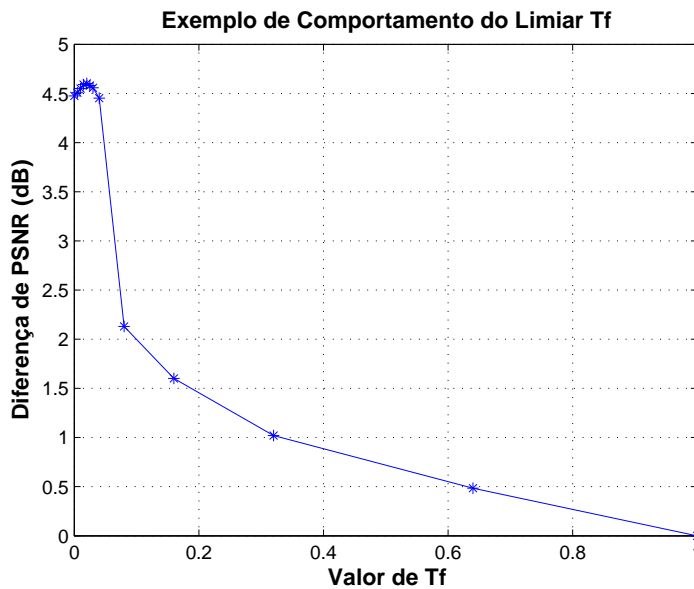


Figura 3.20: Exemplo do comportamento do limiar T_f . A linha do eixo horizontal retrata o limite inferior de qualidade estabelecido pela qualidade do quadro borrado. O ponto de máximo observado nesse caso é aproximadamente $T_f = 0,02$.

permanecendo borrados. Um gráfico ilustrando o comportamento típico do parâmetro T_f tendo como eixo vertical a qualidade objetiva da imagem e o eixo horizontal foi traçado e pode ser visto na Figura 3.20.

Geralmente a qualidade do quadro super-resolvido $F'_{SR}[n]$ tem uma tendência inicialmente ascendente, surgindo logo um ponto de quebra (ainda muito próximo ao eixo vertical) e decaindo em direção à curva de limite inferior que retrata a qualidade do quadro borrado $F'_B[n]$ (retratado na Figura 3.20 como o próprio eixo horizontal). Na maioria dos casos testados neste projeto, a curva de qualidade das sequências super-resolvidas raramente ficou abaixo da curva dos quadros interpolados como será visto no Capítulo 4. Um valor mais adequado para as sequências de vídeo testadas foi determinado empiricamente e os respectivos resultados serão mostrados no Capítulo a seguir.

3.3.3 Analisando as Métricas Utilizadas na Estimação de Movimento

Intuitivamente, outra informação que poderia ser útil para melhorar a saída do método de super-resolução proposto seria utilizar a métrica de combinação de blocos (SAD ou SSD, nesse

caso) como critério de decisão para a adição de alta-frequência. Para isso, primeiramente foi realizada uma comparação em relação à eficiência da SAD e da SSD. Cada uma das métricas é calculada baseada na relação mostrada na Equação 3.3 referenciada na Subseção 3.2.3. Seus cálculos são ilustrados, respectivamente, como

$$SAD(x) = \sum_i \sum_j |(x_{ij} - x'_{ij})|, \quad (3.8)$$

$$SSD(x) = \left(\sum_i \sum_j |(x_{ij} - x'_{ij})|^2 \right)^{1/2}. \quad (3.9)$$

Como mostrado nas equações acima, dependendo das dimensões do bloco onde se quer calcular a métrica, o resultado pode ser bem distoante pois quanto maior é o bloco, geralmente maior é a SAD/SSD não sendo apropriada a comparação entre blocos de tamanhos diferentes. Para que fosse realizada uma comparação mais justa, as métricas foram normalizadas sendo divididas pela quantidade de *pixels* de cada macrobloco (simbolizado adiante pela multiplicação das dimensões vertical e horizontal do bloco, I e J , respectivamente), ou seja,

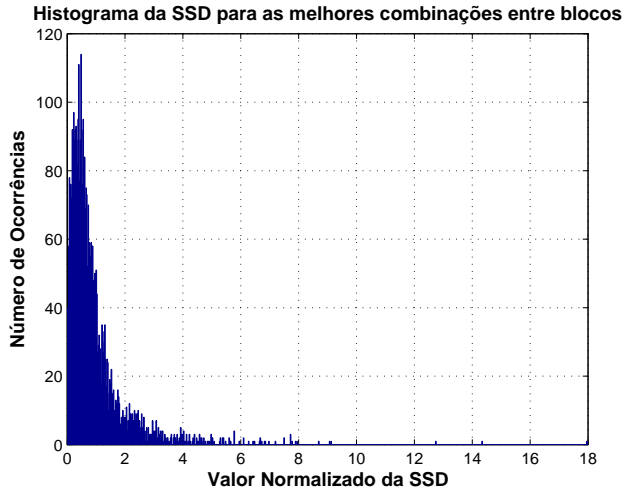
$$SAD_{norm}(x) = \frac{\sum_i \sum_j |(x_{ij} - x'_{ij})|}{IJ}, \quad (3.10)$$

e

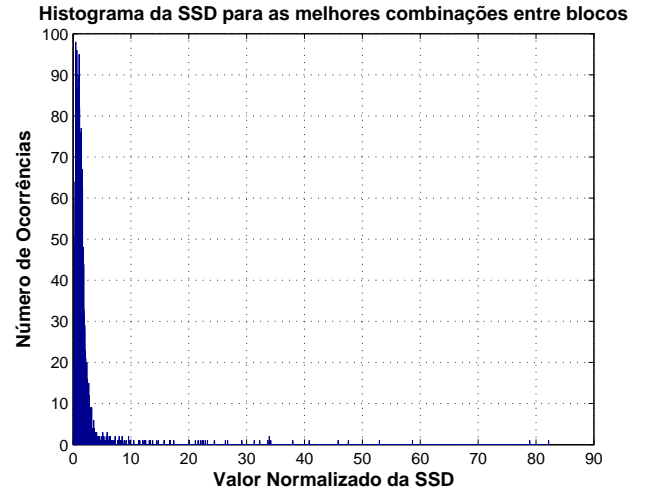
$$SSD_{norm}(x) = \frac{\left(\sum_i \sum_j |(x_{ij} - x'_{ij})|^2 \right)^{1/2}}{IJ}. \quad (3.11)$$

Foram então levantadas estatísticas utilizando SAD e SSD como métricas de comparação na estimação de movimento. Alguns resultados podem ser vistos na Tabela 3.4 a seguir.

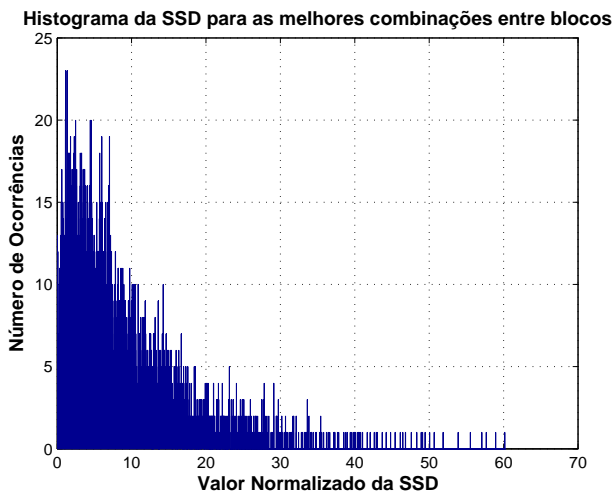
Dada a pouca diferença de performance, a SSD foi arbitrariamente escolhida como métrica na estimação de movimento neste projeto. Outra estatística foi então levantada retratando os valores normalizados da SSD para cada melhor combinação de macroblocos entre quadros chave e não-chave. Os resultados encontrados se encontram nas Figuras 3.21 a 3.22.



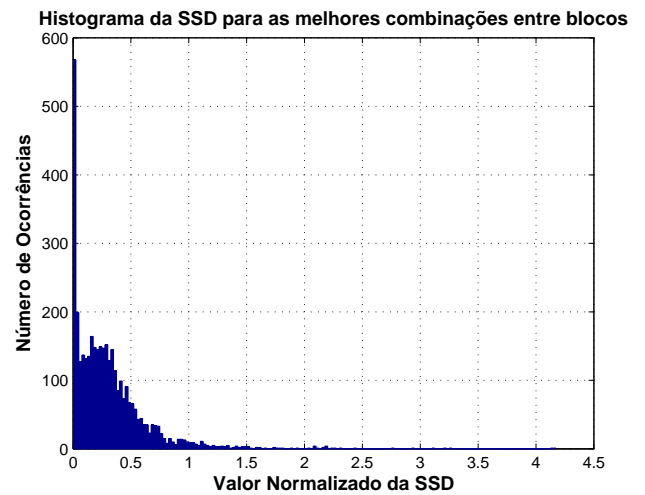
(a)



(b)



(c)



(d)

Figura 3.21: Histograma mostrando os valores normalizados da SSD para as melhores combinações entre blocos dos quadros chave e não-chave para as sequências (a) *Foreman*, (b) *Coastguard*, (c) *Mobile* e (d) *Mother and Daughter*.

Métrica/Sequência	Foreman	Coastguard	Mobile	Mother	Akyio	Football
SAD	34,513	32,0487	23,1840	40,1351	40,9714	31,7167
SSD	34,7690	32,0000	23,0999	40,1578	40,9791	31,7063

Tabela 3.4: Tabela exemplificando os resultados de qualidade objetiva ao fazer super-resolução utilizando cada uma das métricas para estimação de movimento (SAD ou SSD). Perceba que as duas técnicas produzem resultados muito próximos e, portanto, não é concluído um melhor método para esta etapa.

