



# FUSÃO DE VÍDEOS UTILIZANDO DECOMPOSIÇÕES MULTIRRESOLUÇÃO ESPAÇO-TEMPORAIS

Jonathan Nogueira Gois

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Eduardo Antônio Barros da  
Silva  
Carla Liberal Pagliari

Rio de Janeiro  
Setembro de 2016

FUSÃO DE VÍDEOS UTILIZANDO DECOMPOSIÇÕES MULTIRRESOLUÇÃO  
ESPAÇO-TEMPORAIS

Jonathan Nogueira Gois

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO  
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE  
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE  
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A  
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA  
ELÉTRICA.

Examinada por:

---

Prof. Eduardo Antônio Barros da Silva, Ph.D.

---

Prof. Carla Liberal Pagliari, Ph.D.

---

Prof. Sergio Lima Netto, Ph.D.

---

Prof. Lisandro Lovisolo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL  
SETEMBRO DE 2016

Gois, Jonathan Nogueira

Fusão de vídeos utilizando decomposições multirresolução espaço-temporais/Jonathan Nogueira Gois. – Rio de Janeiro: UFRJ/COPPE, 2016.

XX, 113 p.: il.; 29,7cm.

Orientadores: Eduardo Antônio Barros da Silva  
Carla Liberal Pagliari

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2016.

Referências Bibliográficas: p. 86 – 93.

1. Fusão de Vídeos. 2. Transformadas multirresolução espaço-temporais. 3. Filtragem temporal usando compensação de movimento. I. Silva, Eduardo Antônio Barros da *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*A Deus e à minha família.*

# Agradecimentos

Em primeiro lugar, gostaria de agradecer a Deus pelo fim de mais esta jornada. Também sou imensamente grato por minha família, pelo apoio dos meus pais e de minha irmã durante toda a graduação e mestrado. Agradeço à minha namorada, Dayane, pelo tempo e pela compreensão nos momentos difíceis.

Não menos importante, agradeço a quem dividiu de perto esta jornada comigo. Em especial aos amigos do laboratório: Felipe “Curicica”, Markus Lima, Tadeu Ferreira, Anderson “Tchê”, Pedro Donadio, Renam Castro, Gabriel “Baiano”, Luiz “Maestro”, José Fernando e todos os que minha memória não permitiu anotar o nome. Também demonstro minha gratidão aos amigos que me ajudaram no momento de lazer, nos momentos em que precisei de um ombro amigo, em especial à Camilla de Paula, Thiago Pedroza, Beatriz de Ávila, Renan Lúcio, Bruna Oliveira, Fabiano Martins e Almiro Bispo.

Por fim, gostaria de agradecer aos meus orientadores, Eduardo e Carla. Que tecnicamente e pessoalmente me ajudaram durante todos os 6 anos de orientação. Agradeço por terem aceitado mais este desafio em me orientar.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## FUSÃO DE VÍDEOS UTILIZANDO DECOMPOSIÇÕES MULTIRRESOLUÇÃO ESPAÇO-TEMPORAIS

Jonathan Nogueira Gois

Setembro/2016

Orientadores: Eduardo Antônio Barros da Silva  
Carla Liberal Pagliari

Programa: Engenharia Elétrica

A diversidade de sensores de imageamento digital aumentou muito nas últimas décadas. A informação presente em cada um destes novos sensores pode ser importante em diferentes áreas, tais como a área militar, a médica, entre outras. Para o usuário, a análise conjunta das informações oriundas de diferentes sensores pode ser uma tarefa difícil. Desta forma, os métodos de fusão de imagens/vídeos promovem a combinação das diferentes informações em apenas uma única saída. Esta saída deve ser mais significativa do que cada informação isolada, facilitando a análise das informações combinadas para o usuário.

O avanço da tecnologia permitiu a fusão de vídeos, e não somente imagens, de uma mesma cena adquiridos com diferentes sensores. A fusão de vídeos é pouco abordada na literatura, enquanto que a fusão de imagens de cada quadro do vídeo é uma técnica comumente utilizada. Entretanto, esta abordagem desconsidera a informação temporal. Nesta dissertação, a informação temporal é considerada empregando a informação do movimento dos objetos em cena juntamente com decomposições multirresolução espaciais, já utilizadas na literatura de fusão de imagens. A fusão espaço-temporal de vídeos tem por objetivo preservar e salientar tanto a informação temporal quanto a espacial.

Este trabalho apresenta os métodos desenvolvidos de fusão de vídeos, oriundos das faixas espectrais da luz visível e do infravermelho, empregando a informação temporal. Os resultados obtidos indicam que, ao inserir a informação do movimento dos objetos em cena, a qualidade objetiva do vídeo fundido aumenta. Além disso, os métodos propostos diminuem o efeito de erros de registro das imagens de entrada.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

## VIDEO FUSION USING SPATIOTEMPORAL MULTIREOLUTION DECOMPOSITIONS

Jonathan Nogueira Gois

September/2016

Advisors: Eduardo Antônio Barros da Silva

Carla Liberal Pagliari

Department: Electrical Engineering

The diversity of digital imaging sensors has greatly increased in recent decades. The information conveyed by these sensors can be important in different areas, such as military, medical, among others. The joint analysis of information from different sensors can be a difficult task. Thus, the methods of fusing images / videos promote the combination of different information in only a single output. This output should produce a better result than each isolated information, facilitating the analysis of the combined information for the user.

The advancement of technology has allowed, not only the fusion of images of the same scene acquired with different sensors, but also the fusion of image sequences (video). The fusion of videos is rarely addressed in the literature, while the fusion of images of each video frame is a commonly used technique. However, this approach ignores the temporal information. In this dissertation, the temporal information is considered using the motion information of the objects in the scene along with multi-resolution spatial decomposition, already used in image fusion literature. The spatio-temporal video fusion aims to preserve and highlight both temporal and spatial information.

This work presents the methods developed for video fusion, originated from the spectral ranges of visible light and infrared, assisted by the temporal information. The results indicate that when the motion information is considered, the objective quality of the fused video increases. In addition, the proposed methods reduce the registration errors of the input images.

# Sumário

Lista de Figuras	xii
Lista de Tabelas	xiv
Lista de Símbolos	xviii
Lista de Abreviaturas	xix
<b>1 Introdução</b>	<b>1</b>
1.1 Organização . . . . .	4
<b>2 Métricas Objetivas de Avaliação de Vídeos Fundidos</b>	<b>5</b>
2.1 Métricas Espaciais . . . . .	6
2.1.1 Informação Mútua . . . . .	7
2.1.2 Métrica de Xydeas e Petrović . . . . .	8
2.1.3 Medida de artefatos de fusão . . . . .	9
2.1.4 Medida de Informação Perdida . . . . .	10
2.1.5 Métrica de Piella . . . . .	11
2.2 Métricas Espaço-temporais . . . . .	12
2.2.1 Informação Mútua de diferença entre quadros . . . . .	14
2.2.2 Avaliação de Qualidade Dinâmica . . . . .	14
2.3 Validação . . . . .	16
2.4 Conclusões . . . . .	18
<b>3 Fusão Multirresolução de Imagens</b>	<b>20</b>
3.1 Técnicas de Fusão de Imagens Multirresolução . . . . .	21
3.1.1 Transformada <i>Wavelet</i> Discreta . . . . .	23
3.1.2 Transformada <i>Contourlet</i> . . . . .	25
3.1.3 Transformada <i>Wavelet</i> não-decimada . . . . .	27
3.2 Regras de Fusão . . . . .	30
3.3 Banco de Dados de Vídeos para Fusão . . . . .	31
3.3.1 Descrição do conjunto de teste . . . . .	33

3.4	Comparação do Desempenho dos Algoritmos de Fusão Espacial . . . . .	34
3.4.1	Métodos comparativos . . . . .	34
3.4.2	Transformada <i>Wavelet</i> Discreta . . . . .	35
3.4.3	Transformada <i>Contourlet</i> . . . . .	38
3.4.4	Transformada <i>Wavelet</i> não-decimada com Fatoração Espectral . . . . .	40
3.5	Conclusões . . . . .	43
<b>4</b>	<b>Decomposição em Subbandas Temporais ao Longo da Trajetória do Movimento</b>	<b>45</b>
4.1	Trabalhos relacionados à fusão de vídeo . . . . .	45
4.2	Filtragem Temporal usando Compensação de Movimento . . . . .	46
4.3	O Esquema <i>Lifting</i> . . . . .	49
4.4	Esquema <i>Lifting</i> para a Fatoração Espectral . . . . .	53
4.5	Esquema <i>Lifting</i> Haar usando Compensação de Movimento . . . . .	54
4.6	Esquema <i>Lifting</i> Le Gall 5-3 usando Compensação de Movimento . . . . .	55
4.7	Esquema <i>Lifting</i> Cohen-Daubechies-Feauveau 9-7 usando Compensação de Movimento . . . . .	57
4.8	Considerações Finais . . . . .	57
<b>5</b>	<b>Algoritmos de Fusão Espaço-temporal</b>	<b>59</b>
5.1	Decomposição Tridimensional . . . . .	59
5.2	Algoritmo de Fusão usando Esquema <i>Lifting</i> e Compensação de movimento . . . . .	62
5.2.1	Decomposição em Sub-bandas Temporais . . . . .	62
5.2.2	Decomposição em Sub-bandas Espaço-Temporais . . . . .	70
5.3	Transformada <i>Wavelet</i> com Movimento Compensado . . . . .	71
5.3.1	Transformada <i>Wavelet</i> Temporal sem Compensação de Movimento . . . . .	74
5.4	Análise dos resultados . . . . .	75
5.5	Considerações Finais . . . . .	82
<b>6</b>	<b>Conclusão</b>	<b>83</b>
6.1	Trabalhos Futuros . . . . .	84
	<b>Referências Bibliográficas</b>	<b>86</b>
<b>A</b>	<b>Validação das métricas espaço-temporais</b>	<b>94</b>
A.1	Avaliação Espacial . . . . .	95
A.2	Avaliação Temporal . . . . .	96
A.3	Avaliação Espaço-temporal . . . . .	97

<b>B</b>	<b>Resultados de Avaliação Objetiva dos Vídeos Fundidos</b>	<b>98</b>
B.1	Decomposição em Sub-bandas Temporais . . . . .	98
B.2	Decomposição em Sub-bandas Espaço-temporais . . . . .	101
B.3	Extensão Temporal dos Coeficientes Multirresolução . . . . .	103
<b>C</b>	<b>Amostras Fundidas</b>	<b>108</b>
C.1	<i>Camouflage</i> . . . . .	109
C.2	<i>Guanabara Bay</i> . . . . .	110
C.3	<i>Trees</i> . . . . .	111
C.4	<i>Hangar</i> . . . . .	112
C.5	<i>Lab</i> . . . . .	113

# Lista de Figuras

1.1	Diagrama de fusão de imagens de dois sensores diferentes. A mesma cena pode ser conter diferentes informações para cada sensor. . . . .	1
1.2	As imagens (a) e (b) são provenientes de sensores da região da luz visível e do infravermelho respectivamente; (c) e (d) são exemplos de resultados médicos; Os diferentes usos do mesmo sensor são exemplificadas em (e) e (f); As últimas imagens, (g) e (h), mostram uma imagem com baixa resolução colorida e uma imagem de alta resolução em tons de cinza. . . . .	2
1.3	Comparação entre as informações da região do espectro de luz visível e do infravermelho. Na imagem da esquerda, detalhes da face, das roupas e da estrutura da imagem são notados. Na imagem da direita, é possível perceber os contornos dos objetos de maneira a revelar dois objetos escondidos. Estas imagens são parte da sequência <i>Lab Take 1</i> presentes no banco de vídeos VLIRVDIF. . . . .	3
2.1	Contribuição das informações de $I_A$ e $I_B$ em $I_F$ . . . . .	9
2.2	Artefatos de fusão presentes em $I_F$ . . . . .	10
2.3	Informação perdida devido a fusão. . . . .	10
2.4	Comparação entre as imagens fundidas usando entradas registradas ou com erro de registro. Foi imposto, artificialmente, um erro de registro de 10 pixels. É possível observar que bordas duplas são mais presentes em (b). . . . .	13
2.5	Operador Sobel estendido para a dimensão temporal. . . . .	15
3.1	Comparação da resolução espaço-frequencial entre a STFT e DWT. . . . .	21
3.2	Diagrama para exemplificar os passos de um dado algoritmo de fusão. . . . .	22
3.3	Comparação entre as descrições em bases <i>wavelet</i> e <i>contourlet</i> de imagens naturais. . . . .	25

3.4	Esquema <i>Contourlet</i> : O processo de filtragem dupla é composta por a filtragem Laplaciana ( <i>Laplacian Pyramid</i> ) seguida pela filtragem direcional ( <i>Directional Filter Bank</i> ). Outro nível de filtragem é mostrado após o decimador (2, 2) que representa a subamostragem das linhas e colunas pelo fator 2. . . . .	26
3.5	Esquema LP para um nível de decomposição e reconstrução. . . . .	27
3.6	Partição frequencial do DFB com $l = 3$ , com isso, tem-se $2^3 = 8$ sub-bandas frequências. . . . .	27
3.7	UWT-SF usando dois níveis de decomposição. Após aplicar o primeiro nível de decomposição, os detalhes da imagem fundida são extraídos da fatoração espectral e da regra de fusão <b>F</b> . Esta representação denota uma decomposição unidimensional, porém a decomposição n-dimensional pode ser obtida através da aplicação desse esquema em cada dimensão. . . . .	29
3.8	<i>Thumbnail</i> do banco de vídeos. . . . .	33
4.1	Predição do movimento através dos vetores de movimento. . . . .	47
4.2	Ilustração da compensação de movimento de um pixel $(m, n)$ . . . . .	48
4.3	Compensação de movimento <i>backward</i> . . . . .	49
4.4	Esquema <i>Lifting</i> Clássico. . . . .	51
4.5	<i>Lazy Wavelet</i> . . . . .	51
4.6	Inversa do esquema <i>Lifting</i> . . . . .	53
4.7	Estrutura <i>Lifting</i> não-decimada usando os passos provenientes da estrutura decimada. . . . .	54
4.8	Esquemas <i>lifting</i> usando compensação de movimento e o banco de filtros Haar. Neste caso, o filtro de predição $P(z)$ é igual a $1/\sqrt{2}$ e o filtro de atualização $U(z)$ é igual a $\sqrt{2}$ . A esquerda está o esquema responsável pela decomposição. A direita, o esquema que realiza a síntese das sub-bandas. . . . .	54
5.1	Exemplos de quadros fundidos usando a técnica 3D-DWT. . . . .	62
5.2	Exemplos de quadros fundidos usando a técnica 3D-UWT. . . . .	62
5.3	Diagrama de fusão usando Decomposição Temporal em Sub-bandas. O primeiro passo é a decomposição temporal usando MCTF. Após a decomposição, as sub-bandas são combinadas através da regra de fusão temporal. . . . .	64
5.4	Diagrama de fusão usando decomposição em sub-bandas espaço-temporais. Nesta abordagem, primeiramente é aplicada a decomposição em sub-bandas temporais. Após isto, cada sub-banda temporal é fundida através de alguma técnica de fusão de imagens. . . . .	70

5.5	Diagrama da Transformada <i>Wavelet</i> com Movimento Compensado usando um quadro passado e um futuro. . . . .	72
5.6	Cena <i>Camouflage</i> :quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda). Pode-se notar diferentes planos na cena. . .	77
5.7	Cena <i>Camouflage</i> : amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)). . . . .	77
5.8	Cena <i>Guanabara Bay</i> : quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda). . . . .	78
5.9	Cena <i>Guanabara Bay</i> : amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)). . . . .	78
5.10	Cena <i>Trees</i> : quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda). . . . .	79
5.11	Cena <i>Trees</i> : amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)). . .	79
5.12	Cena <i>Hangar</i> : quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda). . . . .	80
5.13	Cena <i>Hangar</i> : amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)).	80
5.14	Cena <i>Lab</i> : quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda). . . . .	80
5.15	Cena <i>Lab</i> : amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)). . .	81
5.16	Histograma comparativo do valor médio do quadro avaliado ao longo da cena <i>Camouflage</i> . Na imagem da esquerda, é mostrado um histograma do melhor resultado segundo a métrica MI-IFD do Capítulo 3. Na direita, o melhor resultado mostrado neste Capítulo. . . . .	81

# Lista de Tabelas

2.1	Correspondência subjetiva com as métricas de avaliação objetivas de fusão de imagens mostradas. . . . .	17
2.2	Resultados da validação das métricas espaciais incluindo a validação AUC. . . . .	18
2.3	Inclinação das curvas MOS $\times$ valor predito pelas métricas objetivas. Esta inclinação foi alcançada através de um ajuste linear. Os pares de valores foram normalizados a fim de que a maior correlação seja expressa no coeficiente angular igual a 1. . . . .	18
3.1	Descrição da parte do banco usado neste trabalho. . . . .	33
3.2	Resultado da fusão usando média. . . . .	35
3.3	Principal component analysis fusion results. . . . .	36
3.4	Resultados da cena <i>Camouflage</i> usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	37
3.5	Resultados da cena <i>Guanabara Bay</i> usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	37
3.6	Resultados da cena <i>Trees</i> usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	37
3.7	Resultados da cena <i>Hangar</i> usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	38
3.8	Resultados da cena <i>Lab</i> usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	38
3.9	Resultados da cena <i>Camouflage</i> usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	39
3.10	Resultados da cena <i>Guanabara Bay</i> usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	39
3.11	Resultados da cena <i>Trees</i> usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	39
3.12	Resultados da cena <i>Hangar</i> usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	40

3.13	Resultados da cena <i>Lab</i> usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	40
3.14	Resultados da cena <i>Camouflage</i> usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	40
3.15	Resultados da cena <i>Guanabara Bay</i> usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	41
3.16	Resultados da cena <i>Trees</i> usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	41
3.17	Resultados da cena <i>Hangar</i> usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	41
3.18	Resultados da cena <i>Lab</i> usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica. . . . .	42
3.19	Resultado comparativo final dos métodos multirresolução utilizadas neste Capítulo. . . . .	43
5.1	Comparação entre a DWT-2D e a DWT-3D, ou seja, usando a dimensão temporal na decomposição ou não. . . . .	60
5.2	Comparação da UWT-SF aplicada a cada quadro com a UWT-SF tridimensional (usando o banco de filtros $B_3$ -spline e 4 níveis de decomposição). . . . .	61
5.3	Regras de fusão temporal que serão usadas nos experimentos. . . . .	65
5.4	Fusão da cena <i>Camouflage</i> usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 8 e acurácia unitária. . . . .	67
5.5	Fusão da cena <i>Camouflage</i> usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 4 e acurácia unitária. . . . .	68
5.6	Fusão da cena <i>Camouflage</i> usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 4 e acurácia de 1/2 pixel. . . . .	68
5.7	Fusão da cena <i>Camouflage</i> usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 4 e acurácia de 1/4 pixel. . . . .	69
5.8	Cenas fundidas usando decomposição em sub-bandas temporais aplicando os filtros 5-3 e 9-7. Em todas as cenas, os vetores de movimento foram calculados usando tamanho de bloco igual a 4 e acurácia unitária. . . . .	69