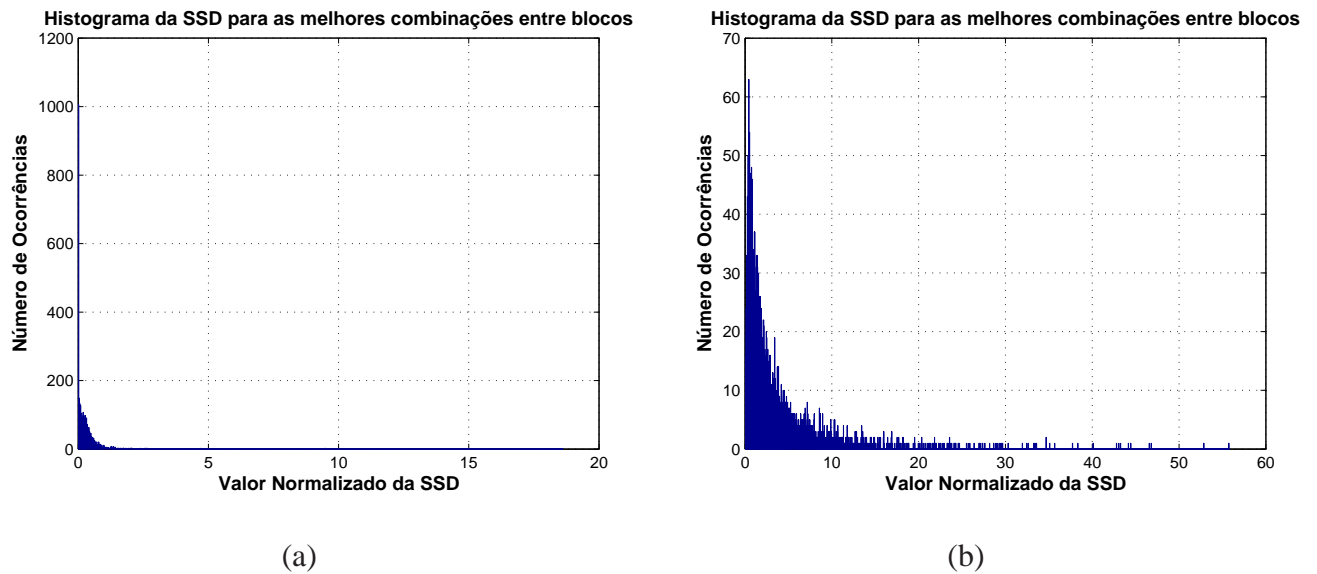


Figura 3.22: Histograma mostrando os valores normalizados da SSD para as melhores combinações entre blocos dos quadros chave e não-chave para as sequências (a) *Akyio* e (b) *Football*.

É possível constatar pelas Figuras supracitadas que apesar do formato recorrente dos histogramas, dependendo da sequência de vídeo, os valores normalizados da SSD se alteram demasiadamente (veja as sequências *Coastguard*, *Mobile* e *Football* em relação às outras). Ainda assim, alguns limiares T_{ssd} foram testados de modo que caso a SSD normalizada de um bloco ultrapassasse este limiar, tal macrobloco não seria super-resolvido.

A Tabela 3.5 mostra que para as três sequências com SSD normalizada mais alta (*Coastguard*, *Mobile* e *Football*), foram utilizados alguns limiares T_{ssd} . Pode-se constatar que mesmo cortando os valores mais altos da SSD (que indicam uma combinação não tão boa), o resultado final retrata alguns ganhos insignificantes porém a partir deste ponto a PSNR começa a cair.

T_{ssd} /Sequência	Coastguard	Mobile	Football
sem limiar	32,0000	23,0999	31,7063
50	31,9892	23,1043	31,7065
40	31,9932	23,0897	31,7071
30	31,9709	23,0621	31,7078
20	31,8994	22,9053	31,7001

Tabela 3.5: Tabela ilustrando o efeito do uso de diversos valores para o limiar T_{ssd} .

Dado o comportamento em relação à diminuição da qualidade objetiva das imagens ao se utilizar limiares para a SSD, a idéia de manter T_{ssd} foi desconsiderada no esquema final da super-resolução proposta. Apesar da não utilização deste passo no método de super-resolução, tal etapa foi citada nesta Subseção para efeito de conhecimento mostrando a tentativa de explorar tal idéia. No Capítulo 5 esta Subseção será citada como fonte de estudos futuros.

3.3.4 Utilizando a Filtragem de Texturas

Esta proposta de refinamento consiste na utilização de uma última filtragem antes da adição da alta-frequência nos quadros não-chave. Após o processo da estimação de movimento e da escolha da melhor combinação entre os macroblocos dos quadros chave e não-chave, a alta-frequência provinda do macrobloco escolhido $B_{HF}[k, l]$ a ser adicionada no bloco $B'_B[i, j]$ pode conter elementos indesejáveis ao quadro não-chave. O intuito deste filtro de texturas, neste documento referido como *filtro sigma*, é tentar aguçar a alta-frequência e minimizar os ruídos e texturas providas dos macroblocos do quadro chave.

Ao se definir a melhor combinação entre macroblocos na estimação de movimento, usualmente a SSD encontrada não é nula, indicando que o casamento entre os dois blocos não foi perfeito. Dado que a métrica que calcula o índice de semelhança entre dois blocos não acusa uma combinação exata nas frequências mais baixas, é de se esperar que as frequências mais altas dos dois blocos também não sejam iguais. Porém, como no lado do decodificador não há informações precisas sobre a alta-frequência do quadro não-chave, a dica mais fiel que se tem disponível para a tarefa da estimação provém dos macroblocos borrados como já explicado nas Subseções

anteriores.

A idéia geral desta etapa é adicionar toda a alta-frequência do macrobloco do quadro chave quando a SSD se mostrar nula ou abaixo de um limiar T_σ pré-estabelecido, e adicionar somente parte da alta-frequência quando o valor da SSD da melhor combinação se encontrar além deste limiar. O diagrama incluindo o bloco do filtro sigma se encontra na Figura 3.23 a seguir.

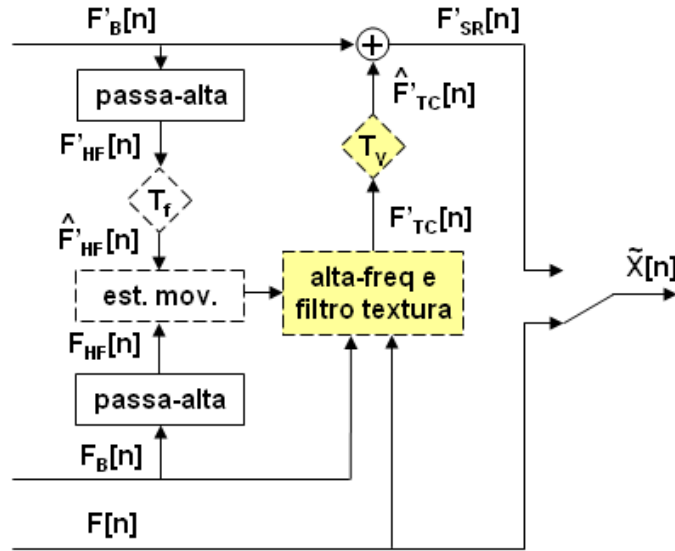


Figura 3.23: Detalhe do diagrama destacando o filtro sigma. A alta-frequência dos quadros chave passam por uma filtragem antes de serem adicionadas ao quadro não-chave dependendo da decisão tomada baseada no valor da SSD normalizada e no limiar T_σ .

Vale lembrar que como visto na Figura 3.23, a alta-frequência que será adicionada ao quadro não-chave provém da alta-frequência complementar descrita na Equação 3.4 na Subseção 3.2.4 e não da alta-frequência filtrada resultante da convolução do quadro chave pela máscara $M_{alta}[n]$ (Equação 3.6). Portanto, o filtro sigma será aplicado ao quadro de altas-frequências estimado ($F'_{HC}[n]$) explicado na Seção 3.2.4.

O filtro sigma consiste em diversas etapas. Em um primeiro momento, é realizada uma varredura em toda a imagem analisando regiões de 3×3 *pixels* do quadro $F'_{HC}[n]$. Se os valores extremos (máximo e mínimo) dos *pixels*, aqui designados como p_{max} e p_{min} , dessa região ultrapassarem o intervalo definido pelo valor médio dessa mesma região, chamado p_{medio} , e por uma ponderação pré-definida designada como β , então o *pixel* central, $p_{central}$, dessa região se mantém intacto. Caso contrário, $p_{central}$ assumirá o valor médio já calculado p_{medio} . Esta relação

então pode ser abreviada matematicamente pela Equação 3.12:

$$\tilde{p}_{central} = \begin{cases} p_{medio}, & \text{se } p_{medio}(1 - \beta) \leq p_{max}, p_{min} \leq p_{medio}(1 + \beta); \\ p_{central}, & \text{caso contrário.} \end{cases} \quad (3.12)$$

Estes *pixels* processados ($\tilde{p}_{central}$) vão sendo armazenados separadamente em uma nova matriz chamada $F'_P[n]$ durante o processo para que *pixels* vizinhos não sofram interferência dos *pixels* já alterados. O fator β pode variar de 0 a 1 porém o valor que produz empiricamente a melhor saída (considerando a qualidade objetiva das imagens super-resolvidas) nas sequências testadas foi $\beta = 0,05$.

Após percorrer todos os *pixels* para encontrar $F'_P[n]$, o próximo passo é realizar uma outra operação envolvendo este quadro $F'_P[n]$, o quadro original de alta-frequência $F'_{HC}[n]$ e uma constante de ponderação pré-estabelecida chamada γ . O quadro resultante dessa operação é designado como $F'_{TC}[n]$ e sua relação pode ser vista na Equação a seguir.

$$F'_{TC}[n] = F'_{HC}[n] - \gamma \cdot F'_P[n]. \quad (3.13)$$

O quadro $F'_{TC}[n]$ possui majoritariamente as altas-frequências de $F'_{HC}[n]$ porém sem alguns detalhes tais como texturas e outros pequenos ruídos. Um exemplo dessa subtração pode ser visto nas Figuras 3.24 e 3.25.

Métrica/Sequência	Foreman	Coastguard	Mobile	Mother	Akyio	Football
sem filtro sigma	34,7690	32,0000	23,0999	40,1578	40,9791	31,7063
com filtro sigma	34,8387	31,9866	23,1035	40,1513	40,9871	31,8895

Tabela 3.6: Tabela exemplificando os resultados de qualidade objetiva ao fazer super-resolução sem e com filtro sigma.

Como dito anteriormente, existe ainda um limar T_σ que define o valor máximo de SSD normalizada para que o macrobloco não sofra influência do filtro sigma pois *a priori* quanto mais baixa a SSD, mais fiel é o casamento entre os macroblocos do quadro chave e não-chave e, conseqüentemente, mais semelhante se deduz também ser a componente de alta-frequência.



Figura 3.24: Exemplo de quadro de alta-frequência $F'_{HC}[n]$.



Figura 3.25: Exemplo de quadro de alta-frequência processada por um filtro sigma $F'_{TC}[n]$.

O limiar definido neste projeto foi estipulado em $T_\sigma = 0,5$. Então, qualquer macrobloco que possuir sua melhor combinação com SSD normalizada ultrapassando esse valor T_σ terá sua alta-frequência retirada de $F'_{TC}[n]$ e não mais de $F'_{HC}[n]$.

A Tabela 3.6 mostra a influência do filtro sigma aqui descrito nas sequências de vídeo testadas. Perceba que dependendo da sequência de vídeo, o filtro sigma produz resultados iguais ou melhores ao método sem ele. Na média, embora existam casos em que não há ganhos nem perdas com o uso desta filtragem, seu uso mostra ser vantajoso pois eventualmente melhora a qualidade da imagem super-resolvida. Por ter obtido ganhos discretos, foi deixada a cargo do interesse do usuário a opção de utilização deste filtro como pós-processamento.

3.4 APLICANDO OS REFINAMENTOS AO MÉTODO BÁSICO DA SUPER-RESOLUÇÃO PROPOSTA

Após a explanação da idéia básica da super-resolução proposta e dos refinamentos sugeridos nas Seções 3.2 e 3.3 anteriores, serão lembrados e resumidos aqui o diagrama de blocos e a explicação do método final de super-resolução espelhando o modo em que o mesmo se encontra implementado atualmente.

No lado do codificador as etapas são as mesmas já explicadas na Seção 3.2. A Figura 3.7 será oportunamente repetida na Figura 3.26 abaixo e mostra o respectivo diagrama de blocos no lado do codificador.

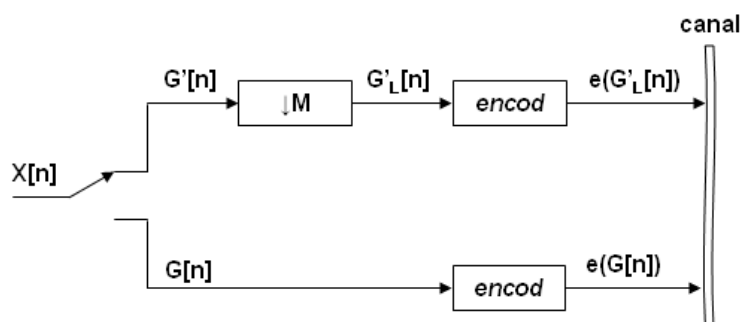


Figura 3.26: Diagrama de blocos completo no lado do codificador.

Como visto, quadros chave $\{G[n]\}$ são simplesmente codificados. Quadros não-chave

$\{G'[n]\}$ são decimados por um fator M e codificados em seguida.

No lado do decodificador, além da idéia exposta na Seção 3.2 existem também os processos explicados na Seção 3.3. Todo o conjunto será compilado sucintamente a seguir e é possível ver o diagrama de blocos de todo o processo no lado do decodificador na Figura 3.27.

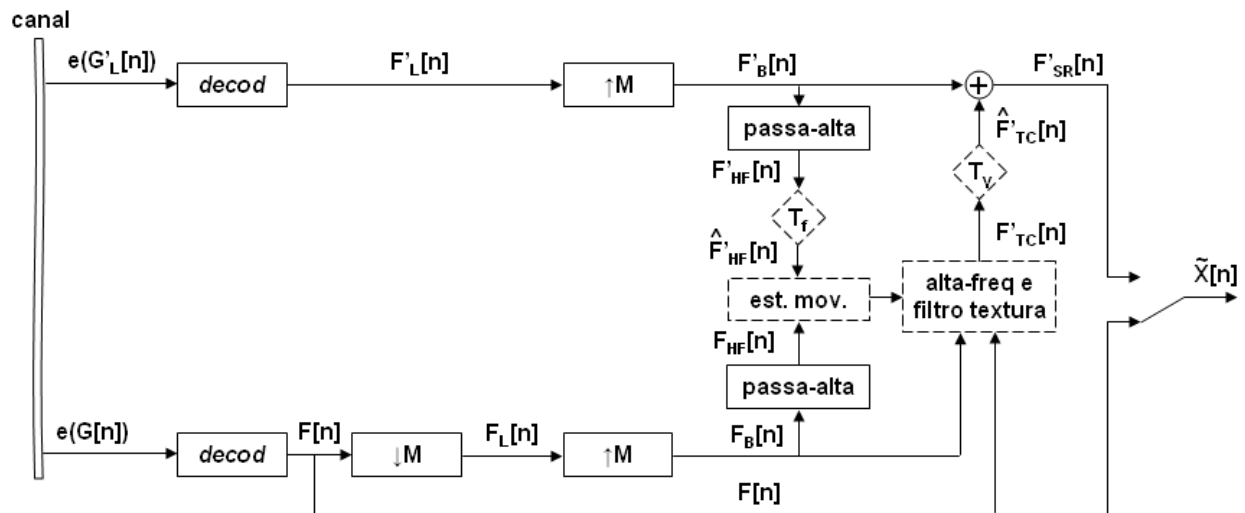


Figura 3.27: Diagrama de blocos completo no lado do decodificador incluindo as etapas de refinamento do método de super-resolução proposto. Os blocos tracejados do diagrama representam as operações realizadas em nível de macrobloco.

Primeiramente, quadros chave e não-chave são decodificados normalmente, cada um em sua resolução atual. Em seguida, quadros não-chave $\{F'_L[n]\}$ são interpolados por um fator M (mesmo do codificador) para voltar a sua resolução original resultando em um quadro não-chave borrado $\{F'_B[n]\}$. Já os quadros chave $\{F[n]\}$ seguem por dois caminhos diferentes. Em um dos caminhos eles são decimados e interpolados com o mesmo fator M para simular a mesma degradação sofrida pelos quadros não-chave gerando um quadro chave borrado designado como $F_B[n]$. O outro caminho conduz os quadros chave $\{F[n]\}$ diretamente até o final do esquema onde serão posteriormente utilizados.

Os quadros chave e não-chave borrados, respectivamente $\{F_B[n]\}$ e $\{F'_B[n]\}$, são desviados para um caminho onde sofrem um processamento por um filtro passa-altas resultando nos quadros $\{F_{HF}[n]\}$ e $\{F'_{HF}[n]\}$. Somente macroblocos dos quadros não-chave que possuam quantidade de alta-frequência acima de um limiar T_f seguirão para a próxima etapa da estimação de movimento.

A estimação de movimento procura o melhor casamento entre um macrobloco do quadro

não-chave $F'_{HF}[n]$ e um macrobloco do quadro chave $F_{HF}[n]$ encontrando o deslocamento mais apropriado quando for calculada a menor SSD entre os dois blocos. A partir das melhores combinações definidas na estimação de movimento, o próximo passo é super-resolver os quadros não-chave a partir da alta-frequência disponível nos quadros chave.

Voltando a observar o fluxo inicialmente seguido pelos quadros chave, realiza-se um processamento envolvendo os quadros chave $\{F[n]\}$, suas respectivas versões borradas $\{F_B[n]\}$ e ainda os vetores de movimento obtidos na estimação de movimento. Nesta etapa, é alcançado um novo quadro não-chave de altas-frequências chamado $F'_{HC}[n]$. Tais quadros podem ser processados por um filtro sigma (opcional) que tenta eliminar as texturas e ruídos indesejáveis. Baseado em um limiar T_σ e nos valores das SSD das melhores combinações entre os macroblocos, os blocos do quadro não-chave podem ser super-resolvidos diretamente com a alta-frequência do quadro $F'_{HC}[n]$ (caso a SSD tenha valor abaixo de T_σ) ou do quadro $F'_{TC}[n]$ (caso a SSD tenha valor acima de T_σ). Após decidir sobre as componentes de alta-frequência que serão adicionadas aos macroblocos dos quadros não-chave borrados $F'_B[n]$ o próximo passo é super-resolvê-los fazendo as adições de alta-frequência de fato.

No final do diagrama é possível ver a saída do sistema $\{\tilde{X}[n]\}$ representando a sequência de vídeo composta de quadros chave $\{F[n]\}$ e quadros não-chave super-resolvidos $\{F'_{SR}[n]\}$.

A seguir serão ilustrados os resultados obtidos através do processo de super-resolução proposto neste documento. Mais informações sobre valores de parâmetros, condições de testes e outros detalhes que porventura existirem serão divulgados no Capítulo 4 .

4 EXPERIMENTOS E RESULTADOS PRÁTICOS

Este Capítulo visa apresentar os resultados obtidos através do método de super-resolução proposto e explicado no Capítulo 3. Os detalhes em relação ao cenário considerado, parâmetros utilizados no algoritmo e condições de teste serão mencionados juntamente aos seus respectivos gráficos e figuras.

Todos os testes foram realizados utilizando o programa *MatLab* versão 7.4 e 6.5 [36]. A única parte do código fonte portada para um arquivo *mex* (linguagem *C*) foi a etapa da estimação de movimento por se tratar de um processo extremamente complexo. Para simular os diversos graus de degradação devido à codificação do vídeo, foi utilizado o padrão *H.264 JM 10.2* [37] como *codec* por atender bem às necessidades deste projeto e ainda por ser o padrão de codificação estado-da-arte.