5.9	Nesta abordagem, foi usado a decomposição em sub-bandas temporais e então a UWT-SF para a decomposição em sub-bandas espaciais. Na primeira etapa, foi usado o banco de filtros Haar, por obter os melhores resultados na última Seção. Na etapa de compensação de movimento, foi escolhido tamanho de bloco igual a 4 e 1/2 pixel de acurácia. A regra temporal escolhida foi a <i>meanmean</i> . . . . .	71
5.10	Melhores resultados usando a MCWT. Neste experimento, foi usado tamanhos de bloco iguais a 4,8 e 16; acurácia unitária e 2 níveis de decomposição. . . . .	73
5.11	Melhores resultados usando MCWT e UWT. Neste experimento, foi usado tamanhos de bloco iguais a 4,8 e 16; acurácia unitária e 4 níveis de decomposição. . . . .	73
5.12	Melhores resultados usando a Transformada <i>Wavelet</i> Temporal sem Compensação de Movimento e DWT. . . . .	74
5.13	Melhores resultados usando a Transformada <i>Wavelet</i> Temporal sem Compensação de Movimento e UWT. . . . .	74
5.14	Melhores resultados usando as técnicas espaço-temporais propostas. Na descrição dos métodos, a abreviação <i>pel</i> indica a acurácia da compensação de movimento. . . . .	75
5.15	Resutaldo comparativo entre os métodos espaciais e os métodos espaço-temporais. . . . .	76
B.1	Fusão da cena <i>Guanabara Bay</i> computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel. . . . .	98
B.2	Fusão da cena <i>Trees</i> computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.	99
B.3	Fusão da cena <i>Hangar</i> computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.	99
B.4	Fusão da cena <i>Lab</i> computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.	100
B.5	Fusão da cena <i>Guanabara Bay</i> computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel.	101
B.6	Fusão da cena <i>Hangar</i> computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel. . . . .	101

B.7	Fusão da cena <i>Trees</i> computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel. . . . .	102
B.8	Fusão da cena <i>Lab</i> computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel. . . . .	102
B.9	Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária. . . . .	103
B.10	Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária. . . . .	104
B.11	Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária. . . . .	105
B.12	Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária. . . . .	106
B.13	Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária. . . . .	107

# Lista de Símbolos

$N^{AB/F}$	Medida de Artefato de Fusão, p. 9
$M_{AB/F}$	Informação Mútua, p. 7
$Q_p$	Métrica de Piella, p. 12
$Q_{AB/F}$	Métrica Xydeas e Petrović, p. 8
$Q_{ST-PC}$	Índice de Congruência de Fase Espaço-temporal, p. 6
$Q_{ST}$	Índice de Similaridade Estrutural Espaço-temporal , p. 6

# Lista de Abreviaturas

1D	Unidimensional, p. 28
3D-DWT	<i>Tridimensional Discrete Wavelet Transform</i> , p. 59
3D-UWT	<i>Tridimensional Undecimated Wavelet Transform</i> , p. 59
3D	Tridimensional, p. 46
AUC	<i>Area under the ROC Curve</i> , p. 17
CVT	<i>Curvelet Transform</i> , p. 20
ConT	<i>Contourlet Transform</i> , p. 25
DFB	<i>Dirertional Filter Bank</i> , p. 25
DFT	<i>Discrete Fourier Transform</i> , p. 20
DQ	<i>Dynamic Quality Evaluation</i> , p. 14
DWT	<i>Discrete Wavelet Transform</i> , p. 20
FIR	<i>Finite Impulse Response</i> , p. 50
IFD	<i>Interframe Difference</i> , p. 14
LP	<i>Laplacian pyramid</i> , p. 25
LS	<i>Lifting Scheme</i> , p. 50
MCTF	<i>Motion Compensated Temporal Filtering</i> , p. 47
MCWT	<i>Motion Compensated Wavelet Transform</i> , p. 72
MI-IFD	<i>Mutual information from interframe differences</i> , p. 14
MOS	<i>mean opinion score</i> , p. 18
MVR	<i>Motion Vector Rule</i> , p. 64

NSCT	<i>Nonsubsampled Contourlet Transform</i> , p. 26
PCA	<i>Principal Component Analysis</i> , p. 31, 34
ROC	<i>Receiver Operating Characteristic</i> , p. 17
SF	<i>Spectral Factorization</i> , p. 29
SSIM	<i>Structural Similarity index</i> , p. 6
STFT	<i>Short-Time Fourier Transform</i> , p. 20
SVC	<i>Scalable Video Coding</i> , p. 47
SVH	<i>Sistema Visual Humano</i> , p. 6
UWT-SF	<i>Undecimated Wavelet Transform with Spectral Factorization</i> , p. 29
UWT	<i>Undecimated Wavelet Transform</i> , p. 20
VLIRVDIF	<i>Visible-Light and Infrared Video Database for Image Fusion</i> , p. 4

# Capítulo 1

## Introdução

A área de imageamento digital vem crescendo rapidamente nas últimas duas décadas, por conta da grande disponibilidade de sensores [1]. A diversidade de tecnologias permite vários benefícios em vários campos tais como as áreas médicas, militar, de sensoriamento remoto, entre outras. Concomitante ao crescimento da disponibilidade de diferentes sensores, a disponibilidade de armazenamento para tais informações também cresceu. Isto permite que, além da informação espacial disponível na imagem, a informação temporal seja contemplada. Com isso, pode-se armazenar várias imagens que registram diferentes momentos da cena que possam conter distintas informações.

Dada a variedade de informações, a visualização pode ser prejudicada pelo excesso. A solução para este problema pode ser obtida usando técnicas de fusão de imagens/vídeos [2, 3]. Este processo combina todas as entradas em apenas uma saída mais representativa. Isto é, a imagem final contém mais informações do que cada entrada individualmente (Figura 1.1).

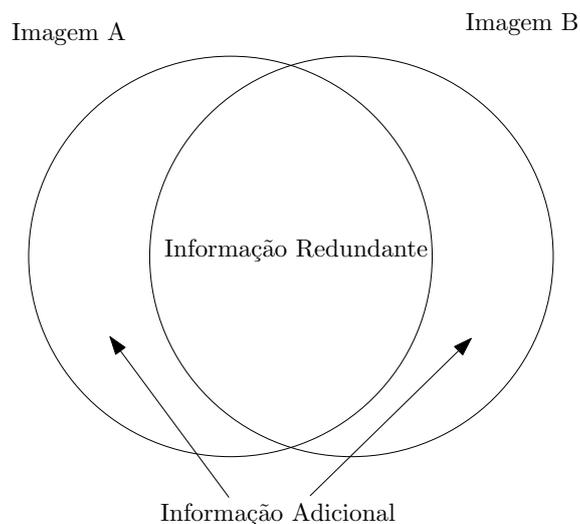


Figura 1.1: Diagrama de fusão de imagens de dois sensores diferentes. A mesma cena pode conter diferentes informações para cada sensor.

Essa informação adicional é dependente dos sensores escolhidos. Por exemplo, nas Figuras 1.2(a) e 1.2(b) são mostradas, respectivamente, imagens de sensores sensíveis à região espectral do infravermelho e da luz visível, muito úteis na área militar. Já as Figuras 1.2(c) e 1.2(d) dão exemplo de tipos de imagens úteis na área médica tais como uma ressonância magnética e um raio-x. Imagens da mesma cena representadas em diferentes planos focais podem ser vistas nas Figuras 1.2(e) e 1.2(f). Por fim, algumas imagens do banco de dados Quickbird [4] são exemplificadas nas Figuras 1.2(g) e 1.2(h).

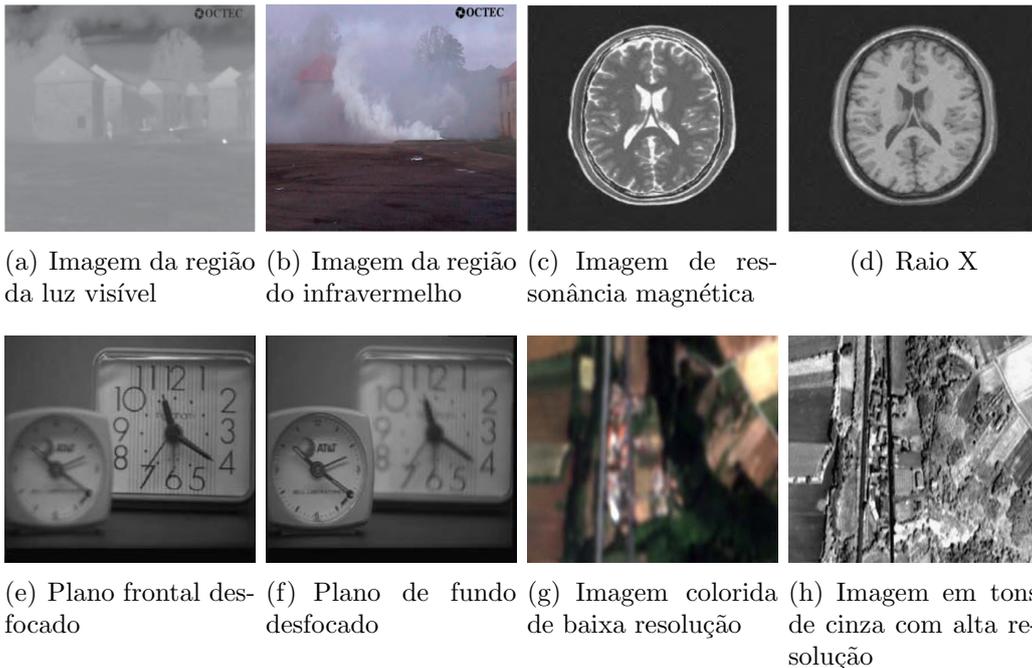


Figura 1.2: As imagens (a) e (b) são provenientes de sensores da região da luz visível e do infravermelho respectivamente; (c) e (d) são exemplos de resultados médicos; Os diferentes usos do mesmo sensor são exemplificadas em (e) e (f); As últimas imagens, (g) e (h), mostram uma imagem com baixa resolução colorida e uma imagem de alta resolução em tons de cinza.

A variedade de informação presente em cada imagem pode ser ilustrada usando o par de imagens da Figura 1.3 [5]. A imagem que contém as informações provenientes da região do infravermelho mostra o contorno dos objetos e, ainda, alguns objetos “escondidos”, tais como facas e outras armas. Enquanto a outra imagem, que nos dá informação da região de luz visível, nos mostra os detalhes da cena, tais como as aparências das pessoas e dos objetos.

Depois de algumas décadas, muitas técnicas de fusão de imagens foram desenvolvidas [6–12]. Geralmente, o processo de fusão pode ser categorizado em três níveis: do pixel, das características e dos níveis de decisão [13]. Estas categorias ainda podem ser reduzidas a duas: a nível do pixel-caraterística e a nível de decisão [14]. O nível pixel-característica realiza a fusão usando pixels, blocos de pixels, imagens



(a) Imagem da região da luz visível      (b) Imagem da região do infravermelho

Figura 1.3: Comparação entre as informações da região do espectro de luz visível e do infravermelho. Na imagem da esquerda, detalhes da face, das roupas e da estrutura da imagem são notados. Na imagem da direita, é possível perceber os contornos dos objetos de maneira a revelar dois objetos escondidos. Estas imagens são parte da sequência *Lab Take 1* presentes no banco de vídeos VLIRVDIF.

segmentadas, etc. Por outro lado, o nível de decisão combina as entradas usando descritores da imagem [13].

Como objetivo comum a todas as técnicas de fusão de imagens, está sempre o realce da informação espacial. Ao propor técnicas de fusão de vídeo, tem-se interesse também no realce da informação temporal. Por isso, pode-se também categorizar as técnicas de fusão de vídeo em dois aspectos [15]: fusão baseada em um único quadro e fusão espaço-temporal. Enquanto a primeira usa as técnicas de fusão de imagem, a fusão espaço-temporal inclui o tempo como entrada.

Embora as técnicas de fusão da imagem se mostrem populares na literatura, poucos são os trabalhos que contemplam a fusão de vídeos atualmente. Por essa razão, este trabalho tem como foco técnicas de fusão de vídeos que preservem o movimento dos objetos em cena.

Há diversas áreas da ciência onde as informações temporal e espacial devem ser analisadas conjuntamente, ou seja, devem ser preservadas na etapa de fusão. Por exemplo, na área militar, a detecção de alvo é um desafio. Como objetivo, é tão importante localizar a trajetória do alvo ao longo do tempo (informação temporal) quanto os detalhes do alvo (informação espacial) [16]. Usando os vídeos da região do infravermelho e da luz visível, podemos extrair detalhes importantes. O infravermelho traz a informação dos contornos dos objetos em cena, facilitando a localização mesmo em ambientes com fumaça, neblina, iluminação deficiente, excesso de luz, etc. Por outro lado, o vídeo com informações da região da luz visível nos dá os detalhes dos objetos tais como feições, texturas, cores, etc.

O mesmo problema acontece na área de sensoriamento remoto [17]: um vídeo infravermelho ou com visão noturna permite realçar áreas quentes mesmo detrás de fumaça, neblina e outros obstáculos. A câmera de luz visível nos dá os detalhes

da cena. Também nesta área, há importância tanto na informação temporal, que descreve a dinâmica de localização dos alvos, quanto na informação espacial, que são os detalhes da cena. Há necessidade de fusão de imagens [18] e vídeos [19] também na área médica.

Neste trabalho é proposto um método para realizar a fusão espaço-temporal de vídeos. Por isso, para realizar as simulações, é necessário um banco de vídeos que contenha uma ampla variedade tanto de informação temporal quanto de informação espacial. O algoritmo proposto aqui será testado usando o banco de dados VLIRVDIF (*Visible-Light and Infrared Video Database for Image Fusion*) [20]. Este banco inclui: diferentes condições de iluminação, objetos oclusos em uma das vistas, uma vasta variedade de objetos em diferentes temperaturas e diferentes planos de cena/fundo em distâncias que variam de 10m a 300m.

## 1.1 Organização

No Capítulo 2 são apresentadas as métricas objetivas de avaliação de qualidade de fusão de imagens/vídeos. A motivação do estudo das métricas objetivas é evitar o uso de testes subjetivos, pois o uso de especialistas apresenta grande dificuldade operacional. Estes testes subjetivos, apesar de demonstrarem a maior confiabilidade dos resultados, são muito caros e muito difíceis de serem implementados por requerer especialistas de cada área. Então, neste capítulo serão mostradas formas de implementar uma avaliação de qualidade objetiva e as ideias por trás de cada uma delas.

No Capítulo 3, são testados vários métodos de fusão de imagens para criar um *benchmark* e podermos fazer comparação com os métodos propostos.

Para realizar a extensão dos métodos espaciais de fusão serão necessárias ferramentas que incluam a informação temporal. O Capítulo 4 traz o arcabouço teórico que proporcionará a adição da compensação de movimento nas decomposições multirresolução espaço-temporais.

No Capítulo 5 são apresentados algoritmos propostos para realizar a fusão multirresolução espaço-temporal e seus respectivos resultados.

Finalmente, no Capítulo 6, são sumarizados os resultados e apresentadas sugestões de trabalhos futuros.

## Capítulo 2

# Métricas Objetivas de Avaliação de Vídeos Fundidos

Fusão de imagens é um processo que cria uma nova imagem a partir de duas outras ou mais. Espera-se que a imagem final seja melhor do que cada entrada individualmente segundo algum critério. Portanto, é preciso avaliar a qualidade da imagem imagem final é após a fusão. A partir desta avaliação, pode-se testar vários métodos de fusão e escolher o melhor entre eles. Ainda é possível fazer um ajuste fino dos parâmetros dos algoritmos para resultar em uma melhor imagem fundida.

Como a fusão de imagens/vídeos pode ser usada em distintas áreas, deve-se encontrar uma medida que possa ser aplicada em todas elas. Além disso, na maioria das aplicações, devemos maximizar a experiência humana ao avaliar o resultado da fusão. Para isso, uma avaliação subjetiva de qualidade é essencial. Entende-se por avaliação subjetiva de qualidade a fornecida por utilizadores do resultado da fusão (nesse caso), que atribuem uma nota a cada sequência. Por as notas dos testes dependerem em da avaliação humana, muitos fatores podem alterar as notas dadas aos testes [21] e tornar a transposição subjetiva-objetiva bastante complicada.

Desejamos realizar esta transposição de maneira que a métrica objetiva tenha resultados próximos aos das avaliações subjetivas. E, com isso, sem a necessidade custosa de inúmeros especialistas em cada área, conseguiremos resultados objetivos bem próximos dos subjetivos, podendo prosseguir com a criação e desenvolvimento de algoritmos nesta área.

As métricas para a avaliação de fusão de vídeo podem ser classificadas em três tipos [22]:

- Métricas Espaciais
- Métricas Temporais
- Métricas Espaço-temporais

O primeiro tipo é focado nos métodos de avaliação de imagens fundidas usando-os diretamente nos vídeos fundidos. O segundo analisa apenas as variações temporais com o intuito de quantificar a qualidade do vídeo. Na literatura, este tipo de avaliação é bem raro. O último tipo une os dois primeiros, adicionando a análise da qualidade temporal à análise da qualidade espacial [23].

Para mostrar a importância de se analisar a consistência temporal, Zhang *et al.* [22] mostraram com testes subjetivos que as variações temporais afetam a avaliação mais que as variações espaciais (na maioria dos casos). Entretanto, quando as qualidades temporal e espacial são priorizadas conjuntamente, as avaliações subjetivas melhoram ainda mais.

Alguns trabalhos de Zhang *et al.* [3, 15, 22, 24] propõem novas métricas como o índice de similaridade estrutural espaço-temporal ( $Q_{ST}$ ) e o índice de congruência de fase espaço-temporal ( $Q_{ST-PC}$ ), ambos baseados nas características espaço-temporais e ajustadas à avaliação subjetiva. Entretanto, essas métricas citadas não são disseminadas na literatura, são muito custosas computacionalmente e os autores não disponibilizaram nenhuma implementação delas.

Neste Capítulo, serão estudados dois dos três tipos citados de métricas: espacial e espaço-temporal. Ele irá concentrar-se nas métricas mais populares e significantes da literatura.

## 2.1 Métricas Espaciais

Ao analisar a estatística de uma imagem, é interessante notar alguns pontos específicos que podem interferir diretamente na experiência perceptual da imagem. Em codificação de imagens, algumas métricas consolidadas de avaliação mensuram a qualidade da imagem após o processo de compressão. Estas métricas fazem menção às características importantes do SVH (Sistema Visual Humano). Por exemplo, uma métrica popular nesta área é o SSIM (*Structural Similarity index*) [25]. Nesta métrica, os autores propõem que a preservação das bordas dos objetos da imagem é uma boa medida para quantificar a qualidade espacial. A razão de computar a qualidade dessa maneira é a alta correlação desta característica com a experiência do SVH.

Por isso, para se falar das métricas espaciais, é necessário entender as preferências do SVH. Neste trabalho, não será proposto nenhum tipo de métrica. Porém, é necessário entender a motivação de cada uma para a interpretação de seus resultados.

No contexto da fusão, a característica estrutural também se mostra muito importante. Tendo que os vídeos usados neste trabalho são gravados a partir de duas câmeras distintas, podem ocorrer pequenos erros na etapa de registro acarretando em feitos indesejados na estrutura dos objetos do vídeo fundido tais como bordas

duplas, objetos deslocados ou duplicados e ainda falsas bordas.

Outros pontos a serem salientados são as características intrínsecas das imagens/vídeos tais como iluminação, brilho, contraste, representação das cores, entre outros [26]. Esses pontos podem melhorar ou piorar a avaliação subjetiva.

Nas próximas Subseções, serão mostradas as métricas mais populares na área de fusão de imagens.