As sequências de vídeo testadas foram: *Foreman*, *Coastguard*, *Mobile Calendar*, *Mother and Daughter*, *Akyio* e *Football* apresentadas visualmente na Figura 4.1 que se segue (são ilustrados o primeiro quadro de cada uma das sequência). Com o propósito de validar o método, tais sequências foram escolhidas por possuírem diversos graus de quantidade de movimento, elementos novos/recorrentes em cena e quantidade/intensidade de alta-frequência impondo assim diversos níveis de dificuldade à super-resolução proposta.

A resolução natural dos quadros é de 352×288 *pixels* (*CIF*) sendo decimados por um filtro bilinear com um fator de redução $M = 2$ para cada dimensão alcançando quadros sub-amostrados com $1/4$ da resolução original, ou seja, 176×144 *pixels* (*QCIF*). Para os testes foi considerada somente a componente de luminância (sistema *YUV* [38]) dos quadros.

Vale lembrar que todas as figuras e curvas mostradas a seguir se referem somente aos resultados da super-resolução nos quadros não-chave, não sendo nunca contabilizadas a PSNR dos quadros chave pois isso encobriria os resultados reais do método.



(a)



(b)



(c)



(d)



(e)



(f)

Figura 4.1: Sequências de vídeo: (a) *Foreman*, (b) *Coastguard*, (c) *Mobile*, (d) *Mother and Daughter*, (e) *Akyio* e (f) *Football*.

4.1 ALTERANDO A DENSIDADE DE QUADROS CHAVE

Os primeiros testes foram realizados variando a densidade de quadros chave em relação aos quadros não-chave. Para todas as sequências foram sempre utilizados 30 quadros ao total (chave e não-chave) e as densidades são indicadas levando em consideração a quantidade de quadros chave dentro dos 30 quadros totais. Foram testadas então quatro densidades diferentes: 1/30, 3/30, 6/30 e 15/30, sendo o numerador a quantidade de quadros chaves no teste. Como dito anteriormente, a distribuição de quadros chave é temporalmente uniforme.

Além dos detalhes já citados, foram fixados alguns outros parâmetros sendo eles:

- Método de interpolação: filtragem bicúbica;
- Tamanho de macroblocos: 16x16 e 8x8 *pixels*;
- Parâmetro de Quantização (QP) do *H.264*: 30.

Os resultados objetivos se encontram ilustrados nos gráficos das Figuras 4.2 a 4.7.

4.1.1 Análise de resultados

É possível perceber pelas Figuras desta Seção que mesmo com uma densidade muito baixa de quadros chaves na sequência, na maioria dos casos (com exceção das sequências *Coastguard* e *Football*), os quadros não-chave conseguem ser super-resolvidos de forma a sempre resultarem em imagens com uma qualidade objetiva maior do que a imagem borrada disponível no lado do decodificador.

Considerando o cenário destes testes, o tamanho de macrobloco que apresentou melhor desempenho foi 16x16 *pixels*. O filtro sigma apesar de na média apresentar discretos ganhos em relação ao método sem esta filtragem, foi incluído nos gráficos para apreciação. Como os parâmetros do filtro sigma (β , γ e T_σ) utilizados foram os mesmos para todas as sequências então é certo que com um ajuste mais direcionado para cada tipo de sequência, o desempenho melhora de forma mais significativa.

A sequência *Football* apresentou desempenho bem inferior ao restante das sequências por possuir baixa correlação entre-quadros (ver Tabelas 3.1 e 3.2), quadros com alto grau de

renovação de elementos em cena e objetos naturalmente muito borrados (ocasionado pela velocidade da ação da cena real não comportada pelo dispositivo de captura). Portanto, a baixa densidade de quadros chave e a pouca quantidade de alta-frequência disponível resultaram em poucas e más referências para super-resolver os quadros não-chave. Um fator agravante, também percebido nos resultados da sequência *Coastguard* é a presença de muitas texturas (gramado do campo e superfície da água) dificultando o processo de super-resolução.

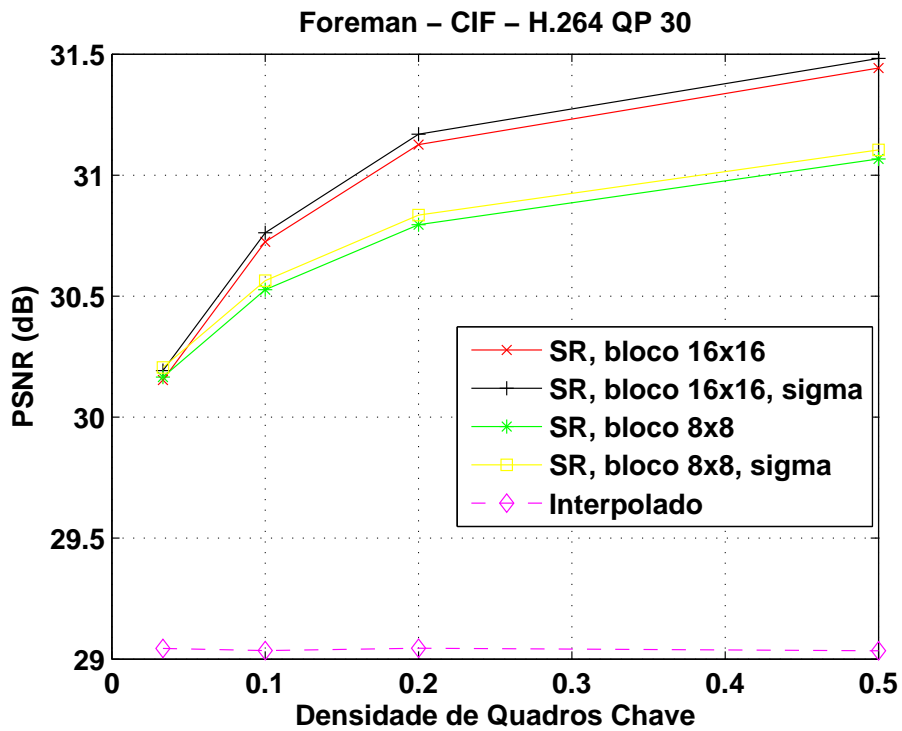


Figura 4.2: Gráfico com os resultados variando a densidade de quadros chave para a sequência *Foreman*.

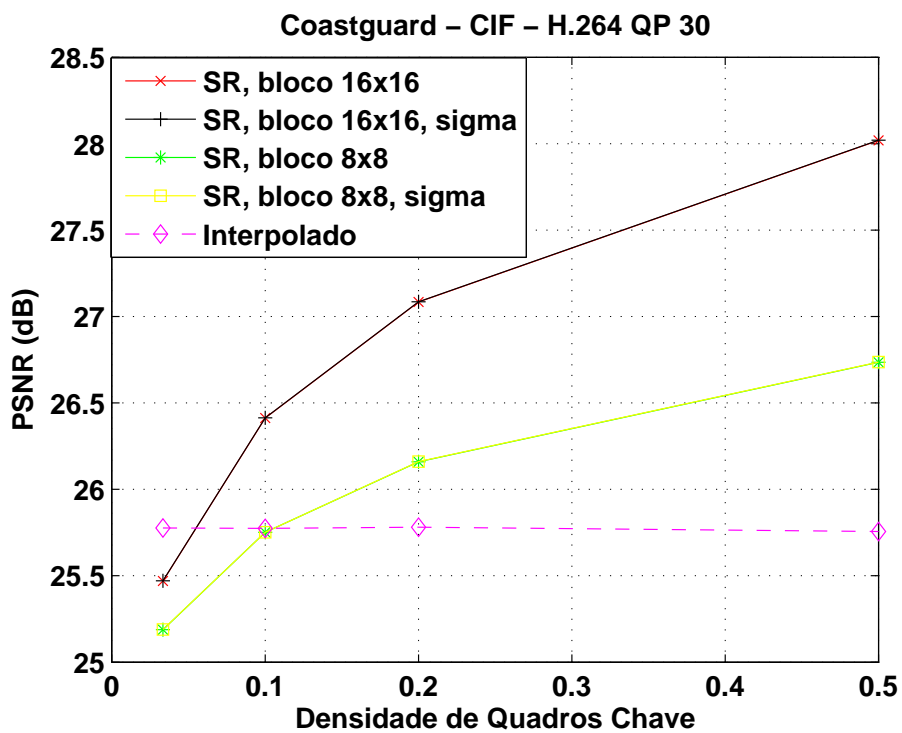


Figura 4.3: Gráfico com os resultados variando a densidade de quadros chave para a sequência *Coastguard*.

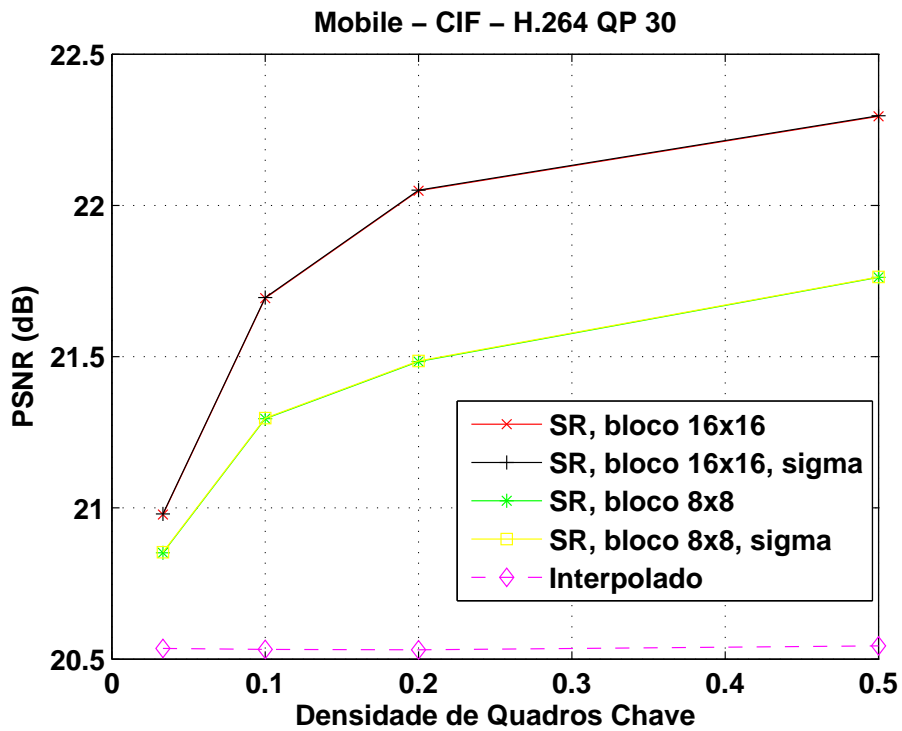


Figura 4.4: Gráfico com os resultados variando a densidade de quadros chave para a sequência *Mobile Calendar*.