### 2.1.1 Informação Mútua

A respeito da informação espacial, geralmente, tem-se interesse em mensurar a quantidade de informação transferida entre as entradas e a saída. Em teoria da informação é possível realizar esta medida através do conceito de informação mútua. Usando isto, pode-se interpretar a quantidade de informação transferida como um índice de qualidade para a fusão [27–32].

A métrica de Informação Mútua ( $M_{AB/F}$ ) é definida a partir da transferência de informação entre duas imagens de entrada  $I_A$  e  $I_B$  e a imagem fundida  $I_F$ . A métrica pode ser definida, matematicamente, como:

$$M_{AB/F} = \mathcal{I}(I_A; I_F) + \mathcal{I}(I_B; I_F) \quad (2.1)$$

A Equação 2.1 é dependente da informação mútua entre cada par  $(I_A, I_F)$  e  $(I_B, I_F)$ . Cada termo pode ser computado por:

$$\mathcal{I}(I_X, I_F) = \sum_{i=1}^L \sum_{j=1}^L p_{I_X I_F}[i, j] \log_2 \frac{p_{I_X I_F}[i, j]}{p_{I_X}[i] p_{I_F}[j]} \quad (2.2)$$

onde  $L$  é o número de *bins* usados para o cálculo do histograma. Com esse histograma, serão estimadas as funções de densidade de probabilidade.

Pode-se ainda normalizar a métrica usando a entropia das entradas [28], i.e.,

$$M_{AB/F} = \frac{\mathcal{I}(I_A; I_F) + \mathcal{I}(I_B; I_F)}{H(I_A) + H(I_B)} \quad (2.3)$$

Ao longo desta dissertação, será usado sempre a versão normalizada da métrica, tendo a finalidade de uma comparação mais direta entre os métodos. Note que esta métrica é calculada para cada imagem. Para realizar a avaliação de qualidade em um vídeo, será usada a média dos resultados de cada quadro, ou seja, de cada imagem que compõe o vídeo.

## 2.1.2 Métrica de Xydeas e Petrović

A proposta de Xydeas e Petrović [33] é utilizar a informação estrutural presente no gradiente para a medida de qualidade da imagem fundida. Este fato é utilizado porque a preservação da estrutura é fator determinante para o SVH. Para o cálculo do gradiente, é utilizado o operador de Sobel [34] nas imagens  $I_A$ ,  $I_B$  e  $I_F$ . Com este operador é possível calcular o gradiente e avaliar a quantidade de informação transferida para as bordas da imagem fundida. Para a análise do vídeo, será usada a mesma estratégia anterior, a média entre as medidas de cada quadro será utilizada como métrica.

A métrica de Xydeas e Petrović, conhecida como  $Q_{AB/F}$ , é baseada na informação estrutural, calculada com o operador de Sobel que retorna o gradiente espacial de cada imagem. A comparação é feita através da direção e do valor absoluto de cada gradiente da imagem de entrada e da imagem final. Primeiro, pode-se computar o gradiente espacial como:

$$g_{i,j}^X = \sqrt{I_X^h(i,j)^2 + I_X^v(i,j)^2}, \quad (2.4)$$

$$\alpha_{i,j}^X = \tan^{-1} \left( \frac{I_X^v(i,j)}{I_X^h(i,j)} \right), \quad (2.5)$$

onde  $I_X^h$  e  $I_X^v$  são as componentes horizontal e vertical do gradiente, respectivamente, computadas através da filtragem da imagem pelo operador de Sobel. Na Equação 2.6,  $g_{i,j}^X$  e  $\alpha_{i,j}$  representam o valor absoluto do gradiente e a direção das bordas. Essas medidas são dadas por:

$$G_{i,j}^{XF} = \begin{cases} \frac{g_{i,j}^F}{g_{i,j}^X}, & \text{se } g_{i,j}^X > g_{i,j}^F \\ \frac{g_{i,j}^X}{g_{i,j}^F}, & \text{nos outros casos.} \end{cases} \quad (2.6)$$

$$A_{i,j}^{XF} = 1 - \frac{|\alpha_{i,j}^X - \alpha_{i,j}^F|}{\pi/2}, \quad (2.7)$$

onde  $X = A$  ou  $B$ . Pode-se calcular uma medida para a preservação das bordas de cada imagem para a imagem fundida através da seguinte expressão:

$$Q_{i,j}^{XF} = \frac{\Gamma_g}{\sqrt{(1 + e^{k_g(G_{i,j}^{XF} - \sigma_g)})}} \frac{\Gamma_\alpha}{\sqrt{(1 + e^{k_\alpha(A_{i,j}^{XF} - \sigma_\alpha)})}}, \quad (2.8)$$

onde  $\Gamma_g$  and  $\Gamma_\alpha \approx 1$  são calculados analiticamente impondo que a métrica seja

normalizada para ter máximo igual a 1, ou seja, fazendo  $Q^{XF} = 1$  quando  $G^{XF}$  e  $A^{AF}$  também valem 1. Os outros parâmetros da sigmoide são obtidos a partir de testes subjetivos. De acordo com [2, 23, 33, 35], os valores com mais alta correspondência com os testes realizados são para  $(k_g, \sigma_g, k_a, \sigma_a) = (-11; 0.7; -24; 0.8)$ .

Computando a transferência para as duas imagens  $A$  e  $B$ , a métrica de avaliação normalizada  $Q_{AB/F}$  é dada por:

$$Q_{AB/F} = \sum_{i=1}^M \sum_{j=1}^N \frac{Q_{i,j}^{AF} w_{i,j}^A + Q_{i,j}^{BF} w_{i,j}^B}{w_{i,j}^A + w_{i,j}^B}, \quad (2.9)$$

onde os pesos  $w_A[i, j]$  e  $w_B[i, j]$  são iguais a  $(g_A(i, j))^L$  e  $(g_B(i, j))^L$ , respectivamente. Tipicamente, usa-se  $L = 1$ .

Graficamente, pode-se ilustrar [2] a informação transferida para a imagem fundida através da Figura 2.1.

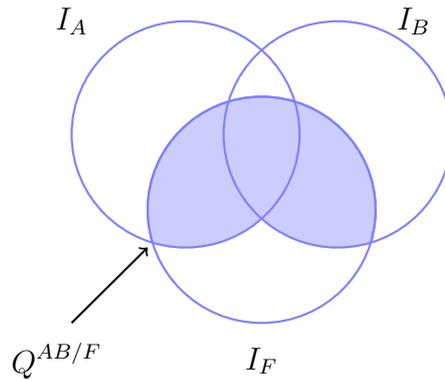


Figura 2.1: Contribuição das informações de  $I_A$  e  $I_B$  em  $I_F$ .

### 2.1.3 Medida de artefatos de fusão

Estendendo a métrica mostrada anteriormente, pode-se extrair mais informações usando a Figura 2.1. Idealmente, espera-se que a imagem fundida contenha informações advindas apenas das duas imagens de entrada. Porém, outras informações aparecem. Esta informação adicional pode ser chamado como artefatos de fusão. Graficamente, estes artefatos estão localizados na área sombreada da Figura 2.2.

Usando a definição das Equações 2.8 e 2.9, pode-se formalizar a definição da medida de artefato de fusão  $N^{AB/F}$ .

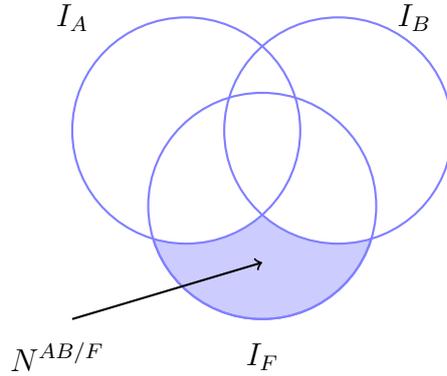


Figura 2.2: Artefatos de fusão presentes em  $I_F$ .

$$N_{i,j} = \begin{cases} 2 - Q_{i,j}^{AF} - Q_{i,j}^{BF}, & \text{se } g_{i,j}^F > g_{i,j}^A \text{ e } g_{i,j}^F > g_{i,j}^B \\ 0, & \text{nos demais casos.} \end{cases} \quad (2.10)$$

$$N^{AB/F} = \frac{\sum_{i=1}^M \sum_{j=1}^N N_{i,j} (w_{i,j}^A + w_{i,j}^B)}{\sum_{i=1}^M \sum_{j=1}^N (w_{i,j}^A + w_{i,j}^B)} \quad (2.11)$$

### 2.1.4 Medida de Informação Perdida

Como mais um desdobramento da métrica  $Q_{AB/F}$ , pode-se também mensurar a informação estrutural perdida ao realizar a fusão. A métrica  $N_{AB/F}$  é definida [2] como pode-se ver na Figura 2.3.

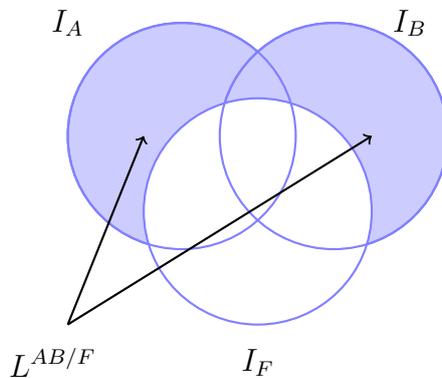


Figura 2.3: Informação perdida devido a fusão.

A expressão que descreve essa informação é dada por:

$$r_{i,j} = \begin{cases} 1, & \text{se } g_{i,j}^F > g_{i,j}^A \text{ ou } g_{i,j}^F > g_{i,j}^B \\ 0, & \text{nos demais casos.} \end{cases} \quad (2.12)$$

$$L^{AB/F} = \frac{\sum_{i=1}^M \sum_{j=1}^N r_{i,j} [(1 - Q_{i,j}^{AF})w_{i,j}^A + (1 - Q_{i,j}^{BF})w_{i,j}^B]}{\sum_{i=1}^M \sum_{j=1}^N w_{i,j}^A + w_{i,j}^B}. \quad (2.13)$$

Com esta definição e as anteriores, pode-se entender toda a informação estrutural extraída da preservação dos gradientes da imagem final.

### 2.1.5 Métrica de Piella

Com o intuito de salientar pontos importantes para o SVH, a métrica de Piella [28] propõe iniciar com a ideia do índice SSIM [25]. Igualmente às métricas de Xydeas e Petrović, a métrica de Piella propõe o cálculo da preservação estrutural calculada usando janelamento.

Começando pela maneira em que o índice de qualidade de imagens SSIM foi definido por Wang e Bovik [25], temos que:

$$Q_0(I_X, I_F) = \frac{4\sigma_{xf}\bar{x}\bar{f}}{(\bar{x}^2 + \bar{f}^2)(\sigma_x^2 + \sigma_f^2)} \quad (2.14)$$

onde  $\bar{x}$  e  $\bar{f}$  são as médias de  $I_X$  e  $I_F$ , respectivamente. As outras variáveis são definidas por:

$$\sigma_x^2 = \frac{1}{MN - 1} \sum_{i=1}^M \sum_{j=1}^N (I_X(i, j) - \bar{x})^2, \quad (2.15)$$

$$\sigma_f^2 = \frac{1}{MN - 1} \sum_{i=1}^M \sum_{j=1}^N (I_F(i, j) - \bar{f})^2, \quad (2.16)$$

$$\sigma_{xf}^2 = \frac{1}{MN - 1} \sum_{i=1}^M \sum_{j=1}^N (I_X(i, j) - \bar{x})(I_F(i, j) - \bar{f}). \quad (2.17)$$

onde  $M, N$  são as dimensões das imagens em questão.

Sabendo que as imagens são sinais, geralmente não-estacionários [25, 28], temos que calcular o índice SSIM usando uma janela deslizante  $w$  que percorre a imagem

por completo. Definimos, então, a métrica como:

$$Q_0(I_X, I_F) = \frac{1}{|W|} \sum_{w \in W} Q_0(I_x, I_F|w), \quad (2.18)$$

onde  $W$  é o conjunto de todas as janelas,  $|W|$  é sua cardinalidade.

Pode-se denotar  $s(I_A|w)$  como alguma saliência na imagem  $I_A$  na janela  $w$  (igualmente  $s(I_B|w)$  com respeito a  $I_B$ ). A sugestão apresentada para mensurar a qualidade de fusão [28] é usar a variância de cada bloco como esta saliência. Após isso, definimos  $C(w) = \max(s(I_A|w), s(I_B|w))$  como a saliência global da janela  $w$ . Com isso, pode-se computar a medida de qualidade de fusão como:

$$Q_w(I_A, I_B, I_F) = \sum_{w \in W} C(w) \frac{s(I_A|w)Q_0(I_A, I_F|w) + s(I_B|w)Q_0(I_B, I_F|w)}{s(I_A|w) + s(I_B|w)}. \quad (2.19)$$

A métrica de Piella  $Q_p$ , também conhecida como índice de qualidade de fusão dependente de borda (do inglês *edge-dependent fusion quality index*), é definido como:

$$Q_P(I_A, I_B, I_F) = Q_w(I_A, I_B, I_F)^{1/2} Q_w(\nabla I_A, \nabla I_B, \nabla I_F)^{1/2}, \quad (2.20)$$

onde  $\nabla I_A$ ,  $\nabla I_B$  e  $\nabla I_F$  são as *imagens de gradientes* de  $I_A$ ,  $I_B$  e  $I_F$  respectivamente. Estas imagens são a norma dos gradientes das imagens em questão.

## 2.2 Métricas Espaço-temporais

Muitos trabalhos foram publicados a respeito da degradação espacial das imagens fundidas e seus efeitos no SVH [2, 25, 28]. Entretanto, a informação temporal também é relevante subjetivamente [22] e não pode ser negligenciada.

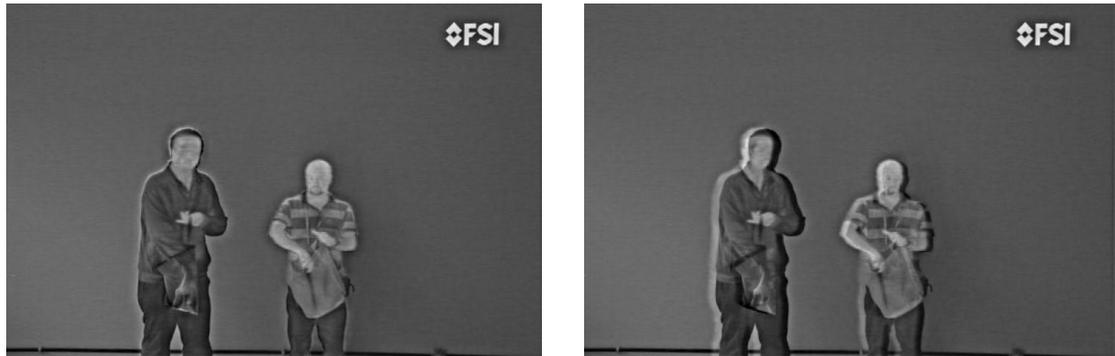
Os vídeos fundidos trazem características diferentes das imagens naturais, na maioria dos casos. E por isso, alguns efeitos indesejados ao SVH podem aparecer [22]. Para reduzir esses efeitos, primeiramente, tem-se que defini-los e entender suas consequências ao usuário.

Ao definir métricas espaço-temporais, deve-se levar em conta os seguintes pontos:

### 1. Erro de registro:

O processo de registro é responsável por alinhar e escalar as imagens capturadas a fim que haja equivalência pixel a pixel de um mesmo ponto de vista. Este processo faz uma suposição importante que todos os objetos em cena

estejam no mesmo plano. Como em vídeos naturais há mais de um plano, ao se realizar a fusão, o erro de registro se torna aparente. Na imagem fundida, há aparições de bordas duplas, ou falsas bordas, ou ainda outro efeito indesejado. Na Figura 2.4 é mostrada uma imagem onde é imposto um erro de registro (ilustrado aqui como um deslocamento horizontal de 10 pixels). Nesta Figura é possível entender os efeitos do desalinhamento dos objetos em cena na imagem fundida.



(a) Imagem fundida a partir de imagens registradas.

(b) Imagem fundida a partir de imagens com erro de registro (deslocamento horizontal de 10 pixels).

Figura 2.4: Comparação entre as imagens fundidas usando entradas registradas ou com erro de registro. Foi imposto, artificialmente, um erro de registro de 10 pixels. É possível observar que bordas duplas são mais presentes em (b).

## 2. Conteúdo espacial:

Os artefatos espaciais que podem ocorrer nas imagens fundidas [2, 25, 28] podem ter seus efeitos agravados ao longo do tempo. Em outras palavras, mudanças drásticas na estatística das imagens podem resultar em efeitos desagradáveis ao usuário [36].

## 3. Conteúdo temporal:

Métodos de fusão de vídeo devem garantir uma boa representação do movimento, fato desconsiderado ao realizar apenas a fusão espacial. Deve-se garantir também que o processamento temporal não adicione artefatos.

Os pontos enumerados servem de guia para um estudo das métricas espaço-temporais. De maneira geral, as métricas mostradas na literatura seguem as diretrizes anteriores, exceto quanto ao erro de registro. Este ponto é englobado pelas métricas de maneira indireta na preservação do conteúdo espacial. A seguir, serão estudadas métricas espaço-temporais descritas e amplamente divulgadas na literatura.

### 2.2.1 Informação Mútua de diferença entre quadros

A métrica baseada na informação mútua de diferença entre quadros (MI-IFD – *Mutual information from interframe differences*) é popularmente utilizada para a avaliação de vídeos fundidos [3, 22–24, 28]. Tal como a métrica de informação mútua, esta analisa apenas a quantidade de informação transferida entre os vídeos de entrada e o vídeo fundido. Para adicionar a informação temporal, é usado as diferenças entre dois quadros consecutivos.

Rockinger *et al.* [37] define a diferença entre quadros IFD(*interframe difference*) como:

$$dS_X(i, j, t) = S_X(i, j, t) - S_X(i, j, t + 1), \quad (2.21)$$

onde  $S_X$  é um dos vídeos ( $S_A$ ,  $S_B$  ou  $S_F$ ).

Define-se a variável aleatória multidimensional  $\mathbf{S} = (dS_A, dS_B)$ , onde uma realização desta variável é um par composto pelos pixels de um IFD de  $S_A$  e um IFD de  $S_B$ . Da mesma maneira, define-se a variável aleatória multidimensional  $\mathbf{U} = (dS_A, dS_B, dS_F)$  e  $\mathbf{F} = (dS_F)$ . Formalmente, podemos definir a métrica MI-IFD como:

$$\text{MI-IFD} = I(\mathbf{S}; \mathbf{F}) = H(\mathbf{S}) + H(\mathbf{F}) - H(\mathbf{U}). \quad (2.22)$$

onde  $I(\cdot)$  é a informação mútua e  $H(\cdot)$  é a entropia das variáveis aleatórias em questão.

Qualitativamente, pode-se acompanhar a dinâmica desta métrica através das variações das entropias [37]. Referindo-se a Equação 2.22, se o nível de cinza da sequência fundida varia e esta mudança não pode ser explicada pelas sequências de entrada, a entropia  $H(\mathbf{U})$  reduz a informação mútua. Para representar a inconsistência temporal, pode-se acompanhar a evolução da diferença  $H(\mathbf{U}) - H(\mathbf{S})$ . Quando menor este termo, maior estabilidade temporal.

### 2.2.2 Avaliação de Qualidade Dinâmica

Essa métrica é uma extensão direta da métrica de Xydeas e Petrović. A métrica  $Q_{AB/F}$  é computada através do total de bordas preservadas após a fusão, onde essa medida é obtida através do cálculo dos gradientes. A avaliação dinâmica de qualidade (DQ – *Dynamic Quality Evaluation*) é a extensão direta desta métrica acrescentando a preservação das bordas ao longo do eixo temporal também.

Por isso, dá-se início do desenvolvimento desta métrica com a ideia da métrica  $Q_{AB/F}$ . Neste caso, é adicionado o cálculo dos gradientes temporais que podem ser definidos como:

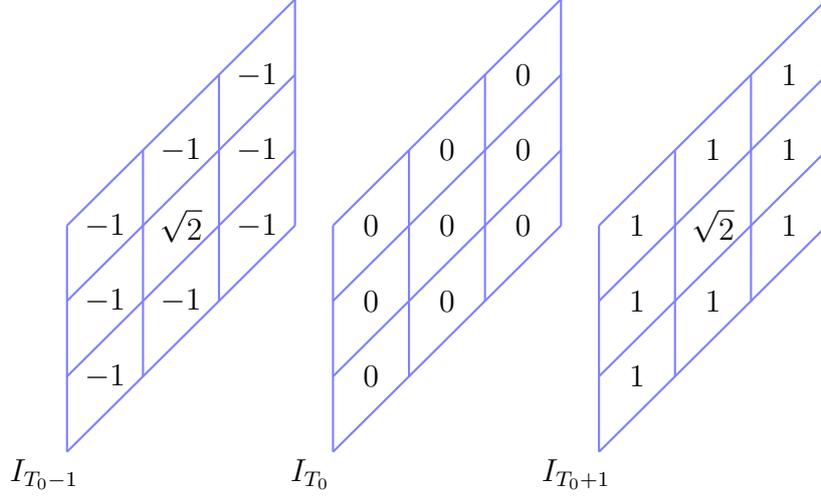


Figura 2.5: Operador Sobel estendido para a dimensão temporal.

$$gt_{T_0} = |I_{T_0-1} * S_{T-1} + I_{T_0+1} * S_{T+1}|, \quad (2.23)$$

onde  $S_{T_0-1}$  e  $S_{T_0+1}$  são os operadores de Sobel (descritos na Figura 2.5);  $I_{T-1}$  e  $I_{T+1}$  são os quadros anterior o posterior ao quadro atual.

De forma idêntica ao caso espacial, a matriz  $T^{XF}$  pode ser escrita como:

$$T_{T_0}^{XF}(i, j) = \begin{cases} \frac{gt_{T_0}^F(i, j)}{gt_{T_0}^X(i, j)}, & \text{se } gt_{T_0}^X(i, j) > gt_{T_0}^F(i, j), \\ \frac{gt_{T_0}^X(i, j)}{gt_{T_0}^F(i, j)}, & \text{para os demais casos.} \end{cases} \quad (2.24)$$

onde, novamente,  $X$  representa a sequência  $A$  ou  $B$ .

Tendo a definição do gradiente temporal, pode-se computar a preservação ao longo do tempo:

$$Qt_{T_0}^{XF}(i, j) = \frac{\Gamma}{\sqrt{(1 + e^{k_g(T_{T_0}^{XF}(i, j) - \sigma_g)})}}, \quad (2.25)$$

$k_g$ ,  $\sigma_g$  são obtidos através de testes subjetivos ( $(k_g, \sigma_g) = (-11; 0.7)$ ) e  $\Gamma = \sqrt{1 + e^{k_g(1 - \sigma_g)}}$  (aprox. 0.98).

Para computar o índice de avaliação de qualidade dinâmica, deve-se adicionar um processamento adicional à análise do gradiente temporal. O SVH é mais sensível aos movimentos de maior amplitude. Então, a métrica deve ser pesada para privilegiar

a manutenção destes movimentos na sequência fundida [23]. Este peso é dado por:

$$w_{i,j}^X = g_{i,j}^X + |gt_{i,j}^X| \quad (2.26)$$

Os pesos descritos na Equação 2.26 serão usados para unir a representação da informação temporal e espacial das sequências  $A$  e  $B$ . Ainda assim, é necessário combinar os resultados da preservação espacial e temporal de cada sequência individualmente. Em Petrović *et al.* [23] foram unidos os fatores de preservação em um único fator através da expressão:

$$Q_{i,j}^{XF} = \sqrt{Qs_{i,j}^{XF} + Qt_{i,j}^{XF}}, \quad (2.27)$$

e, finalmente, pode-se sintetizar a preservação das bordas temporais e espaciais na métrica DQ que avalia a qualidade dinâmica como:

$$DQ = \frac{1}{N_T} \sum_{T=1}^{N_T} \left\{ \frac{\sum_{i=1}^M \sum_{j=1}^N Q_{n,m}^{AF} w_{n,m}^A + Q_{n,m}^{BF} w_{n,m}^B}{\sum_{i=1}^M \sum_{j=1}^N w_{n,m}^A + w_{n,m}^B} \right\}, \quad (2.28)$$

onde  $N_T$  é o número de quadros do vídeo.

## 2.3 Validação

Todas as métricas mostradas até aqui seguem uma mesma premissa que é a coerência com os critérios perceptuais. Em outras palavras, estas métricas precisam ser correlacionadas com uma avaliação de acordo com a visão humana. Isto pode ser feito através dos testes subjetivos, como já foi dito.

Há alguns testes subjetivos na área de fusão [35, 38] que servirão de guia para confiabilidade das métrica usadas neste trabalho. O primeiro experimento a ser discutido [35] foi realizado com mais de 100 avaliadores e 9 algoritmos de fusão distintos. Para avaliar a correlação entre os testes e o resultado das métricas, foram utilizados dois métodos [31]: o *Correct Ranking (CR)* e a relevância subjetiva ( $r$ ).

Ambos os métodos são baseados no vetor de preferência subjetiva ( $p_n$ ) definido como:

$$p_n = \begin{bmatrix} p_n^0 \\ p_n^1 \\ p_n^2 \end{bmatrix} = \begin{cases} [1, 0, 0]^T, & \text{avaliador } n \text{ não tem preferência.} \\ [0, 1, 0]^T, & \text{avaliador } n \text{ prefere a imagem fundida pelo esquema 1.} \\ [0, 0, 1]^T, & \text{avaliador } n \text{ prefere a imagem fundida pelo esquema 2.} \end{cases} \quad (2.29)$$

e também na vetor de preferência objetiva ( $q_n$ ):

$$q_n = \begin{bmatrix} q_n^0 \\ q_n^1 \\ q_n^2 \end{bmatrix} = \begin{cases} [1, 0, 0]^T, & \text{if } Q_1 \approx Q_2, \\ [0, 1, 0]^T, & \text{if } Q_1 > Q_2, \\ [0, 0, 1]^T, & \text{if } Q_1 < Q_2. \end{cases} \quad (2.30)$$

O *Correct Ranking* ( $CR$ ) é dado por:

$$CR = \frac{1}{N} \sum_{\forall n} q_n^T p_n. \quad (2.31)$$

A relevância subjetiva ( $r$ ) é descrita de maneira similar incluindo a certeza relativa das avaliações subjetivas. Ou seja, altera-se a descrição binária do vetor  $p_n$  por uma versão normalizada da distribuição das preferências  $t_n$  e isso é mensurado de maneira que

$$r = \sum_{\forall n} q_n^T t_n. \quad (2.32)$$

Dados estes critérios de validação, pode-se prosseguir para a análise das métricas apresentadas. Em [35] é apresentado um teste subjetivo deste tipo. Os resultados da validação estão descritos na Tabela 2.1.

	$Q_p$	$M_{AB/F}$	$Q_{AB/F}$
$r$	0.737	0.742	0.833
$CR$	63.3%	62.5%	72.5%

Tabela 2.1: Correspondência subjetiva com as métricas de avaliação objetivas de fusão de imagens mostradas.

De outra fonte, o trabalho [38] mostra outra validação. Na Tabela 2.2 estão disponíveis os resultados. Note que mais um fator de correspondência foi acrescentado, o AUC (*Area under the ROC Curve*). Este fator utiliza a mensuração da área sob a curva de características operacionais do receptor, ROC (*Receiver Operating Characteristic*) [38]. Neste experimento, 120 grupos de imagens fundidas separados em 8 subgrupos foram usados. O número de avaliadores não foi informado.

Em outro trabalho, Zhang *et al.* [22] realizou testes com 12 avaliadores, 4 vídeos fundidos e 4 diferentes algoritmos. Este trabalho é interessante pois apresenta resultados para vídeos. Os avaliadores julgaram o vídeo de acordo com sua qualidade temporal, espacial e espaço-temporal, formando 3 grupos de resultados distintos.

Infelizmente, os fatores de correspondência não foram calculados no trabalho. Um dos principais motivos é a pequena quantidade de avaliações neste experimento.

	$Q_p$	$M_{AB/F}$	$Q_{AB/F}$
$r$	0.493	0.642	0.573
$CR$	60.8%	64.2%	71.7%
$AUC$	0.754	0.625	0.795

Tabela 2.2: Resultados da validação das métricas espaciais incluindo a validação AUC.

Para contornar este fato, pode-se fazer uma análise qualitativa.

Usando o conjunto de pontos formados pelo escore médio de opinião - MOS (*Mean Opinion Score*) e os resultados das métricas objetivas também disponíveis em [22], pode-se calcular a correlação entre elas através de um ajuste linear deste par de pontos normalizados.

Os ajustes lineares estão disponíveis no Apêndice A. Os resultados comparativos estão na Tabela 2.3.

	Questão Subjetiva		
	Qualidade Espacial	Qualidade Temporal	Qualidade Espaço-temporal
coef. angular ( $Q_p, MOS$ )	0.063478	0.051995	0.051355
coef. angular ( $Q_{AB/F}, MOS$ )	0.171873	0.174318	0.174208
coef. angular (MI-IFD, MOS)	<b>1.119968</b>	<b>0.822405</b>	<b>0.838932</b>
coef. angular (DQ, MOS)	0.510302	0.446472	0.448519

Tabela 2.3: Inclinação das curvas MOS  $\times$  valor predito pelas métricas objetivas. Esta inclinação foi alcançada através de um ajuste linear. Os pares de valores foram normalizados a fim de que a maior correlação seja expressa no coeficiente angular igual a 1.

Nesta análise, pode-se notar uma correlação com os testes subjetivos maior com a métrica MI-IFD para todas as três questões. Novamente, esta análise é apenas para quantificar os resultados obtidos pelo o artigo [22]. Como são poucos testes, os resultados também são pouco confiáveis.

## 2.4 Conclusões

Neste Capítulo foram apresentadas algumas métricas de avaliação objetiva para imagens/vídeos. De acordo com a literatura, estas métricas foram escolhidas (e amplamente disseminadas) pela boa correlação com testes subjetivos ou pela baixa complexidade computacional [31].

Os resultados apresentados na Seção 2.3 mostram uma ligeira superioridade da métrica  $M_{AB/F}$  (Equação 2.3) em relação a métrica de Piella. Também é mostrada uma superioridade da métrica  $Q_{AB/F}$  (Equação 2.9). Em relação ao custo computacional, nossos experimentos mostraram que a métrica de Xydeas e Petrović e a

informação mútua são cerca de 40 vezes mais rápidas que a métrica de Piella. Em outras palavras, há um alto custo computacional desta última métrica.

Concluindo este Capítulo, foram comparadas tanto métricas espaciais quanto espaço-temporais. Foram vistas as correlações destas métricas de acordo com testes subjetivos disponíveis na literatura. As ramificações do  $Q_{AB/F}$  não serão utilizadas nesta dissertação. Porém, estas podem medir pontos interessantes na fusão e por isso foram apresentadas.

Dos testes apresentados em [22], foi possível validar (graficamente) a correlação entre o escore médio de opinião (MOS) e as métricas objetivas (principalmente as espaço-temporais). Porém, por apresentarem um número baixo de avaliações, os resultados não são definitivos.

# Capítulo 3

## Fusão Multirresolução de Imagens

Transformadas multirresolução são amplamente usadas na literatura para a análise e fusão de imagens. No âmbito de fusão de imagens, estes métodos são populares pois permitem ressaltar detalhes de diferentes faixas de frequências temporal e espacial. Mais ainda, as representações multirresolução trazem outras características interessantes aos métodos de fusão que iremos discutir mais adiante.

Pelas vantagens deste tipo de abordagem, nos últimos 20 anos, a pesquisa na área de fusão usando decomposição multirresolução vêm crescendo. Pode-se citar as decomposições multirresolução tais como a Transformada de Fourier Discreta (DFT) , Transformada *Wavelet* Discreta (DWT) [39], Transformada *Curvelet* (CVT) [40], Transformada *Contourlet* (ConT) [6] e Transformada *Wavelet* não-decimada (UWT) [41], entre outras.

Em meio a tantas decomposições, é interessante notar as características que as tornam favoráveis no âmbito de fusão de imagens ao nível do pixel, tais como [6, 42, 43]:

### 1. Representação Esparsa:

Na área de processamento de sinais, uma representação esparsa é muito interessante, especialmente em processamento de imagens. Tal representação é capaz de salientar a informação essencial para o SVH em poucos coeficientes da decomposição [6, 44]. Por isso, como a fusão tem por objetivo reunir tais informações, uma representação esparsa pode ser necessária.

### 2. Resolução espaço-frequencial:

Deve-se considerar que algumas características importantes podem ser extraídas ao avaliarmos as características espaciais e frequenciais da imagem [45] conjuntamente. Ao passarmos à área de fusão de vídeos, estaríamos interessados em resolução tempo-espaço-frequencial.

Como exemplo [43, 46], pode-se comparar a representação espaço-frequência das transformadas *Short-time* Fourier (STFT) com a *Wavelet* (Figura 3.1).

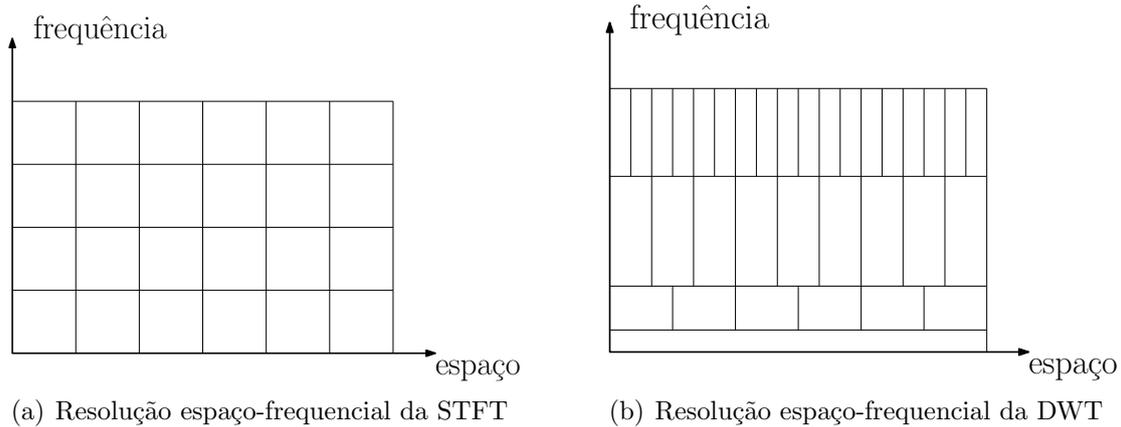


Figura 3.1: Comparação da resolução espaço-frequencial entre a STFT e DWT.

A STFT é usada aqui apenas para ilustrar o problema da resolução espaço-frequencial. Do lado esquerdo, a resolução espacial alcançada para altas frequências é igual a alcançada para baixas frequências. No entanto, do lado direito, a DWT aumenta a resolução temporal de acordo com o aumento do aumento da frequência. Isso se traduz na possibilidade de se representar melhor características pequenas de alta frequência espacial através da transformada.

### 3. Anisotropia:

Imagens naturais são formadas por contornos suaves. Pela suavidade, estes contornos necessitam de uma boa representação dada pela transformada. Por isso, é necessário que a base dada pela transformada represente bem todos estes contornos. Uma base anisotrópica tem as direções ajustadas de acordo com os contornos e pode ser útil para esta representação ser bem sucedida.

Devido a estes fatos, neste Capítulo irão ser estudadas algumas técnicas de fusão de vídeo usando decomposição multirresolução que serão aplicadas sem levar em conta o eixo temporal. Ou seja, em cada quadro do vídeo será aplicada a técnica de fusão de imagens a ser estudada.

## 3.1 Técnicas de Fusão de Imagens Multirresolução

Para analisar imagens, um conceito interessante e muito utilizado é a decomposição multirresolução, como já foi dito. No contexto de fusão, estas técnicas permitem separar cada escala e analisá-la individualmente [47].

Podemos definir uma notação comum a todas as técnicas de decomposição multirresolução para poder estudá-las. Como notação da imagem de entrada tem-se  $x^0$

que é decomposta no conjunto de coeficientes multirresolução  $\mathbf{y}$  pela transformada  $\Psi$  como [42, 47]:

$$\mathbf{y} = \{y^1, y^2, \dots, y^J, x^J\}, \quad (3.1)$$

onde  $x^J$  é chamada de imagem aproximada do nível mais alto de decomposição (mais baixa resolução) e  $y^j, j = 1, 2, \dots, J$  é chamada de imagem de detalhes do nível  $j$ .

A decomposição multirresolução dada pela transformada  $\Psi$  será parte de um algoritmo de fusão. Um diagrama para este algoritmo é dado na Figura 3.2.

Neste diagrama, as imagens de entrada  $I_A$  e  $I_B$  são submetidas à decomposição  $\Psi$  e são representadas pelo conjunto de coeficientes  $y_A$  e  $y_B$ . A partir disso, tem-se que unir estes coeficientes em apenas um conjunto. Para isso, é aplicada a regra de fusão  $\Theta$ . Esta regra resulta nos pesos  $w_A$  e  $w_B$  que são usados para combinar os dois conjuntos de coeficientes. Aplica-se a transformada inversa e a imagem fundida  $I_F$  é obtida.