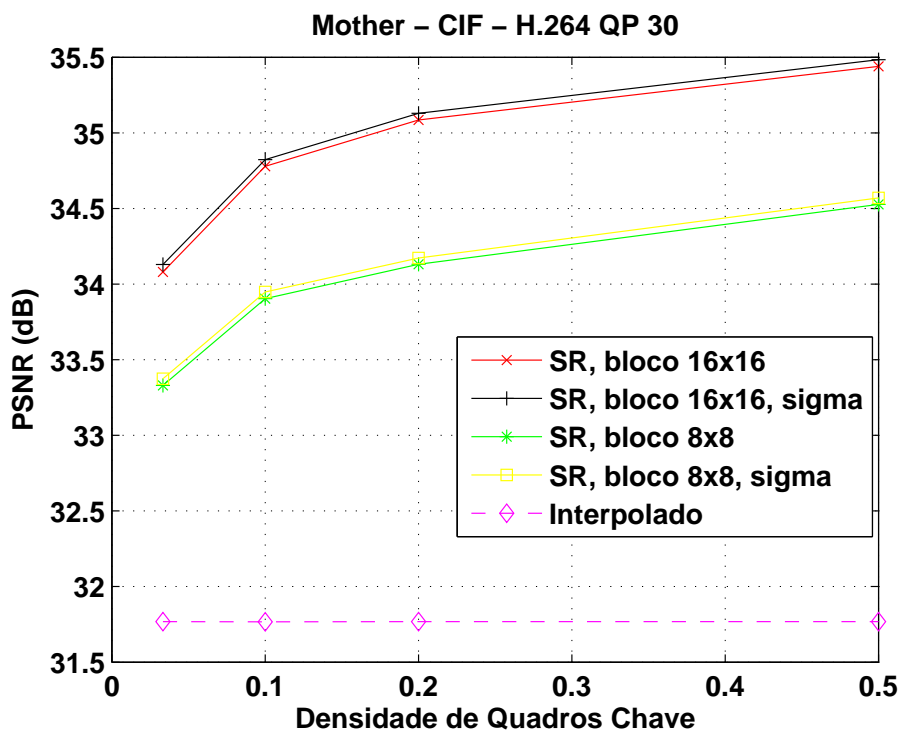


Figura 4.5: Gráfico com os resultados variando a densidade de quadros chave para a sequência *Mother and Daughter*.

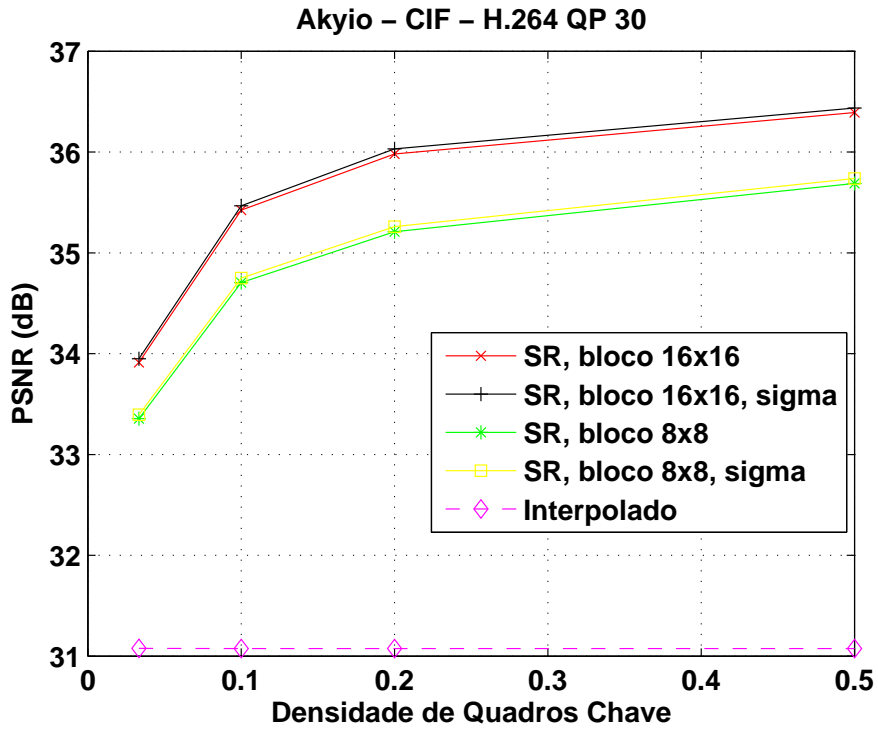


Figura 4.6: Gráfico com os resultados variando a densidade de quadros chave para a sequência *Akyio*.

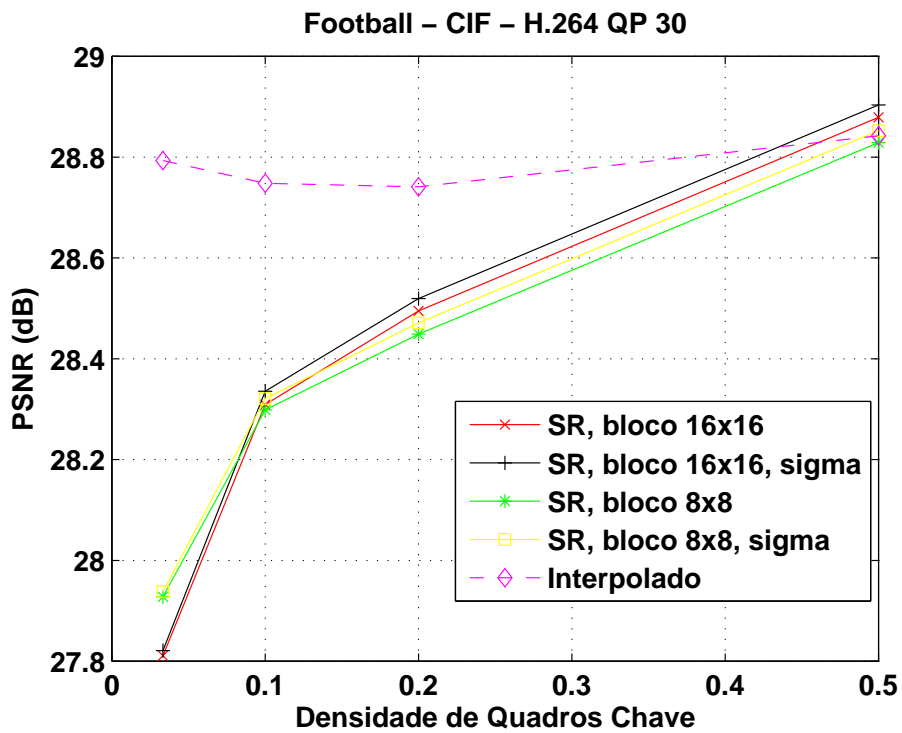


Figura 4.7: Gráfico com os resultados variando a densidade de quadros chave para a sequência *Football*.

4.2 ALTERANDO A COMPRESSÃO DOS QUADROS

Testes foram realizados variando a taxa de compressão aplicada pelo codificador *H.264*. A super-resolução proposta foi testada em sequências degradadas com quatro parâmetros de quantização (QP) diferentes, sendo eles: 10, 20, 30 e 40. Quanto maior o QP, menor é a taxa e maior é a degradação. Nos gráficos serão mostradas as curvas de qualidade (PSNR) em relação a relação a taxa (kbps, do inglês *kilobits per second*).

Além dos detalhes já citados, foram fixados alguns outros parâmetros sendo eles:

- Método de interpolação: filtragem bicúbica;
- Tamanho de macroblocos: 16x16 e 8x8 *pixels*;
- Densidade de quadros chave: 15/30.

Os resultados objetivos se encontram ilustrados nos gráficos das Figuras 4.8 a 4.13.

4.2.1 Análise de resultados

As curvas apresentadas nos gráficos das Figuras desta Seção incluem duas curvas de referência: a curva inferior representando a qualidade do quadro borrado e a curva superior indicando o quadro codificado com o padrão estado-da-arte *H.264 intra* na resolução CIF original. É possível ver que para sequências mais estáticas como *Mother and Daughter* e *Akyio* o desempenho é próximo ao da curva do *H.264 intra*. Sequências de dificuldade mediana tais como *Foreman* e *Coastguard* apresentam resultados objetivos satisfatórios ainda lembrando que com o método de super-resolução proposto, a complexidade no codificador decai consideravelmente em relação à codificação de todos os quadros em suas resoluções originais. Nas sequências *Mobile* e *Football*, sendo vídeos mais difíceis de se comprimir por possuírem grande quantidade de movimento e novos elementos a cada cena, o desempenho é mais modesto não apresentando ganhos tão expressivos porém, no geral, o desempenho é melhor que o caso interpolado mesmo no caso de quadros chave extremamente degradados ($QP = 40$, por exemplo). Ainda no caso das sequências mais complexas, foi percebido também a mudança de desempenho em relação ao

tamanho do macrobloco, sobressaindo o desempenho dos blocos de tamanho 8×8 *pixels* pois os mesmos conseguem abranger melhor os detalhes presentes nessas sequências.

É válido lembrar que a sequência *Football* apresenta desempenho melhor que a curva interpolada pois se trata da maior densidade de quadros testada (proporção de 1:1 entre quadros chave e não-chave), ou seja, aqui são retratadas as várias taxas para os pontos do extremo a direita do gráfico 4.7.