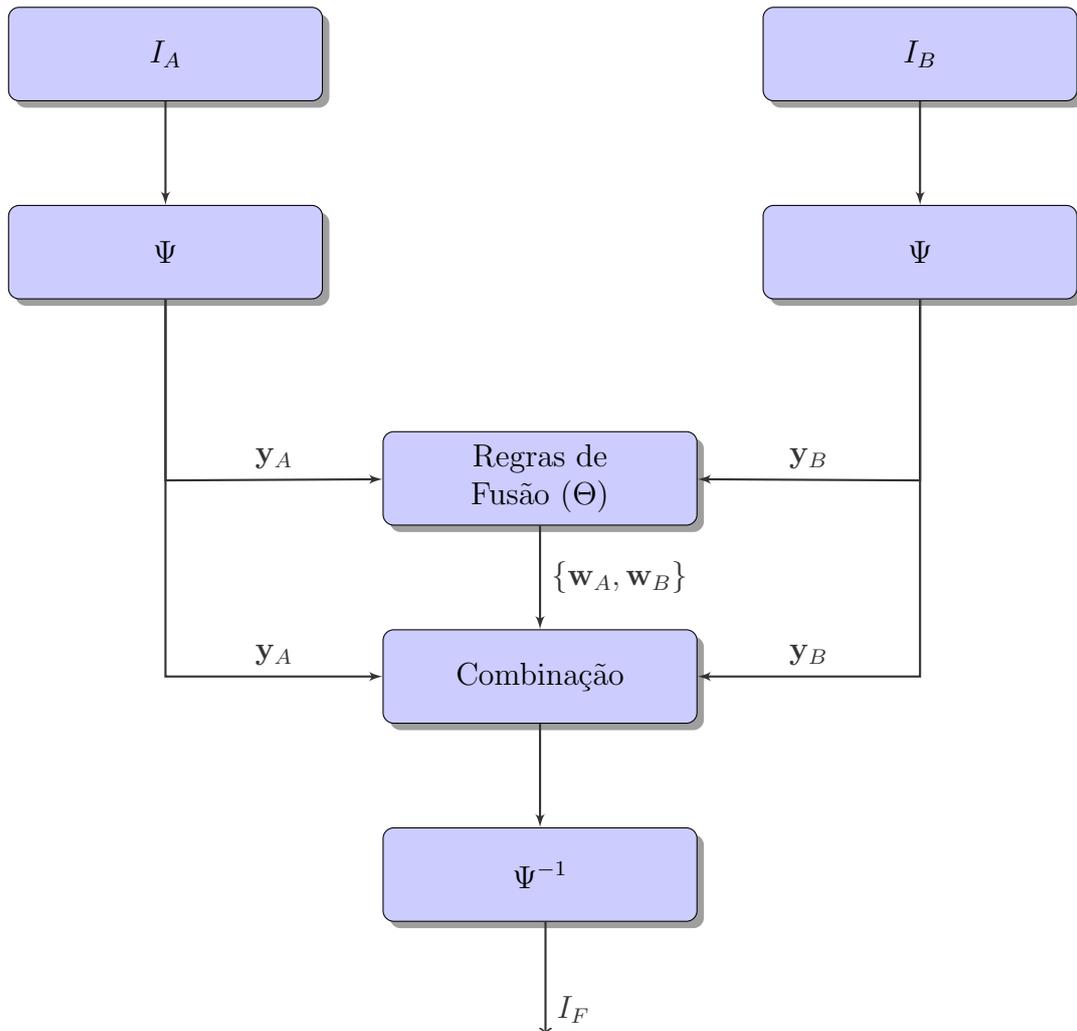


Figura 3.2: Diagrama para exemplificar os passos de um dado algoritmo de fusão.

Seguindo a ordem e a notação vista, primeiramente, serão mostrados métodos de decomposição multirresolução usados em algoritmos de fusão de imagens atualmente e após isso, as regras de fusão que são comumente empregadas.

### 3.1.1 Transformada *Wavelet* Discreta

A Transformada *Wavelet* Discreta (DWT) é amplamente usada e explorada em diferentes áreas desde a década de 90. Em fusão de imagens, geralmente, a DWT é usada como comparação inicial. A DWT proporciona uma análise em diferentes escalas frequenciais e espaciais permitindo uma análise interessante para a fusão de imagens. Essa transformada ainda é ortogonal e invertível. Por conta destas características, há uma ampla variedade de aplicações que vão desde a física quântica [48] à codificação de sinais [49], com forte uso na área de processamento de imagens.

A DWT é baseada na decomposição do sinal em um conjunto de funções base formada por compressões, expansões e translações da função *wavelet* mãe  $\psi[n]$ . Usando um sinal unidimensional  $I[n]$ , podemos descrever a decomposição *wavelet* discreta como:

$$I[n] = \sum_{l=-\infty}^{\infty} \sum_{m=0}^{\infty} y^m[n] \psi_{m,l}[n] \quad (3.2)$$

onde  $m$  é a resolução e  $\psi_{m,l}[n]$  é a função *wavelet* mãe dilatada e/ou deslocada dada pela equação:

$$\psi_{m,l}[n] = 2^{-m/2} \psi[2^m n - l] \quad (3.3)$$

Esta abordagem garante que o sinal é decomposto em *wavelets* normalizadas em uma escala de oitavas  $(2^{-m}n - l)$  [7]. Neste ponto, pode-se entender o que acontece com o sinal  $I[n]$  após a decomposição. Usando um número infinito de coeficientes de detalhes  $y^m$ , pode-se escrever  $I[n]$ . Entretanto, podemos definir um número ( $L$ ) máximo de resoluções para uma análise. Com essa limitação, define-se a função escala  $\phi_{L,l}[n]$  tal que

$$I[n] = \sum_{l=-\infty}^{\infty} x^m \phi_{L,l}[n] + \sum_{l=-\infty}^{\infty} \sum_{m=0}^L y^m \psi_{m,l}[n]. \quad (3.4)$$

Pode-se entender esta operação como a projeção do sinal  $I[n]$  sobre as funções

bases  $\phi_{L,l}[n]$  e  $\psi_{m,l}[n]$  através do produto interno:

$$x^m[n] = \langle I[n], \phi_{m,n}, \rangle \quad (3.5)$$

$$y^m[n] = \langle I[n], \psi_{m,n}, \rangle \quad (3.6)$$

Há uma maneira eficiente de computar os coeficientes de cada escala [43] sem explicitar as funções mãe e escala. Definindo os filtros de análise  $h[n]$  e  $g[n]$  (passa-baixas e passa-altas respectivamente), pode-se escrever os coeficientes de cada escala como:

$$x^{m+1}[n] = \sum_{l=-\infty}^{\infty} h[l - 2n]x^m[l], \quad (3.7)$$

$$y^{m+1}[n] = \sum_{l=-\infty}^{\infty} g[l - 2n]x^m[l]. \quad (3.8)$$

Para garantir que a decomposição tenha reconstrução perfeita, ou seja, que a transformação tenha inversa tal que o sinal original possa ser recuperado sem perdas, é preciso haver  $\tilde{h}[n]$  e  $\tilde{g}[n]$  que satisfaçam a condição

$$H(z)\tilde{H}(z) + G(z)\tilde{G}(z) = 1, \quad (3.9)$$

onde  $H(z)$  e  $G(z)$  são as transformadas Z dos filtros análise  $h[n]$  e  $g[n]$ ;  $\tilde{H}(z)$  e  $\tilde{G}(z)$  são as transformadas Z dos filtros de síntese  $\tilde{h}[n]$  e  $\tilde{g}[n]$ , respectivamente.

Usando os filtros de síntese, podemos definir a transformada inversa e recuperar o sinal de entrada

$$I[n] = \sum_{l=-\infty}^{\infty} \tilde{h}[n - 2l]x^{j+1}[l] + \sum_{l=-\infty}^{\infty} \tilde{g}[n - 2l]y^{j+1}[l]. \quad (3.10)$$

### Transformada *Wavelet* Discreta Multidimensional

No intuito de decompor um sinal multidimensional como uma imagem ou um vídeo, deve-se usar a versão multidimensional da DWT. Porém, a extensão da DWT unidimensional (como mostrado) é simples: deve-se aplicar a decomposição em cada dimensão.

Então, para imagens, pode-se aplicar a DWT nas colunas e nas linhas (2D), e, para vídeos, pode-se aplicar nas colunas, linhas e na dimensão temporal (3D).

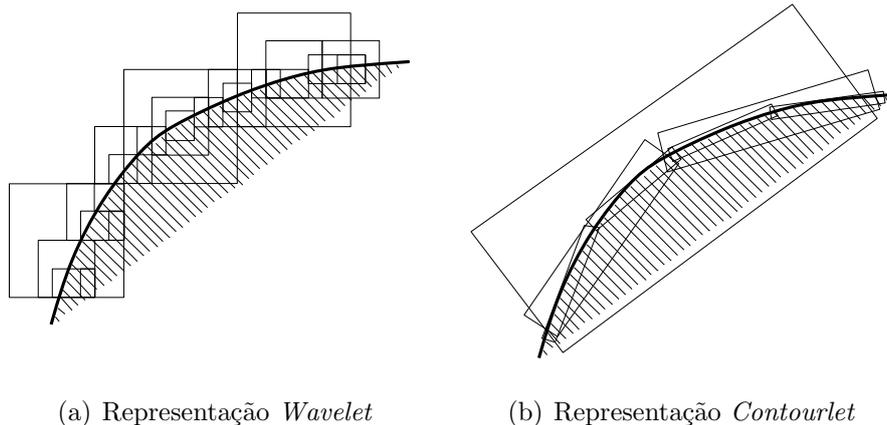


Figura 3.3: Comparação entre as descrições em bases *wavelet* e *contourlet* de imagens naturais.

### 3.1.2 Transformada *Contourlet*

O sistema visual humano é sensível a diversas características. Estas características devem ser levadas em conta quando quisermos realizar um dado algoritmo de fusão. A proposta da Transformada *Contourlet* (ConT) é garantir a eficiência da transformada para imagens naturais. Uma imagem natural é, geralmente, composta por linhas e contornos suaves [6]. A suavidade dos contornos faz com que as bases tradicionais precisem de muitas escalas para garantir uma dada representação (Figura 3.3(a)). Além disso, pode-se notar que as características geométricas das imagens naturais não são diretamente representadas pelas bases tradicionais utilizadas na DWT.

A ConT propõe uma decomposição multiescala/multirresolução direcional (Figura 3.3(b)). Ou seja, ajusta-se cada filtragem na direção dos contornos para garantir uma melhor descrição para os contornos nas bases da transformada *Contourlet*.

Para realizar esta filtragem direcional, Do *et al.* [6] propuseram um esquema baseado em dois passos (Figura 3.4). No primeiro passo, a Pirâmide Laplaciana (Figura 3.5) captura as discontinuidades presentes na imagem. Em seguida, um banco de filtros direcional (DFB) é responsável por conectar os pontos de descontinuidade numa estrutura linear.

Brevemente, podemos descrever as duas etapas de filtragem propostos pela Cont. A Pirâmide Laplaciana (LP) foi proposta por Burt e Adelson em 1983 [50]. A Figura 3.5 representa um nível de decomposição/reconstrução dela. Cada nível de decomposição gera uma versão passa-baixas da entrada e a diferença entre essa versão e a entrada, ou seja, uma versão passa-banda [6]. Ainda na Figura 3.5,  $H$  e  $G$  são os filtros de análise e de síntese respectivamente e  $M$  a matriz de amostragem.

Essencialmente, a ConT desacopla a decomposição multirresolução da decom-

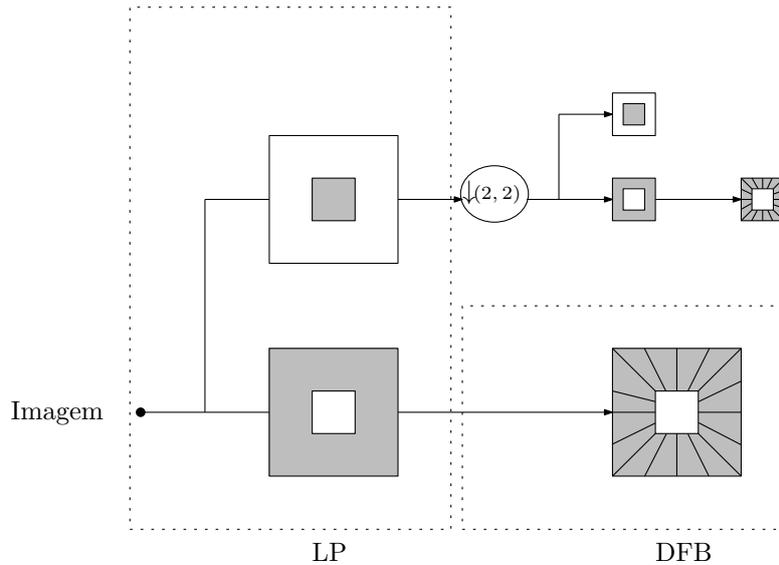


Figura 3.4: Esquema *Contourlet*: O processo de filtragem dupla é composta por a filtragem Laplaciana (*Laplacian Pyramid*) seguida pela filtragem direcional (*Directional Filter Bank*). Outro nível de filtragem é mostrado após o decimador  $(2, 2)$  que representa a subamostragem das linhas e colunas pelo fator 2.

posição direcional. A LP é responsável pela decomposição multirresolução. Para realizar a decomposição direcional é utilizado o esquema de banco de filtros direcional. As duas imagens resultantes do LP são entrada para o DFB. Este passo tem por objetivo realizar o rastreamento dos contornos para uma melhor representação das discontinuidades. O DFB usado na ConT é descrito em detalhes em [6]. Podemos entendê-lo, brevemente, como um banco de filtros com  $2^l$  canais paralelos. A variável  $l$  define o número de direções usadas (Figura 3.6). Dividindo as direções frequencialmente, é possível realizar a filtragem de acordo com as direções das discontinuidades ou dos contornos.

Concluindo, a ConT foi proposta com o intuito de representar as imagens naturais de forma mais precisa. Esta representação é alcançada através dos dois passos de filtragem. Estes passos têm por objetivo mapear os contornos e descrevê-los. Com isso, garante-se um ponto dito no início deste Capítulo: a anisotropia. Comparando esta característica, pode-se ver que a DWT representa bem os detalhes horizontais, verticais e na diagonal (direção  $45^\circ$ ). A Cont consegue mais direções que estas 3 direções, fazendo  $l \geq 2$ .

Para realizar os nossos testes de fusão, será utilizado uma *toolbox* disponibilizada pelos autores em [51].

Uma outra ramificação desta técnica é a Transformada *Contourlet* não-subamostrada (NSCT). Esta técnica contém algumas diferenças em relação ao ConT. A descrição do algoritmo pode ser encontrada em [8]. A *toolbox* é encontrada em [52].

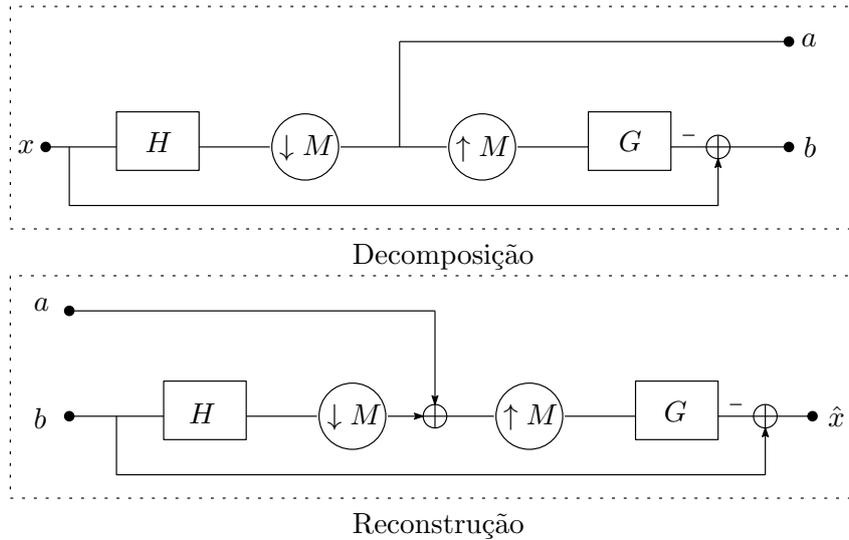


Figura 3.5: Esquema LP para um nível de decomposição e reconstrução.

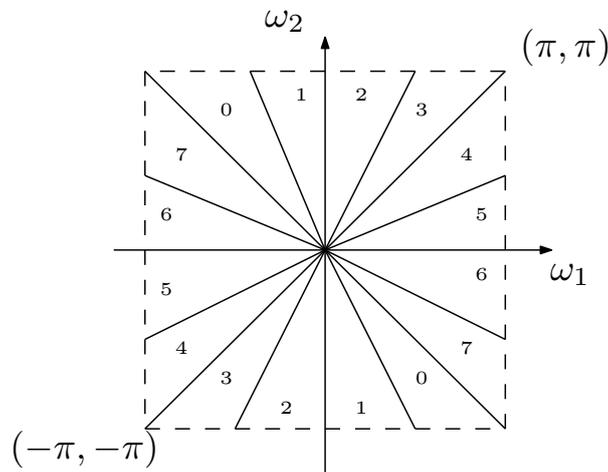


Figura 3.6: Partição frequencial do DFB com  $l = 3$ , com isso, tem-se  $2^3 = 8$  subbandas frequências.

### 3.1.3 Transformada *Wavelet* não-decimada

Em meio aos métodos multirresolução citados, foram listadas algumas características interessantes à fusão de imagens. Como dito, a esparsidade é um ponto interessante. A esparsidade faz com que cada coeficiente da decomposição tenha bastante energia e com isso bastante importância na reconstrução.

Por exemplo, a esparsidade alcançada com a DWT pode trazer desvantagens para a reconstrução com perdas. Após a modificação de poucos coeficientes de uma dada imagem, a reconstrução pode ser comprometida pelo grande número de artefatos criados [53].

Com o intuito de evitar estes artefatos, pode-se abrir mão da esparsidade. Ou seja, pular o passo de subamostragem, gerando mais coeficientes e assim evitar os artefatos na fusão.

Após esta breve motivação, podemos explicar a decomposição proposta pela Transformada *Wavelet* não-decimada (UWT – *Undecimated Wavelet Transform*). Igualmente à DWT, começamos com um sinal unidimensional (1D)  $I[n]$ . Seguindo a descrição multirresolução, a decomposição desse sinal é dada por  $\mathbf{y} = \{y^1, y^2, \dots, y^J, x^J\}$ . Novamente, os coeficientes  $y^j$  são chamados coeficientes de detalhes enquanto  $x_j$  são chamados coeficientes de aproximação.

À luz da definição da DWT, temos que cada resolução pode ser escrita como [53]:

$$x_{j+1}[l] = (\bar{h}^{(j)} * x_j)[l] = \sum_k h[k]x_j[l + 2^j k], \quad (3.11)$$

$$y_{j+1}[l] = (\bar{g}^{(j)} * x_j)[l] = \sum_k g[k]x_j[l + 2^j k], \quad (3.12)$$

onde  $\bar{h}[l] = h[-l]$  e  $h^{(j)}[l] = h[l]$  se  $l/2^j$  é inteiro ou 0 nos outros casos.

A reconstrução é obtida através da seguinte expressão:

$$x_j[l] = \frac{1}{2} \left[ (\tilde{h}^{(j)} * x_{j+1})[l] + (\tilde{g}^{(j)} * y_{j+1})[l] \right], \quad (3.13)$$

onde  $\tilde{h}$  e  $\tilde{g}$  são as versões interpoladas dos filtros passa-baixas e passa-altas respectivamente. Repete-se a operação descrita na Equação 3.13 até que  $j = 0$ . E com isso, obtêm-se  $I[n]$ , pois  $x_0 = I$ .

Novamente, deve-se garantir a reconstrução perfeita. Portanto, como condição necessária para isso, temos que:

$$H(z^{-1})\tilde{H}(z) + G(z^{-1})\tilde{G}(z) = 1. \quad (3.14)$$

Pode-se estender o algoritmo para a versão bi-dimensional [53]:

$$x_{j+1}[k, l] = (\bar{h}^{(j)}\bar{h}^{(j)} * x_j)[k, l], \quad (3.15)$$

$$y_{j+1}^1[k, l] = (\bar{g}^{(j)}\bar{h}^{(j)} * x_j)[k, l], \quad (3.16)$$

$$y_{j+1}^2[k, l] = (\bar{h}^{(j)}\bar{g}^{(j)} * x_j)[k, l], \quad (3.17)$$

$$y_{j+1}^3[k, l] = (\bar{g}^{(j)}\bar{g}^{(j)} * x_j)[k, l]. \quad (3.18)$$

## Fatoração Espectral

A UWT produz uma maior redundância entre os coeficientes em comparação com as transformações estudadas até agora. O efeito causado é o espalhamento da energia quando são utilizados filtros longos. Essa redundância é muito interessante para evitar os artefatos de fusão da reconstrução. Porém, quando os filtros são muito longos, há um excesso desse espalhamento.

Ellmauthaler *et al.* [42] mostraram que é possível contornar este problema e

assegurar este compromisso entre suporte de filtro e espalhamento de energia através da fatoração espectral.

Essencialmente, a Fatoração Espectral (SF) descreve os filtros como:

$$H(z) = (1 + z^{-1})P(z) \quad (3.19)$$

$$G(z) = (1 - z^{-1})Q(z) \quad (3.20)$$

Ainda pode-se descrever para outros níveis de decomposição:

$$H(z^{2^{j-1}}) = (1 + z^{-2^{j-1}})P(z^{2^{j-1}}) \quad (3.21)$$

$$G(z^{2^{j-1}}) = (1 - z^{-2^{j-1}})Q(z^{2^{j-1}}) \quad (3.22)$$

Comumente, a informação saliente é dita como imagem de detalhes (advinda do filtro passa-altas). A fatoração espectral nos permite realizar a fusão destas imagens antes do espalhamento de energia causado por  $Q(z)$ .

Um diagrama de como a fusão de imagens é obtida através da Transformada *Wavelet* não-decimada com Fatoração Espectral (UWT-SF) é mostrado na Figura 3.7.

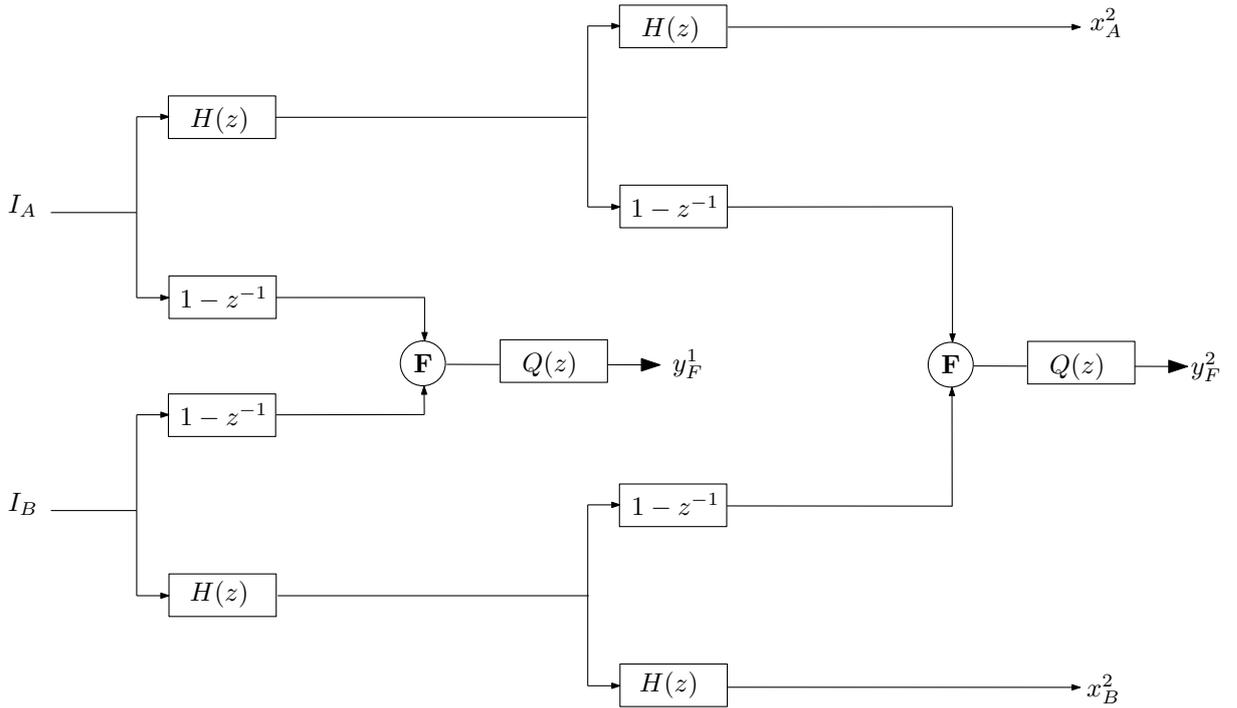


Figura 3.7: UWT-SF usando dois níveis de decomposição. Após aplicar o primeiro nível de decomposição, os detalhes da imagem fundida são extraídos da fatoração espectral e da regra de fusão  $\mathbf{F}$ . Esta representação denota uma decomposição unidimensional, porém a decomposição n-dimensional pode ser obtida através da aplicação desse esquema em cada dimensão.

Nos nossos testes, será usado a UWT-SF para realizar a fusão. Pois, na literatura [12, 42, 54] foi mostrado que a fatoração espectral é uma ferramenta importante

para melhorar os resultados da fusão.

## 3.2 Regras de Fusão

Após as transformadas multirresolução serem apresentadas neste Capítulo, devemos prosseguir no diagrama da Figura 3.2 e estudar as regras de fusão.

Pode-se dividir este bloco de regra de fusão em 3 passos [47]:

1. Combinação:

A combinação é obtida através de uma análise feita por regiões. Ao segmentar as imagens de entrada do algoritmo de fusão, deve-se fazer correspondência entre cada imagem. A região  $\mathcal{R}$  de interesse, pode ser, no caso de fusão ao nível do pixel, o próprio pixel, um bloco de pixels, etc.

2. Medida de atividade:

A medida de atividade  $s$  representa o nível de energia (ou alguma outra medida) de cada região  $\mathcal{R}$ . Em outras palavras, esta medida tem por objetivo calcular o nível de “importância” de cada região localmente.

3. Mapa de decisão:

Após as equivalências no passo de combinação e a mensuração de atividade local de cada imagem, pode-se definir um mapa de decisão. Ou seja, decidir qual pixel/coeficiente/escala de cada imagem estará presente na imagem/vídeo fundido.

Podemos citar exemplos de regras de fusão tais quais [14]:

1. *CM selection (choose max selection)*

A regra *CM selection* é baseada na escolha do valor máximo absoluto entre as entradas. Sendo  $y_A^j[m, n]$  e  $y_B^j[m, n]$  os pixels da entrada  $A$  e  $B$  da resolução  $j$ , respectivamente, os pesos  $w_A^j$  e  $w_B^j$  que formam a imagem fundida são dados por:

$$w_A^j[m, n] = \begin{cases} 1 & \text{if } |y_A^j[m, n]| > |y_B^j[m, n]| \\ 0 & \text{caso contrário.} \end{cases} \quad (3.23)$$

$$w_B^j[m, n] = 1 - w_A^j[m, n]. \quad (3.24)$$

2. *CBA selection (Coefficient-based activity level measurement selection)*

Aprofundando-se mais na informação intrínseca de cada quadro, a regra *CBA*

*selection* calcula os pesos  $w_A^j$  e  $w_B^j$  referentes a cada pixel de uma dada resolução de modo que:

$$w_A^j[m, n] = \begin{cases} 1 & \text{if } \alpha_A^j[m, n] > \alpha_B^j[m, n] \\ 0 & \text{caso contrário.} \end{cases} \quad (3.25)$$

$$w_B^j[m, n] = 1 - w_A^j[m, n]. \quad (3.26)$$

sendo  $\alpha_A^j[m, n]$  e  $\alpha_B^j[m, n]$  uma medida de atividade aplicada às imagens  $A$  e  $B$ . Esta medida de atividade é calculada para cada pixel.

### 3. *RBA selection (Region-based activity measurement selection)*

Diferente da regra anterior, esta usa a medida de atividade em uma dada região  $\mathcal{R}$ . Após isso, os pesos podem ser tomados como mostrados na regra anterior.

### 4. *WBA selection (Window-based activity measurement selection)*

Restringindo ainda mais a região, pode-se tomar a medida de atividade por janelas. Normalmente, são tomadas janelas deslizantes de tamanho 5x5 ou 7x7 pixels. Após isso, os pesos são calculados como na regra *CBA selection*.

Uma maneira natural de realizar a média ponderada entre duas representações pode ser obtida através da Análise de Componentes Principais (PCA) onde os pesos de cada entrada são dados pelos autovalores da matriz de autocovariância.

Enfim, existem inúmeras regras de fusão cuja finalidade seja obter o melhor quadro fundido possível. Com o intuito de manter a mesma regra para todos os exemplos nesta dissertação, será usada a regra *meanmax*. Esta regra é amplamente utilizada na literatura [14, 42, 47] e é constituída por uma regra para cada sub-banda. Para as sub-bandas passa-altas, a regra utilizada é a *CM selection* (Equações 3.23 e 3.24). Para as sub-bandas passa-baixas, a regra utilizada é a média das entradas, ou seja, utilizar pesos iguais a  $w_A = w_B = 1/2$ .

## 3.3 Banco de Dados de Vídeos para Fusão

Infelizmente, existem pouquíssimos bancos de dados disponíveis para testes de algoritmos de fusão de vídeos. Em particular, o interesse desta dissertação é em fusão de vídeos das faixas visível e do infravermelho, para vigilância e aplicações militares, o que resulta em um número de bases ainda mais reduzido.

Pelo número reduzido, pode-se listar os bancos de dados disponíveis:

1. OTCBVS: *Object Tracking and Classification in and Beyond the Visible Spectrum* [55].

Este banco de dados é organizado pela Universidade do Estado de Oklahoma e composto por 12 diferentes conjuntos de vídeos provenientes de vários pesquisadores em visão computacional. Dentre todos os conjuntos, podemos mencionar:

- Dataset 01: *OSU Thermal Pedestrian Database*
  - Interesse: Detecção de pedestres através de imagens termais.
  - 10 sequências (284 imagens), apenas imagens termais, 30 quadros por segundo, tamanho das imagens de  $360 \times 240$ .
- Dataset 02: *IRIS Thermal/Visible Face Database*
  - Interesse: registros simultâneos de imagens termal e visível de faces sob iluminação, posições e expressões variáveis.
  - 4228 pares de imagens, tamanho das imagens de  $320 \times 240$ .
- Dataset 03: *OSU Color-Thermal Database*
  - Interesse: fusão de imagem colorida e termal.
  - 6 sequências (17089 imagens), 30 fps, tamanho das imagens de  $320 \times 240$ .

2. *Eden Project Multi-sensor dataset* [56].

Este banco de dados disponibiliza vídeos de diferentes sensores com o intuito de rastrear alvos. Os vídeos disponibilizados foram manualmente registrados. É composto por apenas um par de vídeos infravermelho e visível e três versões fundidas.

3. *LITIV (Laboratoire D'Interprétation et de Traitement D'Images Et Vidéo)* [57].

O tópico de interesse deste banco é rastreamento de pessoas e objetos. Este contém nove sequências de vídeos infravermelho/visível que variam entre 11 e 88 segundos de duração e ainda as máscaras de segmentação plano de cena/plano de fundo para cada sequência.

4. *VLIRVDIF: (Visible-Light and Infrared Video Database for Image Fusion)* [20].

Este banco de dados foi criado com diferentes condições ambientais, em diferentes lugares e com conteúdo variado. No total são 24 sequências (aproximadamente 3.8Gb de dados) variando entre 15 e 320 segundos, gravadas a 30 quadros por segundo. As imagens tem tamanho  $720 \times 480$ .

Como a proposta desta dissertação é acompanhar o efeito da adição da informação temporal aos algoritmos de fusão, iremos utilizar o banco de dados VLIRVDIF, por conter mais sequências e mais variação temporal. Por isso, iremos descrever o conjunto de dados utilizado para os testes a seguir.

### 3.3.1 Descrição do conjunto de teste

O banco de dados VLIRVDIF é descrito em detalhes em [58]. Porém, foi selecionado um conjunto menor desta base para testes. Com o intuito de ilustrar o conteúdo deste banco, a Figura 3.8 oferece alguns *thumbnails* da base de vídeos.

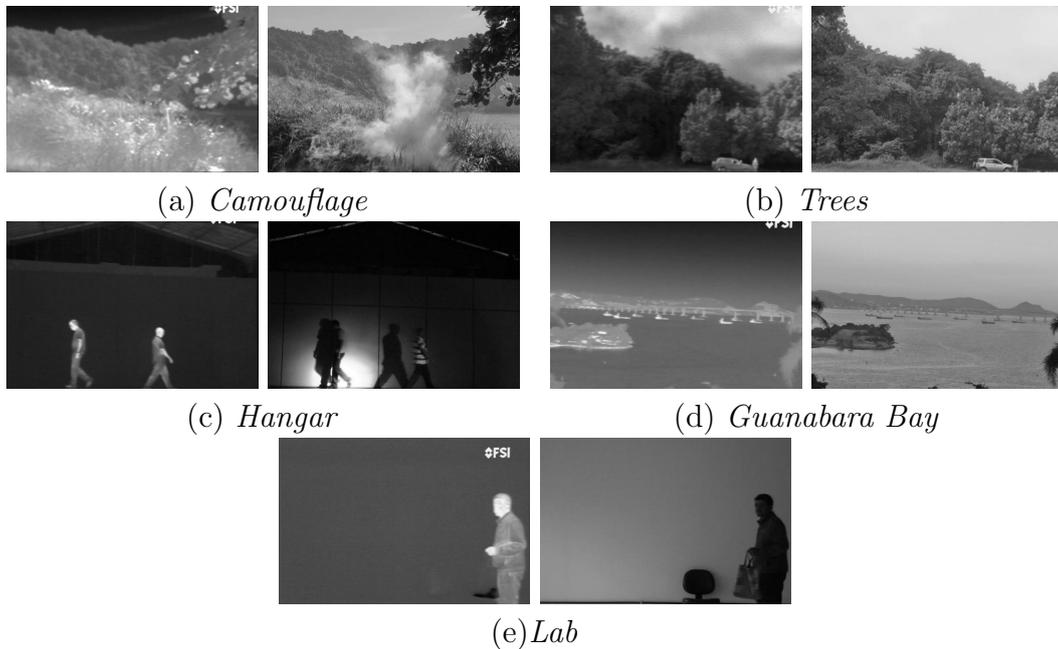


Figura 3.8: *Thumbnail* do banco de vídeos.

Os vídeos serão limitados em um comprimento de 10 segundos para a realização dos testes. Este corte foi feito nas áreas com a presença de algum movimento marcante. Seguindo a enumeração dos quadros e das sequências disponíveis na *webpage* do banco, limitamos o conjunto de testes segundo a Tabela 3.1.

Sequência	Tomada	Primeiro quadro	Último quadro
<i>Camouflage</i>	3	1730	1930
<i>Trees</i>	3	8400	8700
<i>Hangar</i>	1	3500	3800
<i>Guanabara Bay</i>	3	2000	2300
<i>Lab</i>	3	1230	1530

Tabela 3.1: Descrição da parte do banco usado neste trabalho.

Esse conjunto de testes também está disponível para download em [www.smt.ufrj.br/fusion/test](http://www.smt.ufrj.br/fusion/test). Com este conjunto é possível testar diferentes cenários e condições. Pela variabilidade das distâncias das cenas ao plano da câmera, também é possível acompanhar o efeito das diferentes condições de registro nos vídeos fundidos.

## 3.4 Comparação do Desempenho dos Algoritmos de Fusão Espacial

Ao longo deste Capítulo, foram apresentadas diversas transformações multirresolução. Estas foram selecionadas de acordo com o uso e os resultados obtidos na literatura. Primeiramente, foi apresentada a DWT (Seção 3.1.1) por seu amplo e vasto uso em aplicações de sistemas multirresolução. A DWT, assim como os outros métodos multirresolução, permite a análise tanto frequencialmente quanto espacialmente com uma dada acurácia. Isso facilita a seleção das características importantes para a fusão. Porém, em relação à resolução espacial, apenas os contornos horizontais, verticais e diagonais são representados diretamente pelas bases.

Para contornar isto, é sugerido o uso de uma transformada anisotrópica. A ConT foi introduzida e explicada na Seção 3.1.2. Por conta do esquema de filtragem em dois passos, as bordas de imagens naturais podem ser melhor representadas e mais direções espaciais são contempladas.

Seguindo uma outra linha, foi apresentada a UWT-SF. Apesar de uma característica interessante a sistemas multirresolução ser a esparsidade, a UWT propõe uma maior redundância, que evita artefatos no momento da fusão. Para diminuir um pouco a redundância e selecionar as informações salientes de maneira eficaz, foi apresentada a fatoração espectral.

Para a comparação entre todos os métodos mostrados, serão utilizadas as métricas mostradas no Capítulo 2.

É sabido que a DWT apresenta resultados de fusão ruins. Mas ela geralmente é utilizada como comparação de algoritmos de fusão [2, 3, 23, 35, 59, 60]. Iremos utilizar também outras duas técnicas de fusão simples para comparação. São elas a média quadro-a-quadro e a Análise de Componentes Principais (PCA).

### 3.4.1 Métodos comparativos

Para comparar os métodos multirresolução, serão utilizados os dois métodos citados anteriormente. Estas técnicas são usadas, *a priori*, para estabelecer um *benchmark* para os algoritmos de fusão [2, 61].

Por conta da simplicidade, não iremos descrever os algoritmos em detalhes. Entretanto, estes algoritmos foram implementados e os resultados estão a seguir.

#### Fusão por média

Apesar da fusão por média se constituir como um método de pior desempenho, ela será empregada neste trabalho para criar um patamar inferior de qualidade. Sabe-se que, ao realizar a média entre cada quadro, a informação frequencial não

é contemplada. Ou seja, usando a decomposição multirresolução, podemos pensar que apenas o nível  $x^0$  é analisado no algoritmo.

Mais ainda, pelo que foi dito até agora, a média não é uma escolha boa. A explicação desta afirmativa é a motivação por optarmos por métodos multirresolução. Ao aplicarmos a média, as características de estrutura, o compromisso entre resolução frequencial e espacial e a esparsidade da representação são comprometidos.

Os resultados usando esta técnica estão disponíveis na Tabela 3.2.

Fusão através da Média				
Cena	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
<i>Camouflage</i>	0.495564	0.160289	0.421598	0.489455
<i>Guanabara Bay</i>	0.389578	0.250803	0.405866	0.372612
<i>Trees</i>	0.455419	0.364000	0.423686	0.291000
<i>Hangar</i>	0.446791	0.317596	0.400587	0.611836
<i>Lab</i>	0.462739	0.341284	0.415001	0.490310

Tabela 3.2: Resultado da fusão usando média.

Estes resultados garantem um patamar inferior para os métodos que analisaremos a seguir.

### Análise de Componentes Principais

A Análise de Componente Principais (PCA) é um método comumente utilizado como comparativo a qualidade da fusão [61]. Pode-se entendê-lo como uma ligeira modificação do algoritmo de fusão usando média. Computa-se a média ponderada entre os quadros de entrada para se obter o quadro fundido, onde os pesos provém da PCA. Estes são definidos como os autovalores dos autovetores de maior energia. Apesar de denotarmos o método como PCA, na verdade, será usada apenas a componente principal de maior energia.

Pode-se, simplesmente, escrever o quadro fundido  $I_F$  como:

$$I_F = \alpha_1 I_A + \alpha_2 I_B \quad (3.27)$$

onde  $\alpha_1$  e  $\alpha_2$  ( $0 \leq \alpha_1, \alpha_2 \leq 1$ ) são os autovalores da matriz de covariância das entradas  $I_A$  e  $I_B$ , respectivamente.