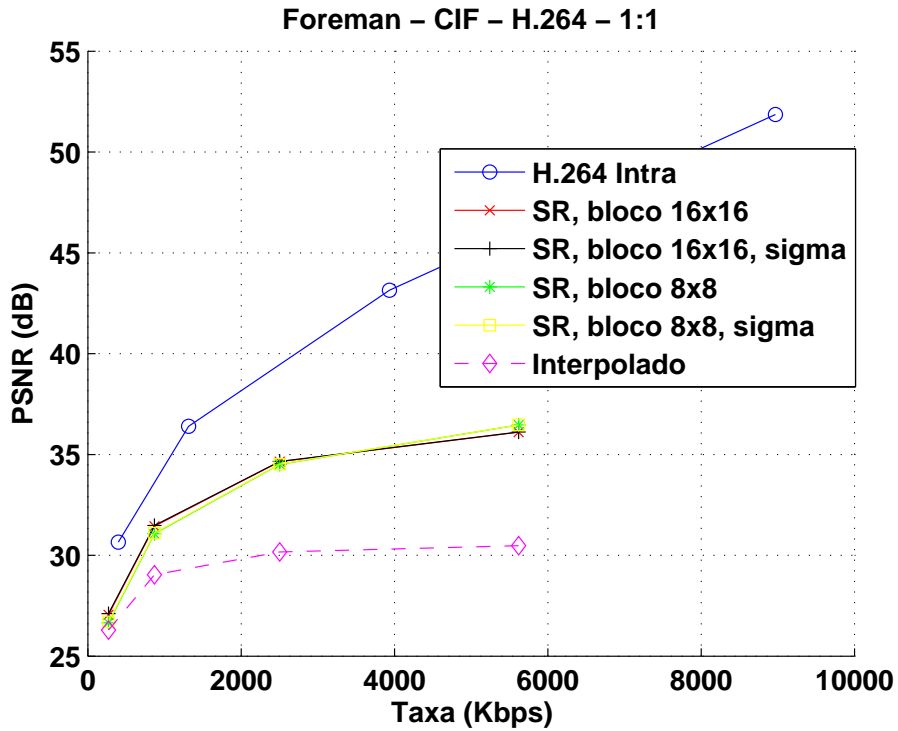


Figura 4.8: Gráfico com os resultados variando o QP na codificação dos quadros para a sequência *Foreman*.

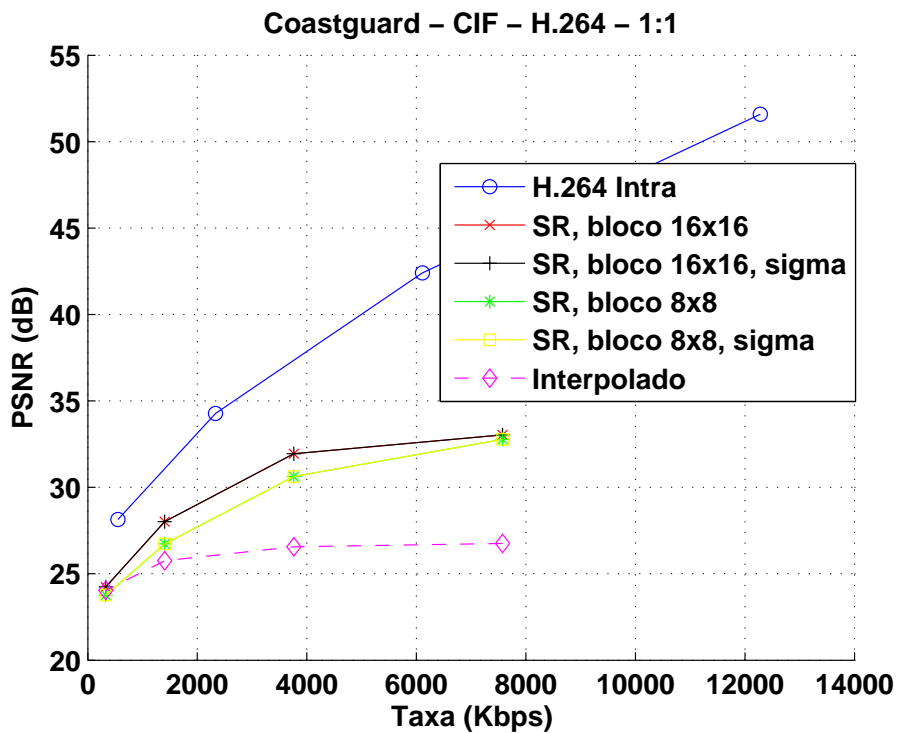


Figura 4.9: Gráfico com os resultados variando o QP na codificação dos quadros para a sequência *Coastguard*.

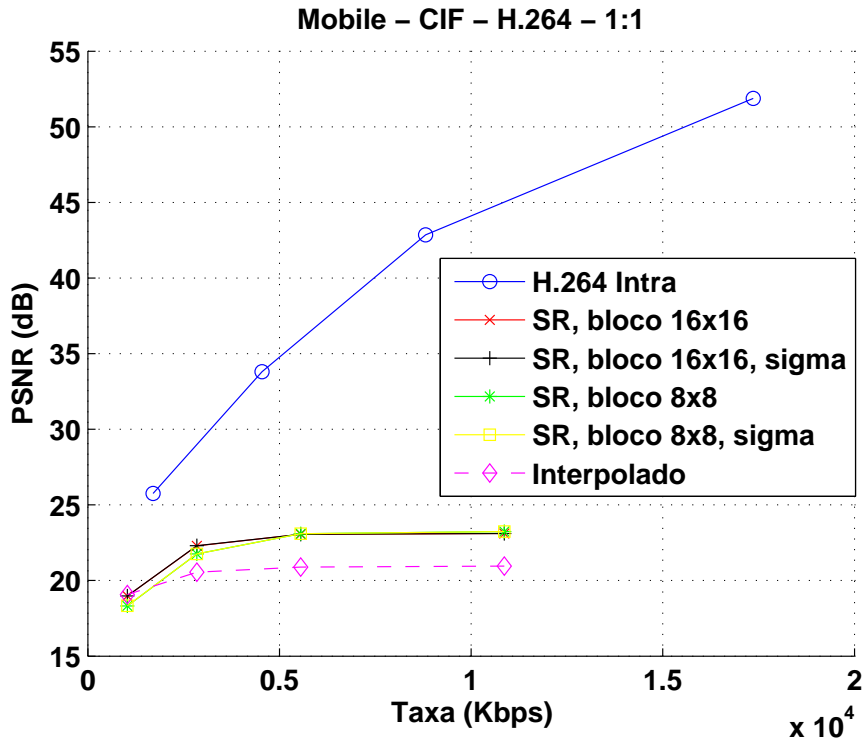


Figura 4.10: Gráfico com os resultados variando o QP na codificação dos quadros para a sequência *Mobile Calendar*.

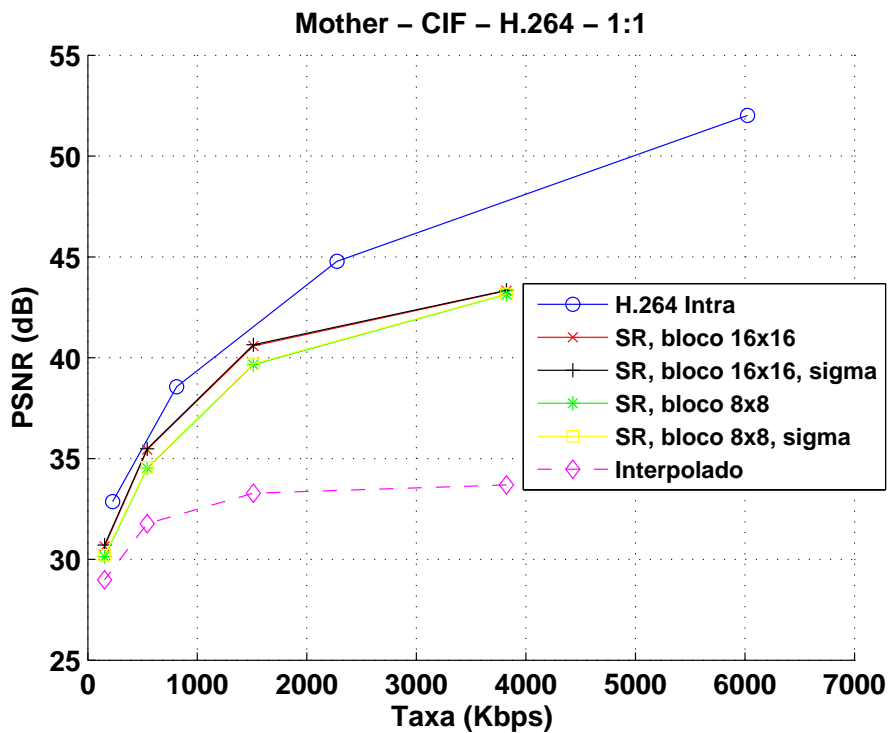


Figura 4.11: Gráfico com os resultados variando o QP na codificação dos quadros para a sequência *Mother and Daughter*.