Os resultados usando esta técnica estão dispostos na Tabela 3.3

### 3.4.2 Transformada *Wavelet* Discreta

O método de fusão usando a transformada *wavelet* discreta é amplamente difundido na literatura desta área. Por esse motivo, há inúmeros bancos de filtros usados para

Fusão a partir do PCA				
Cena	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
<i>Camouflage</i>	0.476697	0.189160	0.405063	0.398807
<i>Guanabara Bay</i>	0.497338	0.504580	0.413030	0.384893
<i>Trees</i>	0.468894	0.368813	0.422938	0.505397
<i>Hangar</i>	0.591775	0.612177	0.426687	1.641015
<i>Lab</i>	0.425543	0.462410	0.345713	0.581341

Tabela 3.3: Principal component analysis fusion results.

implementar a DWT. Entre estes, Daubechies [39] propôs uma família de filtros (que recebe seu nome) que abordaremos primeiramente. Além desta família de filtros, será utilizada a família de filtros biortogonais [43] por ser tão difundida quanto a anterior.

Usando a notação usada no programa Matlab [62], neste teste usaremos os seguintes filtros:

1. Filtros de Daubechies

- db1, db2, db3, db4, db5

2. Filtros Biortogonais

- bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8

O banco de filtros bior1.1 e db1 são iguais. A partir daqui, nos referimos a este banco de filtros como banco de Haar, nome pelo qual ele é amplamente conhecido.

Para prosseguir, tem-se que definir o número de níveis de decomposição. Realizando alguns testes, notamos que os resultados pioram para números de níveis elevados. Por isso, os resultados serão restritos para um número de níveis entre 2 e 4.

Com 3 níveis de decomposição para cada um dos 14 bancos de filtros testados, totalizamos 42 vídeos fundidos para cada uma das 5 sequências. Como o número de resultados é elevado, serão apresentados nas Tabelas 3.4, 3.5, 3.6, 3.7 e 3.8 os relativos aos bancos de filtros com melhores resultados.

Iniciando pela cena *Camouflage*, foram obtidos os resultados presentes na Tabela 3.4. Em comparação com as Tabelas 3.2 e 3.3, esta cena demonstra a superioridade dos métodos de fusão multirresolução à fusão usando a média (Tabela 3.4). As bordas espaciais e temporais obtiveram melhor representação pela DWT, como pode ser observado no ganho das métricas  $Q_{AB/F}$  e  $DQ$ . A avaliação desta cena fundida usando a métrica  $M_{AB/F}$  apresentou resultados inferiores aos resultados obtidos

Camouflage					
Filtro	Nível	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
bior2.2	2	<b>0.501458</b>	0.157273	<b>0.422414</b>	0.735748
bior2.4	2	0.498073	<b>0.157522</b>	0.421873	0.742461
bior6.8	2	0.495994	0.157300	0.421624	<b>0.746381</b>

Tabela 3.4: Resultados da cena *Camouflage* usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

usando a fusão por média. Uma possível causa para esta má avaliação é a presença de artefatos de fusão na reconstrução dos quadros fundidos usando a DWT.

Guanabara Bay					
Filtro	Nível	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	2	0.441072	<b>0.235921</b>	0.414022	0.616701
Haar	3	<b>0.465892</b>	0.222232	<b>0.416807</b>	0.666392
bior2.4	3	0.433094	0.227857	0.412988	<b>0.677884</b>

Tabela 3.5: Resultados da cena *Guanabara Bay* usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

A cena *Guanabara Bay* corrobora o comportamento apresentado anteriormente. Apenas a métrica  $M_{AB/F}$  apresentou resultados (Tabela 3.5) inferiores comparativamente aos mostrados nas Tabelas 3.2 e 3.3.

Trees					
Filtro	Nível	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	4	<b>0.454092</b>	0.333062	0.426256	0.746475
bior2.2	3	0.451476	<b>0.348190</b>	<b>0.426345</b>	0.734005
bior6.8	3	0.448964	0.348106	0.426022	<b>0.757732</b>

Tabela 3.6: Resultados da cena *Trees* usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

No caso da cena *Trees*, os artefatos de fusão causaram um grande impacto nos resultados das métricas  $Q_{AB/F}$  e  $M_{AB/F}$ . Pelo fato da cena *Trees* conter muitos detalhes de alta frequência espacial e temporal, seus artefatos de fusão se mesclaram aos detalhes da cena, degradando os resultados (Tabela 3.6) em comparação aos obtidos anteriormente. Ainda assim, os gradientes temporais mensurados na métrica  $DQ$  foram melhor preservados ao usar a fusão usando a DWT.

A cena *Hangar* apresenta uma característica que não favorece os métodos de fusão multirresolução que é a oclusão de parte da cena. Esta cena contém alguns lugares iluminados (presentes na cena visível) e alguns lugares em completa escuridão

Hangar					
Filtro	Nível	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	4	<b>0.526536</b>	0.262136	<b>0.420056</b>	<b>0.984964</b>
db5	2	0.505414	0.300974	0.414737	0.844421
bior6.8	2	0.502271	<b>0.302012</b>	0.414264	0.834774

Tabela 3.7: Resultados da cena *Hangar* usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

(visíveis apenas na sequência infravermelha). Por essa razão, a DWT apresenta resultados ruins (Tabela 3.7) para as 4 métricas testadas. Os resultados presentes na Tabela 3.3 são superiores neste tipo de cena.

Lab					
Filtro	Nível	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
bior2.2	3	<b>0.546222</b>	0.307164	<b>0.425692</b>	0.762839
Haar	2	0.529817	<b>0.322804</b>	0.423935	0.640996
Haar	3	0.543948	0.314096	0.425516	<b>0.778578</b>

Tabela 3.8: Resultados da cena *Lab* usando DWT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Finalizando, a cena *Lab* apresenta resultados (Tabela 3.8) semelhantes aos resultados da cena *Camouflage* (Tabela 3.4). Apenas a métrica  $M_{AB/F}$  é prejudicada usando o método de fusão usando a DWT, por conta dos motivos já citados.

### 3.4.3 Transformada *Contourlet*

O algoritmo de fusão usando a decomposição através das Transformada *Contourlet* foi computado utilizando os filtros: Le Gall 5-3, Cohen-Daubechies-Feauveau 9-7 [63], Haar [39, 64], pkva, pkva6, pkva8 e pkva12 [65]. O número de direções foi escolhido segundo o trabalho de Li *et al.* [5].

Novamente, pelo número elevado de resultados obtidos, serão apresentados apenas os melhores (para cada métrica) nas Tabelas 3.9, 3.4.3, 3.11, 3.12 e 3.13.

Na Tabela 3.9, são mostrados os resultados obtidos para a cena *Camouflage*. A ConT atinge um ganho para a métrica  $Q_{AB/F}$  de aproximadamente 3% em relação a média, de 7% em relação ao PCA e de 2% para a DWT. As demais métricas indicam uma piora na qualidade deste método. Um fato interessante pode ser observado relativamente a Tabela 3.4, há ganho de aproximadamente 2% na métrica  $M_{AB/F}$ , podendo ser interpretado como uma diminuição da quantidade de artefatos de fusão.

Computando a fusão da cena *Guanabara Bay*, é corroborado o ganho da ConT para a métrica  $M_{AB/F}$  em relação à DWT. Os demais resultados ainda se apresentam

Camouflage						
PFilt	DFilt	Direction	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
5-3	9-7	[1 2]	<b>0.510536</b>	0.158178	0.421489	0.322284
Haar	9-7	[1 2 4]	0.506811	<b>0.160552</b>	0.422781	0.325912
Haar	9-7	[1]	0.507328	0.160549	<b>0.423113</b>	0.347203
9-7	9-7	[1 2]	0.510009	0.157795	0.422034	<b>0.359402</b>

Tabela 3.9: Resultados da cena *Camouflage* usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Guanabara Bay						
PFilt	DFilt	Direction	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	9-7	[1 2 4]	<b>0.451453</b>	0.269412	0.397061	0.243154
Haar	9-7	[1]	0.448256	<b>0.270686</b>	0.395655	0.236709
9-7	9-7	[1 2 4]	0.440738	0.256923	<b>0.411206</b>	0.286070
5-3	9-7	[1 2 4]	0.441270	0.257713	0.410745	<b>0.310173</b>

Tabela 3.10: Resultados da cena *Guanabara Bay* usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

piores do que os mostrados anteriormente como pode ser visto na Tabela .

Trees						
PFilt	DFilt	Direction	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	9-7	[1]	<b>0.477809</b>	0.368932	0.423666	<b>0.385882</b>
Haar	5-3	[1]	0.477679	<b>0.368957</b>	0.423670	0.383957
Haar	pkva	[1 2 4]	0.474423	0.366953	<b>0.424249</b>	0.284298

Tabela 3.11: Resultados da cena *Trees* usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Ao analisar os resultados obtidos pela cena *Trees* usando o método de fusão usando a DWT (Tabela 3.6), foi visto que a presença de muitos artefatos de fusão prejudicaram a preservação das bordas mensurada pela métrica  $Q_{AB/F}$ . Usando a ConT (Tabela 3.11), é possível atingir valores superiores aos obtidos usando a DWT tanto na métrica  $M_{AB/F}$  quanto na métrica  $Q_{AB/F}$ , concluindo então que a Transformada *Contourlet* mitigou os artefatos e preservou melhor os detalhes de alta frequência espacial. Uma exceção é o PCA, que na Tabela 3.3, consegue resultado superior para a métrica  $M_{AB/F}$ .

Como dito na Subseção anterior, as oclusões presentes na cena *Hangar* fazem com que os métodos multirresolução obtenham resultados ruins comparativamente aos outros. Comparando apenas com a fusão usando a DWT (Tabela 3.7), o método de fusão usando a ConT foi superior para as métricas  $M_{AB/F}$  e  $DQ$ .

Hangar						
PFilt	DFilt	Direction	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	9-7	[1]	<b>0.524392</b>	<b>0.363756</b>	<b>0.427706</b>	<b>0.816420</b>

Tabela 3.12: Resultados da cena *Hangar* usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Lab						
PFilt	DFilt	Direction	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar	pkva	[1]	<b>0.543239</b>	0.354459	0.422320	0.389580
pkva	5-3	[1]	0.535998	<b>0.356818</b>	0.420021	0.308165
Haar	5-3	[1 2 4]	0.542644	0.353046	<b>0.422645</b>	0.458911
5-3	5-3	[1 2]	0.537103	0.354757	0.420819	<b>0.465140</b>

Tabela 3.13: Resultados da cena *Lab* usando ConT. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Concluindo a etapa de testes com a transformada *Contoulet*, a cena *Lab* foi testada (Tabela 3.13). Um ganho de qualidade de 10% foi atingido segundo a métrica  $M_{AB/F}$ . Porém, para as demais métricas, não foram obtidos ganhos representativos.

### 3.4.4 Transformada *Wavelet* não-decimada com Fatoração Espectral

Das transformadas citadas, a UWT-SF possui resultados relevantes na literatura [12]. Em relação ao escopo do teste feito, serão utilizados os bancos de filtros: Le Gall 5 – 3 [63],  $B_1$ -spline,  $B_3$ -spline [53] (e as variações isotropic 1 e isotropic 2), near-symmetric 5/7 (‘near\_sym\_a’), 13/19-tap (‘near\_sym\_b’) [66], Haar Starck 1 and Haar Starck 2 [53]. Seguindo os experimentos disponíveis em [12], serão utilizados 4 níveis de decomposição.

Os melhores resultados estão dispostos nas Tabelas 3.14 a 3.18.

Camouflage				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	<b>0.518077</b>	0.148605	<b>0.424082</b>	0.980751
Le Gall 5-3	0.515244	<b>0.149637</b>	0.424061	0.941795
$B_3$ -spline	0.51595	0.145741	0.423470	<b>0.995856</b>

Tabela 3.14: Resultados da cena *Camouflage* usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Começando a análise dos resultados da UWT, é possível notar que a UWT-SF

atinge resultados com avaliação subjetiva superior às demais testadas até aqui. Nesta cena, apenas a métrica  $M_{AB/F}$  obteve resultado inferior aos resultados obtidos pelo método de fusão usando a ConT (Tabela 3.9). Um fato importante a se destacar é que os valores obtidos pela métrica MI-IFD são superiores aos obtidos pelos demais métodos de fusão.

Guanabara Bay				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
$B_3$ -spline	<b>0.460692</b>	<b>0.234711</b>	<b>0.416152</b>	0.866403
Haar Starck 2	0.457654	0.232567	0.416040	<b>0.869633</b>

Tabela 3.15: Resultados da cena *Guanabara Bay* usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Dando continuidade à análise, os resultados da fusão da cena *Guanabara Bay* estão dispostos na Tabela 3.15. Para esta cena, as métricas  $Q_{AB/F}$  e MI-IFD obtiveram os melhores resultados.

Trees				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
$B_3$ -spline	<b>0.469360</b>	0.344809	<b>0.429091</b>	<b>0.971692</b>
Le Gall 5-3	0.465615	<b>0.351628</b>	0.428654	0.955617

Tabela 3.16: Resultados da cena *Trees* usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Corroborando o comportamento dos resultados, o MI-IFD ainda se apresenta melhor na cena *Trees*. Há ganhos ainda nas demais métricas em comparação aos resultados obtidos pela fusão usando a DWT. Em comparação com a fusão usando a ConT, na Tabela 3.11, a UWT-SF obtém resultados ligeiramente inferiores para as métricas  $Q_{AB/F}$  e  $M_{AB/F}$ .

Hangar				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
$B_3$ -spline	<b>0.539683</b>	0.265776	<b>0.424366</b>	<b>1.169460</b>
Le Gall 5-3	0.533475	<b>0.271628</b>	0.422788	1.115678

Tabela 3.17: Resultados da cena *Hangar* usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Para a cena *Hangar*, cujo resultados estão dispostos na Tabela 3.17, há ganhos nas métricas  $Q_{AB/F}$  e na métrica MI-IFD. Há ligeira perda (cerca de 1%) de desempenho para a métrica DQ em comparação com o resultado obtido pela fusão usando

a ConT (Tabela 3.12). Porém, o resultado obtido pela PCA ainda é superior aos resultados obtidos pela UWT-SF.

Lab				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	<b>0.552258</b>	0.301067	<b>0.426866</b>	0.991825
Le Gall 5-3	0.549671	<b>0.302243</b>	0.426465	0.962717
$B_3$ -spline	0.548274	0.294486	0.426164	<b>0.992761</b>

Tabela 3.18: Resultados da cena *Lab* usando UWT-SF. Os resultados em negrito foram os melhores obtidos de acordo com cada métrica.

Concluindo os testes com a cena *Lab*, a UWT-SF obteve melhor avaliação para as métricas  $Q_{AB/F}$ ,  $DQ$  e  $MI-IFD$ . Com a análise das 5 cenas, é notada que, segundo os resultados obtidos pela métrica  $MI-IFD$ , há preferência pela fusão usando a UWT-SF.

### 3.5 Conclusões

Após os testes realizados com as diferentes transformadas multirresolução apresentadas neste Capítulo, foi possível analisar o patamar de qualidade de cada uma através das métricas apresentadas. A Tabela 3.19 sumariza os melhores resultados para cada métrica. É importante notar que, nesta Tabela, foram suprimidos os resultados da fusão usando o PCA pois esta análise tem por objetivo comparar os resultados de fusão usando transformadas multirresolução.

Melhores Resultados				
Método	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Camouflage				
UWT-SF Haar Starck 2	<b>0.518077</b>	0.148605	<b>0.424082</b>	0.980751
ConT Haar/9-7 [1 2 4]	0.506811	<b>0.160552</b>	0.422781	0.325912
UWT-SF $B_3$ -spline	0.515950	0.145741	0.423470	<b>0.995856</b>
Guanabara Bay				
DWT Haar lv 3	<b>0.465892</b>	0.222232	<b>0.416807</b>	0.666392
ConT Haar/9-7 [1]	0.448256	<b>0.270686</b>	0.395655	0.236709
UWT-SF $B_3$ -spline	0.460692	0.234711	0.416152	<b>0.866403</b>
Trees				
ConT Haar/9-7 [1]	<b>0.477809</b>	0.368932	0.423666	0.385882
Cont Haar/5-3 [1]	0.477679	<b>0.368957</b>	0.423670	0.383957
UWT-SF $B_3$ -spline	0.469360	0.344809	<b>0.429091</b>	<b>0.971692</b>
Hangar				
UWT-SF $B_3$ -spline	<b>0.539683</b>	0.265776	0.424366	<b>1.169460</b>
ConT Haar/9-7 [1]	0.524392	<b>0.363756</b>	<b>0.427706</b>	0.816420
Lab				
UWT-SF Haar Starck 2	<b>0.552258</b>	0.301067	<b>0.426866</b>	0.991825
ConT Pkva/5-3 [1]	0.535998	<b>0.356818</b>	0.420021	0.308165
UWT-SF $B_3$ -spline	0.548274	0.294486	0.426164	<b>0.992761</b>

Tabela 3.19: Resultado comparativo final dos métodos multirresolução utilizadas neste Capítulo.

Pode-se ainda notar que a UWT-SF foi superior às demais para as métricas  $Q_{AB/F}$  e para MI-IFD. Para a métrica  $M_{AB/F}$ , a Transformada *Contourlet* obteve melhores resultados.

Na Tabela 3.19 foram mostrados os melhores resultados para as técnicas multirresolução. Comparando-a à Tabela 3.3, nota-se um detalhe interessante. Os resultados obtidos para a sequência *Hangar* são muito díspares dos resultados mostrados aqui. A eficácia do método PCA nesta cena se dá pela quantidade de oclusões de cada entrada. Em outras palavras, as cenas podem ser combinadas utilizando processos

de segmentação/colagem pois há muitas áreas presentes em apenas uma entrada e ausentes na outra.

Por fim, neste Capítulo foi montado um *benchmark* eficaz para a comparação dos algoritmos propostos mais adiante. É importante notar que há uma representação melhor de acordo com a métrica. A eficácia desta medida de qualidade está diretamente ligada ao processo de validação, como visto no Capítulo 2. Não é possível definir nem um conjunto de parâmetros nem uma técnica capaz de maximizar todos eles conjuntamente.

# Capítulo 4

## Decomposição em Subbandas Temporais ao Longo da Trajetória do Movimento

No Capítulo 3 foram mostradas algumas técnicas para executar a fusão de vídeos usando a informação espacial. Em geral, o método usando a UWT-SF mostrou-se melhor que os outros nos vídeos testados.

Entretanto, esta técnica é usada quadro a quadro. Em outras palavras, a informação temporal não é levada em conta. A fim de considerar a dimensão temporal, pode-se apenas estender a decomposição multirresolução para mais uma dimensão, ou seja, utilizar transformadas multirresolução tridimensionais.

Também pode-se incluir a informação temporal aplicando os bancos de filtros apenas na dimensão temporal.

Outra opção é usar vetores de movimento para descrever o comportamento dos objetos em cena. Os vetores de movimento representam o deslocamento da posição de um objeto de um quadro a outro, ou seja, dado o vetor de movimento e um quadro de referência, pode-se estimar o quadro requerido. Detalhes deste uso serão vistos posteriormente.

Neste Capítulo iremos citar os trabalhos anteriores nesta área e formar um arcabouço teórico a fim de desenvolver algoritmos que incluam a descrição do movimento.

### 4.1 Trabalhos relacionados à fusão de vídeo

Nos últimos anos, houve muitos progressos na área de fusão de imagens [54]. Porém, para a fusão à aplicação em vídeos, existem poucos algoritmos e nenhum ainda consolidado. Por isso, é interessante entender as atuais propostas neste assunto.