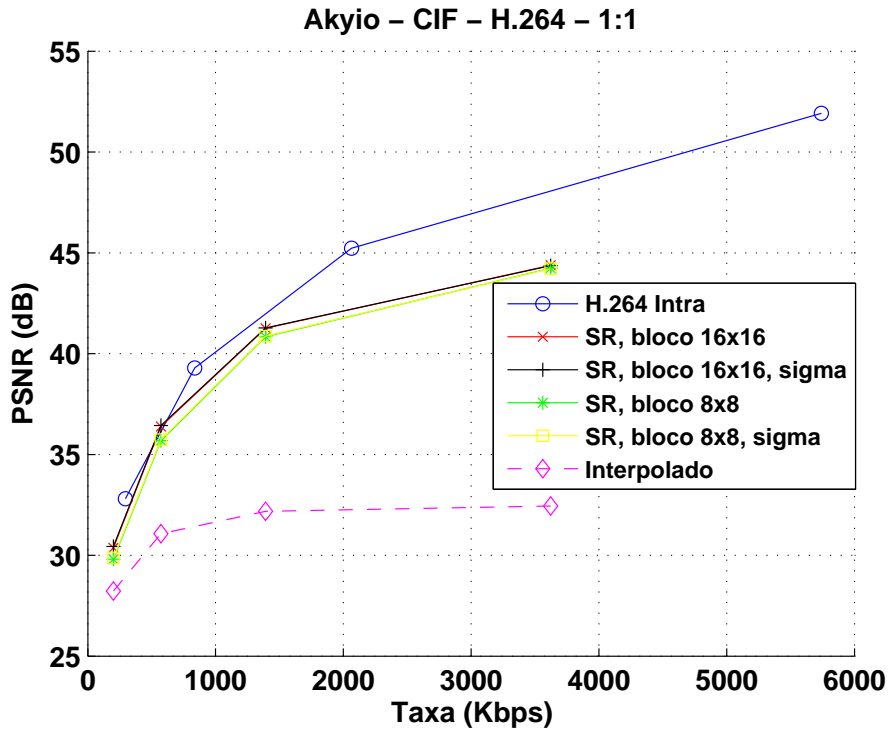


Figura 4.12: Gráfico com os resultados variando o QP na codificação dos quadros para a sequência *Akyio*.

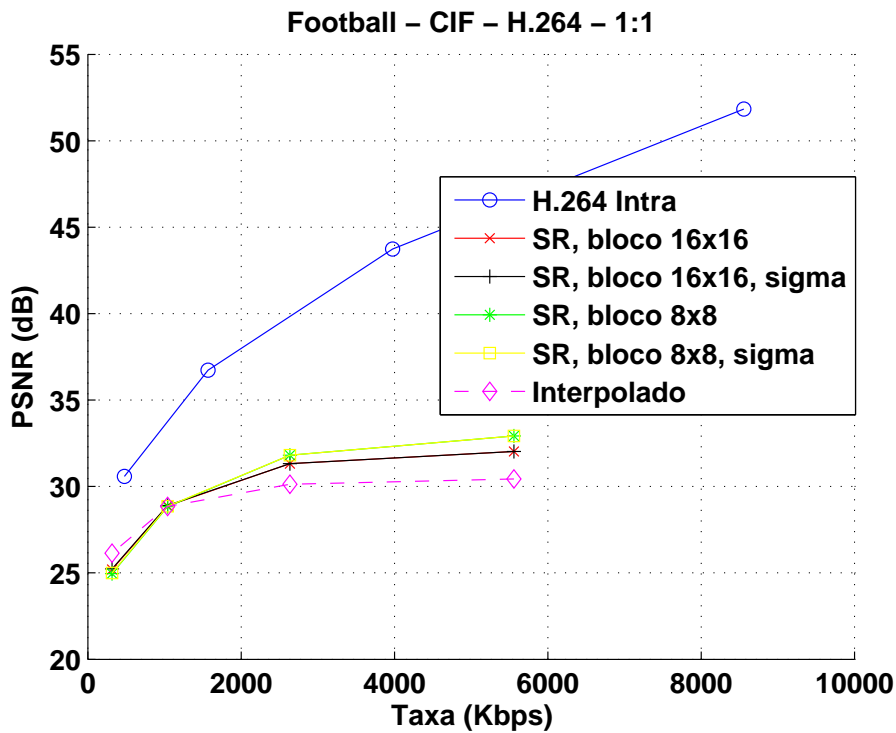


Figura 4.13: Gráfico com os resultados variando o QP na codificação dos quadros para a sequência *Football*.

4.3 RESULTADOS SUBJETIVOS

Alguns resultados subjetivos serão apresentados nessa Subseção para avaliação visual da imagem final super-resolvida. Serão apresentadas sempre três imagens do quadro não-chave para comparação: (a) original, (b) interpolada e (c) super-resolvida. As Figuras 4.14 a 4.17 mostram resultados subjetivos das respectivas sequências: *Foreman*, *Coastguard*, *Mobile Calendar* e *Football*.

Os parâmetros para super-resolver tais sequências foram os seguintes:

- Método de interpolação: filtragem bicúbica;
- Tamanho de macroblocos: 16x16 *pixels*;
- Densidade de quadros chave: 15/30;
- Parâmetro de quantização (QP) do *H.264*: 20;

4.3.1 Análise de resultados

Observando as Figuras desta Seção é fácil perceber que mesmo para as sequências de maior grau de dificuldade (grande quantidade de movimento e novos elementos a cada cena), as imagens super-resolvidas são apresentadas sem nenhum tipo de artefato, efeito de bloco ou outros detalhes visualmente incômodos. A alta-frequência adicionada nos quadros borrados resultando nos quadros super-resolvidos de fato possuem um bom casamento pois não distoam da componente original. Perceba também que bordas, letras, algumas texturas e detalhes em geral impossíveis de serem identificados no quadro interpolado são claramente visíveis nos quadros super-resolvidos.



(a)



(b)



(c)

Figura 4.14: Quadro 6 da sequência *Foreman*: (a) original, (b) interpolado e (c) super-resolvido.



(a)



(b)



(c)

Figura 4.15: Quadro 6 da sequência *Coastguard*: (a) original, (b) interpolado e (c) super-resolvido.



(a)



(b)



(c)

Figura 4.16: Quadro 4 da sequência *Mobile Calendar*: (a) original, (b) interpolado e (c) super-resolvido.



(a)



(b)



(c)

Figura 4.17: Quadro 4 da sequência *Football*: (a) original, (b) interpolado e (c) super-resolvido.

5 CONCLUSÃO

O presente trabalho propôs uma técnica de super-resolução de vídeo não-Bayesiana baseado em compensação de movimento aplicada a um cenário onde estão presentes quadros de diferentes resoluções [19]. Um exemplo onde se pode observar sequências de vídeo de resolução mista é no novo paradigma de codificação de vídeo dito de complexidade reversa [10, 11]. Este novo esquema de codificação tem o intuito de diminuir a complexidade do codificador e transferi-la para decodificador e se mostra mais apropriado para aplicações onde ocorre a produção de vídeos a partir de equipamentos com poucos recursos computacionais e/ou energéticos.

Considerando o esquema exposto, o cenário permite que certos quadros de uma sequência de vídeo, chamados quadros não-chave, sejam decimados resultando em uma resolução espacial reduzida, enquanto os outros quadros da sequência, chamados quadros chave, permanecessem em sua resolução original. Os quadros não-chave são então posteriormente interpolados de volta a sua resolução original. Dado que as componentes de alta-frequência são perdidas durante as filtragens de redimensionamento, a idéia central desse projeto foi super-resolver os quadros não-chave recuperando sua alta-frequência a partir dos quadros chave. Dado que os quadros chave possuem todas as componentes de frequência e são, em geral, altamente correlacionados com seus vizinhos, foram utilizadas técnicas para estimar a alta-frequência dos quadros não-chave a partir da informação contida nos quadros chave.

A partir da técnica proposta de super-resolução e dos refinamentos estudados, diversos testes foram realizados produzindo bons resultados. Dentro da aplicação de codificação de complexidade reversa, a estimação de movimento entre quadros foi eliminada no lado do codificador do *codec* H.264 e todos os quadros foram codificados *intra* em suas respectivas resoluções (originais ou reduzidas). Já no lado do decodificador foi aplicada a super-resolução exposta neste manuscrito. Diversos parâmetros do sistema foram alterados como a densidade de quadros chave, a degradação sofrida pelos quadros devido à quantização, tamanho de macroblocos, tipos de algoritmos de redimensionamento etc. Os resultados apresentados neste

documento retratam os efeitos finais da técnica proposta.

Os obstáculos encontrados durante o desenvolvimento e execução deste projeto espelham majoritariamente as dificuldades em tratar sequências de vídeo que contém alta quantidade de movimento e alta renovação de objetos em cena. Uma das vantagens do método proposto é a qualidade subjetiva da imagem super-resolvida pois não ocorrem efeitos de bloco, inclusão de artefatos ou outros efeitos indesejáveis mesmo quando a PSNR (métrica objetiva) de fato não retrata o melhor dos resultados. A baixa- e a média-frequência dos quadros não-chave são preservadas nesse método dado que são componente fiéis ao quadro original. Por não se apoiar no paradigma Bayesiano, o método proposto não requer treinamento nem assume distribuições de probabilidade. Ainda, esta técnica pode ser completamente concentrada no lado do decodificador estando de acordo com o paradigma de codificação de complexidade reversa considerado.

5.1 TRABALHOS PUBLICADOS

Com o estudo e desenvolvimento deste projeto alguns artigos foram publicados em conferências na área de processamento de imagens e um outro artigo será em breve publicado em uma revista especializada. Será explicada de forma sucinta a idéia de cada um dos artigos aceitos. O primeiro artigo [39] retratou uma etapa inicial deste projeto de modo que para super-resolver um quadro não-chave era utilizada uma procura em um banco de dados formado com macroblocos dos quadros chave e suas respectivas altas-frequências. O segundo artigo [40] mostrou o resultado da super-resolução já aplicando a estimação de movimento em um contexto mais realista de codificação expondo testes e resultados mais completos. O terceiro artigo [34] publicado mostrou o sistema de super-resolução aqui proposto aplicado em um cenário real de sistema de codificação distribuída e com escalabilidade espacial. O quarto artigo [41] que será ainda publicado em revista mostra novamente a super-resolução aqui proposta como meio de auxiliar a geração de informação lateral em *codecs* com escalabilidade temporal e espacial.

Como visto, o método foi amplamente exposto e validado na comunidade científica mostrando ter utilidade em aplicações em contextos atuais e realistas, e ainda possuindo flexibilidade para ser utilizada em diversos tipos de cenários além do aqui retratado. Como já exposto anteriormente,

este método não tem pretensão de se erguer como uma técnica de codificação propriamente dita concorrendo diretamente com os métodos de codificação já conhecidos. Porém, como mostrado nos artigos supracitados, a super-resolução exposta neste documento pode ser útil principalmente como uma ferramenta na recuperação e estimação de imagens em quaisquer cenários onde as sequências de vídeo possuam quadros de resolução mista.

5.2 TRABALHOS FUTUROS

Alguns aspectos deste projeto podem ser desenvolvidos e explorados mais a fundo certamente provendo espaço para trabalhos futuros nessa área. Levando em consideração que dependendo do cenário onde se deseje aplicar essa técnica de super-resolução, diversas melhorias podem ser realizadas sendo direcionadas aos problemas e limitações de cada aplicação. Portanto, aqui serão indicados somente alguns pontos interessantes a serem explorados.

- Estudo das melhores combinações de macroblocos. Embora neste projeto não se tenha alcançado um melhor resultado utilizando a informação da SSD como descrito na Subseção 3.3.3, talvez ainda seja possível convergir em uma função de custo que possa indicar uma decisão mais adequada quanto à adição da alta-frequência e suas ponderações. É possível ainda analisar a existência de alguma outra métrica mais apropriada para avaliar as melhores combinações principalmente levando em consideração que a estimação de movimento é feita somente em parte das componentes dos quadros (baixas- e médias-frequências).
- Análise da viabilidade da implementação de tamanhos de blocos adaptativos. Embora esta idéia tenha sido cogitada (apesar de não estar descrita neste manuscrito) um fator que deve ser contornado é como decidir sobre o melhor tamanho de bloco. Para isso, é preciso antes escolher um critério de avaliação eficiente do que seria o melhor casamento entre blocos, recaindo no item descrito acima. Definido o critério de comparação, seria mais razoável tentar estimar os melhores blocos utilizando diferentes tamanhos.
- Usar técnicas de super-resolução Bayesianas junto ao método proposto. Aqui é possível tentar dois caminhos: o primeiro deles é super-resolver um quadro com duas técnicas

em paralelo e depois fundir os resultados, a segunda alternativa seria super-resolver um quadro com a técnica aqui proposta e depois tentar super-resolvê-lo novamente utilizando as técnicas Bayesianas.

- Utilizar quadros já super-resolvidos para super-resolver outros. A princípio esta idéia não foi levada a frente dado que toda a alta-frequência adicionada aos quadros não-chave provinha dos quadros chave. Essa concepção pode mudar um tanto a partir do momento que caso o item anterior seja válido e se faça possível fundir resultados, as altas-frequências dos quadros não-chave deixam de ser somente combinações lineares das componentens presentes nos quadro chave. Dessa forma, a alta-frequência dos quadros super resolvidos pode ser aproveitada para super-resolver os outros quadros não-chave.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ITU-T. *ITU-T Recommendation H.263, Video coding for low bit rate communication*. [S.l.], March 1996.
- [2] ITU-T. *ITU-T Recommendation H.264, Advanced Video Coding for Generic Audiovisual Services*. [S.l.], May 2003.
- [3] ISO/IEC. *ISO/IEC 13818 - Generic Coding of Moving Pictures and Associated Audio Information*. [S.l.], 1995.
- [4] ISO/IEC. *ISO/IEC 14496-2, Coding of Audio-Visual Objects - Part 2: Visual*. [S.l.], 2001.
- [5] AARON S. RANE, E. S. B. G. A. Transform-domain wynerziv codec for video. *SPIE Visual Communications and Image Processing Conference*, v. 5308, p. 520–528, 2004.
- [6] RAO, K. R.; HWANG, J. J. *Techniques and Standards for Image, Video and Audio Coding*. [S.l.]: Prentice Hall PTR, 1996.
- [7] GIBSON, J. D. et al. *Digital Compression for Multimedia: Principles and Standards*. [S.l.]: Morgan Kaufmann Publishers, 2006.
- [8] RICHARDSON, I. E. G. *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*. [S.l.]: Wiley, 2003.
- [9] GIROD, B. et al. Distributed video coding. In: *Proc. IEEE Special Issue on Advances in Video Coding and Delivery*. [S.l.: s.n.], 2005. p. 1–12.
- [10] PURI, R.; RAMCHANDRAN, K. PRISM: A “reversed” multimedia coding paradigm. *Proc. IEEE International Conference on Image Processing*, v. 1, p. I-617–20, Setembro 2003.
- [11] MUKHERJEE, D.; MACCHIAVELLO, B.; QUEIROZ, R. L. A simple reversed-complexity Wyner-Ziv video coding mode based on a spatial reduction framework. *Proc. of SPIE Visual Communications and Image Processing*, v. 6508, p. 65081Y1–65081Y12, Janeiro 2007.

- [12] RICHARDSON, I. E. G. *Video Codec Design: Developing Image and Video Compression Systems*. [S.l.]: Wiley, 2002.
- [13] JAIN, J. R.; JAIN, A. K. Displacement measurement and its application in interframe image coding. *IEEE Transactions on Communications*, COM-29, p. 1799–1808, December 1981.
- [14] LI, R.; ZENG, B.; LIOU, M. L. A new three-step search algorithm for block motion estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 4, p. 438–442, August 1994.
- [15] PO, L. M.; MA, W. C. A novel four-step search algorithm for fast block motion estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 6, p. 313–317, June 1996.
- [16] DINIZ E. A. B. DA SILVA, S. L. N. P. S. R. *Processamento digital de sinais: projeto e análise de sistemas*. Porto Alegre: Bookman, 2004. 590p. p. ISBN 85-363-0418-9.
- [17] OPPENHEIM R.W. SCHAFFER, J. B. A. *Discrete-Time Signal Processing*. 2.ed. ed. Upper Saddle River, NJ, EUA: Prentice-Hall, 1997. ISBN 0-13-814757-4.
- [18] LATHI, B. P. *Modern Digital and Analog Communication Systems*. 3.ed. ed. New York, NY, USA: Oxford University Press, Inc., 1998. ISBN 0195110099.
- [19] MACCHIAVELLO, B.; MUKHERJEE, D.; QUEIROZ, R. de. A statistical model for a mixed resolution Wyner-Ziv framework. *Picture Coding Symposium*, Lisboa, 2007.
- [20] KATSAGGELOS, A. K.; MOLINA, R.; MATEOS, J. *Super Resolution of Images and Video*. [S.l.]: Morgan and Claypool Publishers, 2007. (Synthesis Lectures on Image, Video, and Multimedia Processing). ISBN 978-1598290844.
- [21] PARK, M. K. P. S. C.; KANG, M. G. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.*, v. 20, p. 21–36, 2003.
- [22] TSAI, R. Y.; HUANG, T. S. Multiframe image restoration and registration. *Adv. Comput. Vis. Image Process*, v. 1, p. 317–339, 1984.
- [23] ROBERTS, R. A.; MULLIS, C. T. *Digital Signal Processing*. [S.l.]: Addison-Wesley, 1987.

- [24] PAPOULIS, A. Generalized sampling expansion. *IEEE Trans. Circuits Syst. Video Technol.*, v. 24, p. 652–654, 1977.
- [25] IRANI, M.; PELEG, S. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *J. Vis. Commun. Image Represent.*, v. 4, p. 324–335, 1993.
- [26] ELAD, M.; FEUER, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.*, v. 6, p. 1646–1658, 1997.
- [27] LERTRATTANAPANICH, S.; BOSE, N. K. High resolution image formation from low resolution frames using delaunay triangulation. *IEEE Trans. Image Process.*, v. 11, p. 1427–1441, 2002.
- [28] SEGALL R. MOLINA, A. K. K. C. A.; MATEOS, J. Bayesian resolution enhancement of compressed video. *IEEE Trans. Image Process.*, v. 13, p. 898–911, 2004.
- [29] BRUCKSTEIN, M. E. A.; KIMMEL, R. Down scaling for better transform compression. *IEEE Trans. Image Process.*, v. 12, p. 1132–1144, 2003.
- [30] MOLINA A. K. KATSAGGELOS, L. D. A. R.; MATEOS, J. Towards a new video compression scheme using super-resolution. *Proc. SPIE Conf. Vis. Commun. Image Process.*, v. 6077, p. 1–13, 2006.
- [31] FREEMAN, T. R. J. W. T.; PASZTOR, E. C. Example based super-resolution. *IEEE Comput. Graphics Appl.*, v. 22, p. 56–65, 2002.
- [32] MUKHERJEE, D. A robust reversed complexity Wyner-Ziv video codec introducing sign-modulated codes. *HP Labs Technical Report, HPL-2006-80*, Maio 2006.
- [33] MACCHIAVELLO, B.; MUKHERJEE, D.; QUEIROZ, R. L. Iterative side-information generation in a mixed resolution Wyner-Ziv framework. *IEEE Trans. on Circuits and Systems for Video Technology*, 2008.
- [34] MACCHIAVELLO, B. et al. Super-resolution applied to distributed video coding with spatial scalability. *Simpósio Brasileiro de Telecomunicações*, 2008.

- [35] KEYS, R. G. Cubic convolution interpolation for digital image processing. *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*, ASSP-2, n. 6, p. 1153–1160, 1981.
- [36] THE MathWorks. <http://www.mathworks.com/>, último acesso em 22/12/2008 2008.
- [37] WIEGAND, T. et al. Overview of the h.264/avc video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 13, n. 7, p. 560–576, July 2003.
- [38] HUNT, R. W. G. *The Reproduction of Colour*. [S.l.]: John Wiley and Sons, 2004. (Series in Imaging Science and Technology).
- [39] BRANDI, F.; QUEIROZ, R. L. de; MUKHERJEE, D. Super resolution of video using key frames. *Proc. of IEEE International Symposium on Circuits and Systems*, Maio 2008.
- [40] BRANDI, F.; QUEIROZ, R. L. de; MUKHERJEE, D. Super resolution of video using key frames and motion estimation. *Proc. of International Conference on Image Processing*, 2008.
- [41] MACCHIAVELLO, B. et al. Side-information generation for temporal and spatial scalable wyner-ziv codecs. *A ser publicado no EURASIP Journal on Image and Video Processing*, 2009.