Jian Li *et al.* [67] foram os primeiros a investigar a informação do movimento no

âmbito de fusão de vídeo. Neste trabalho, foi proposta uma regra estocástica para a utilização do movimento. Calculados os vetores de movimento, estes são usados para melhor detectar um alvo. Liu *et al.* [68] utiliza a mesma abordagem para segmentar o objeto de interesse e o plano de fundo. A aplicação direta à fusão não é abordada nestes artigos.

Qiang Zhang publicou muitos artigos agrupando a informação de movimento à espacial. Em 2012 [24], foi apresentado um trabalho utilizando a transformada tridimensional (3D) *surfacelet* para realizar a fusão do vídeo. Além deste trabalho, Zhang *et al.* [3, 15, 69] sugeriu a utilização de outras transformadas tridimensionais (*Curvelet* discreta uniforme, filtros 3D-Log-Gabor, *Surfacelet* tridimensional em conjunto com decomposição em valores singulares, respectivamente). Apesar de demonstrar a eficácia dos métodos propostos, os artigos não disponibilizam nem o banco de dados nem os códigos usados, dificultando a replicação dos seus resultados.

Ainda usando transformadas tridimensionais, Liang Xu *et al.* [70] utilizam a transformada 3D *Shearlet* para fundir os quadros do vídeo. Porém, em outro trabalho [59], a fusão de vídeo não é feita através de uma transformada 3D. Neste trabalho, é usada uma transformada 2D, como as aplicadas as imagens. A informação do movimento dos objetos é adicionada na escolha dos coeficientes. Esta técnica será explicada em mais detalhes no Capítulo 5.

## 4.2 Filtragem Temporal usando Compensação de Movimento

Desde o padrão H.261 [71], a redundância temporal é explorada na compressão de vídeos através dos passos de estimação e compensação de movimento. O processo de compensação de movimento tem por objetivo reduzir a energia da diferença entre dois quadros. Com essa ideia, pode-se codificar apenas os resíduos da diferença, reduzindo a quantidade de dados a ser enviada.

Antes de compensar o movimento, tem-se que estimá-lo. Idealmente, tem-se que segmentar todos os objetos da cena. Cada objeto tem seu movimento estimado por um vetor de movimento. Este vetor representa o deslocamento da posição de um objeto entre um quadro e outro. Como o processo de segmentação pode ser difícil, normalmente, segmenta-se o quadro em blocos quadrados tão pequenos quanto necessário. O compromisso entre tamanho do bloco/quantidade de dados tem que ser levado em conta nesta definição.

Na Figura 4.1 é ilustrado o processo de compensação de movimento. O blocos em destaque representam o mesmo objeto. Este se movimenta entre os quadros  $I_{t-1}$  até o quadro  $I_{t+1}$ . O movimento é descrito através dos vetores de movimento. Com

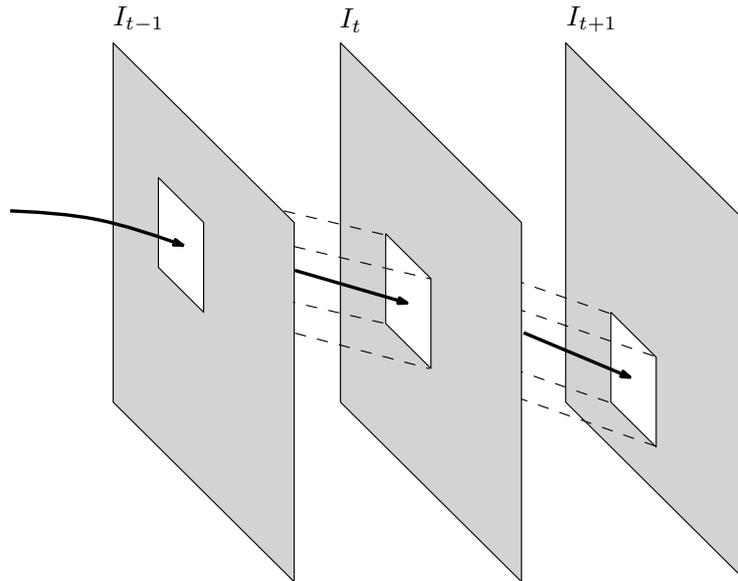


Figura 4.1: Predição do movimento através dos vetores de movimento.

estes e o bloco de referência  $I_t$  é possível estimar o objeto nos demais quadros.

Tendo os vetores de movimento, conhecemos a dinâmica dos objetos ao longo do vídeo. Este conhecimento é interessante para os métodos de fusão. Podemos preservar/unir a descrição do movimento ao criarmos o vídeo fundido.

Além deste tipo de abordagem para descrever o movimento, os avanços da tecnologia de compressão de vídeos trouxeram ferramentas úteis a fusão. Uma destas é a Codificação Escalável de Vídeo (SVC - *Scalable Video Coding*), que permite a codificação em diferentes taxas e resoluções. O ferramental apresentado no padrão H.264/MPEG4 [72] é a ponte necessária para a utilização da informação temporal nos algoritmos de fusão.

A Filtragem Temporal usando Compensação de Movimento (MCTF - *Motion-compensated temporal filtering*) é uma das técnicas em SVC e será a ligação entre a descrição do movimento e a fusão do vídeo.

Os algoritmos de codificação que utilizam MCTF são muito abordadas na literatura [73–76]. Em geral, a filtragem temporal compensada aumenta a compactação de energia da diferença entre quadros [77].

Pode-se ilustrar a MCTF através da Figura 4.2. Note que, após o processo de compensação de movimento, podemos encontrar a equivalência de cada pixel nos demais quadros. Neste exemplo foi usado o quadro no tempo  $t_0$  como referência. Esta ilustração representa um MCTF usando compensação *backward*, ou seja, o quadro compensado é anterior à referência. O vetor de movimento  $\mathbf{d} = (d_m, d_n)$  é responsável por esse deslocamento. Pode-se também calcular os vetores com uma referência anterior (compensação *backward*) ou com ambas as referências (compensação bilateral). Neste último caso, teríamos um par de vetores

de movimento [78].

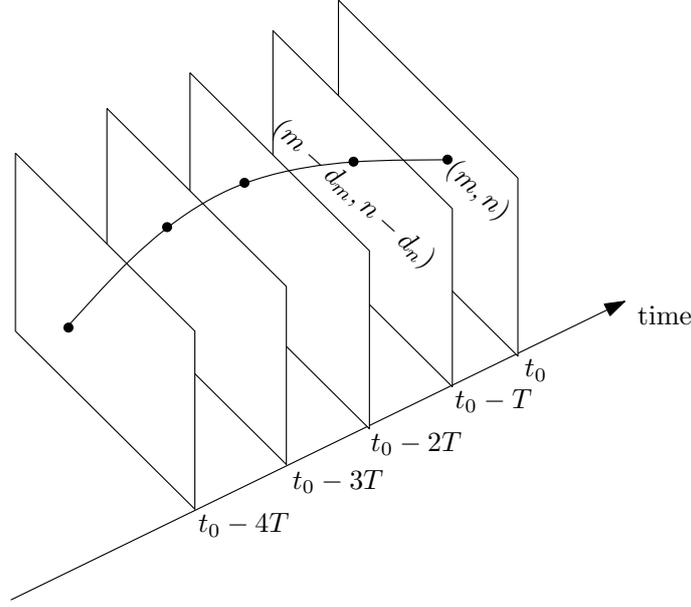


Figura 4.2: Ilustração da compensação de movimento de um pixel  $(m, n)$ .

Matematicamente, pode-se escrever o MCTF como [79]:

$$\begin{aligned}
 y(m, n, t) &= \sum_{k_1} \sum_{k_2} h(k_1, k_2, 0)x(n, m, t) \\
 &+ \sum_{k_1} \sum_{k_2} h(k_1, k_2, 1)x(n - d_m^1, m - d_n^1, t - 1) + \dots \\
 &+ \sum_{k_1} \sum_{k_2} h(k_1, k_2, T)x(n - d_m^T, m - d_n^T, t - T) \quad (4.1)
 \end{aligned}$$

onde  $y(m, n, t)$  é o pixel  $(m, n)$  no instante  $t$ , resultado da filtragem do sinal de entrada  $x(m, n)$  pelo filtro  $h(m, n)$ . Este filtro tridimensional suas dimensões definidas pelos limites dos contadores  $k_1$ ,  $k_2$  e, temporalmente, tem comprimento  $T$ . O deslocamento necessário para realizar a equivalência entre cada pixel do quadro  $i$  e o quadro anterior é dado pelo vetor de movimento  $\mathbf{d}^i = (d_m^i, d_n^i)$ .

Um outro aspecto importante ao MCTF é a relação entre os pixels da imagem de referência e da imagem compensada. Após a estimação de movimento (*forward* ou *backward*), os vetores de movimento categorizam os pixels da imagem compensada/de referência (Figura 4.3) em dois tipos:

1. Pixel Conectado: pixel que foi relacionado a outro na imagem de referência.
2. Pixel Não-conectado: pixel que não foi relacionado a nenhum outro na imagem de referência.

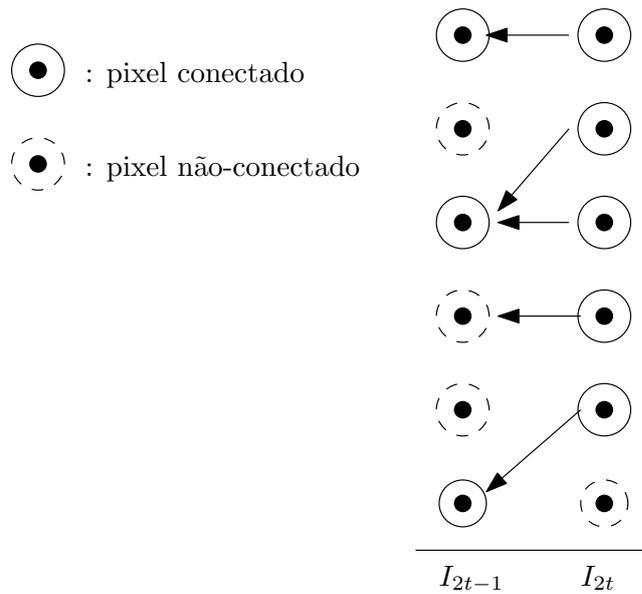


Figura 4.3: Compensação de movimento *backward*.

Este conceito de categorização dos pixels é fundamental para motivar a ferramenta necessária ao MCTF. Voltando ao objetivo inicial dessa técnica de filtragem, queremos usá-la na decomposição multirresolução do vídeo, executar um dado processamento para computar a fusão e então aplicar a transformada inversa, onde também será necessário o MCTF. Quando adicionamos a não-linearidade dada pela compensação de movimento, a inversa pode ser complicada ou não existir. Por exemplo, algum pixel  $(n - d_n, m - d_m)$  pode não existir na inversão e com isso, não há como recuperar um dado pixel categorizado como não-conectado. Na verdade, algumas pesquisas mostram que 3 – 5% dos pixels são categorizados desta maneira [75].

Na próxima Seção, será apresentada uma ferramenta para tratar a não-linearidade causada pela compensação de movimento. Com essa ferramenta, será possível encontrar uma transformada inversa de maneira mais simples do que analiticamente conseguiríamos através da Equação 4.1.

### 4.3 O Esquema *Lifting*

Como foi falado ao discutirmos a DWT, a transformada *wavelet* é uma ferramenta essencial em muitos campos da ciência. Um ponto importante nesta ferramenta é que a torna tão versátil é que ela pode ser implementada usando um banco de

filtros. Em casos em que há não-linearidade como no caso da MCTF, alguns métodos adaptativos podem ser úteis para a criação desses bancos de filtros mas não garantem a invertibilidade [80]. Favor notar que, quando foi incluída uma não-linearidade no MCTF, este ponto já foi comprometido. Ou seja, neste caso a invertibilidade das transformadas é um ponto complicado a ser tratado.

W. Sweldens [81] propôs um método capaz de contornar o problema da inversa (com certas limitações) e que garante reconstrução perfeita, também permitindo filtros variantes no tempo e *wavelets* não-lineares. O esquema *lifting* (LS - *Lifting Scheme*) resulta em muitas vantagens na decomposição multirresolução tais quais [82]:

- Existência da inversa: cada passo de *lifting* é reversível.
- Redução do custo computacional: o LS pode reduzir em até metade o custo computacional dada a transformada *wavelet* original.
- Existência do LS: Daubechies *et al.* [83] mostrou que toda transformada *wavelet* usando filtros com resposta ao impulso finita (FIR - *Finite Impulse Response*) pode ser decomposta em um número finito de passos de *lifting*.
- Extensão do sinal: O LS não necessita de extensão do sinal de entrada. As mesmas amostras da transformada *wavelet* inicial são utilizadas no LS.

Além destas vantagens, o LS tem inúmeras aplicações [82] tais quais *wavelets* com *grids* irregulares, filtros morfológicos, filtragem não-linear entre outros.

O esquema de *lifting* clássico (Figura 4.4) pode ser descrito nas seguintes partes:

1. Decomposição polifásica da entrada  $x[m]$ .

2. Passos do *Lifting*

(a) Passo de predição  $P$ :

$$d[n] = x[2n + 1] - P(x[2n]); \quad (4.2)$$

(b) Passo de atualização  $U$ :

$$c[n] = x[2n] + U(d[n]); \quad (4.3)$$

Acompanhando a Figura 4.4, pode-se destacar a funcionalidade de cada bloco e o significado de cada variável. Nesta Figura está representado um único passo de *lifting*. A primeira etapa é a decomposição polifásica do sinal de entrada  $x[m]$ , conhecido como *Lazy Wavelet* (em destaque na Figura 4.5). Esta etapa separa o

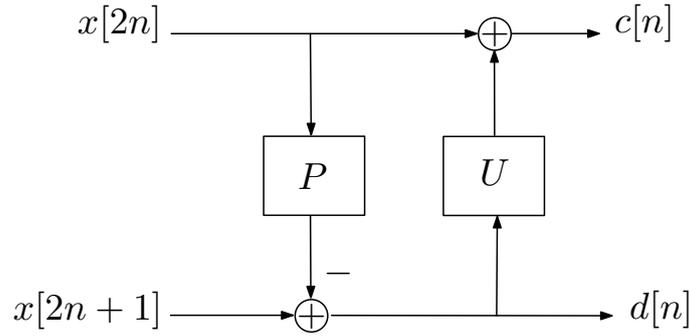


Figura 4.4: Esquema *Lifting* Clássico.

signal in two components  $x_o[n] = x[2n + 1]$  and  $x_e[n] = x[2n]$ , in other words, this decomposition separates the odd and even samples, respectively. After the decomposition of the input signal, the prediction step  $P$  results in the detail signal  $d[n]$ . Using this signal, the update step  $U$  results in the approximation signal  $c[n]$ .

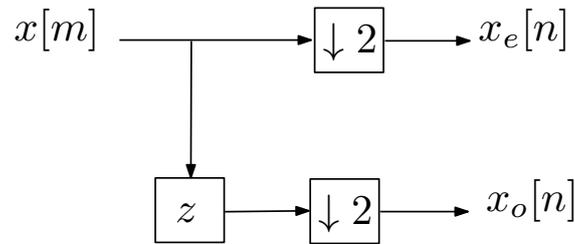


Figura 4.5: *Lazy Wavelet*.

Using the schemes described in Figures 4.4 and 4.6, we see that the process described results in perfect reconstruction for any steps  $P(\cdot)$  and  $U(\cdot)$ . Being a decomposition given in Equations 4.2 and 4.3, the reconstruction can be written as:

$$x[2n] = \underbrace{c[n]}_{x[2n]+U(d[n])} - U(d[n]); \quad (4.4)$$

$$x[2n + 1] = \underbrace{d[n]}_{x[2n+1]-P(x[2n])} + P(x[2n]); \quad (4.5)$$

In the linear case, the prediction and update steps are originated from the factorization of a given filter bank. The filters  $P(z)$  and  $U(z)$  are the center of the LS and defining them is the first step to start the decomposition using LS.

Pode-se escrever o LS da Figura 4.4 em forma matricial:

$$\mathbf{P}(z) = \begin{bmatrix} 1 & 0 \\ -P(z) & 1 \end{bmatrix}, \quad (4.6)$$

$$\mathbf{U}(z) = \begin{bmatrix} 1 & U(z) \\ 0 & 1 \end{bmatrix}. \quad (4.7)$$

Este tipo de descrição permite realizar a equivalência entre as transformadas e o LS correspondente. Ou seja, a forma matricial permite fazer equivalência entre os filtros usando uma dada transformada e o produto  $\mathbf{P}(z)\mathbf{U}(z)$ . E com isso, pode-se fatorar os passos de atualização e predição.

Com a descrição matricial, podemos descrever ainda a estrutura *lifting* inversa. Definimos  $\tilde{\mathbf{P}}(z)$  e  $\tilde{\mathbf{U}}(z)$  como as matrizes de atualização e predição da estrutura inversa, tal que:

$$\tilde{\mathbf{P}}(z) = \begin{bmatrix} 1 & 0 \\ P(z) & 1 \end{bmatrix}, \quad (4.8)$$

$$\tilde{\mathbf{U}}(z) = \begin{bmatrix} 1 & -U(z) \\ 0 & 1 \end{bmatrix}. \quad (4.9)$$

Deste modo, o sinal  $x[m]$  decomposto em componentes polifásicas é reconstruído como:

$$\begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix} = \tilde{\mathbf{P}}\tilde{\mathbf{U}} \begin{bmatrix} c[n] \\ d[n] \end{bmatrix} \quad (4.10)$$

De modo a comprovar que a inversa é dada pelas matrizes ditas, podemos substituir os coeficientes  $c[n]$  e  $d[n]$ , de modo que:

$$\begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix} = \tilde{\mathbf{P}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{P} \begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix}. \quad (4.11)$$

Substituindo as matrizes  $\mathbf{P}(z)$ ,  $\mathbf{U}(z)$ ,  $\tilde{\mathbf{P}}(z)$  e  $\tilde{\mathbf{U}}(z)$  como descritas nas Equações 4.6–4.9, temos que:

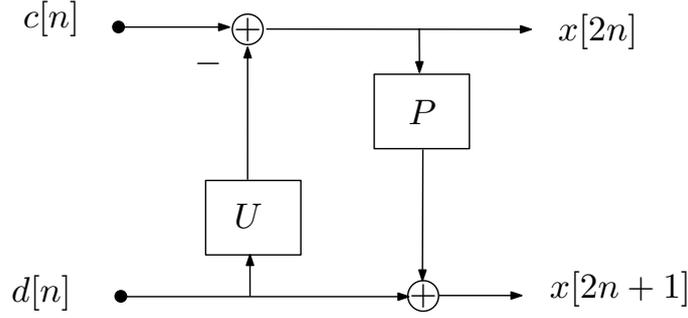


Figura 4.6: Inversa do esquema *Lifting*.

$$\begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ P(z) & 1 \end{bmatrix} \begin{bmatrix} 1 & -U(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & U(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -P(z) & 1 \end{bmatrix} \begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix}; \quad (4.12)$$

$$\begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ P(z) & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -P(z) & 1 \end{bmatrix} \begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix}; \quad (4.13)$$

$$\begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x[2n] \\ x[2n+1] \end{bmatrix}. \quad (4.14)$$

Como o resultado obtido nas Equações 4.12–4.14 é a identidade, prova que as matrizes do esquema *lifting* inverso são válidas. O diagrama do esquema *lifting* inverso está presente na Figura 4.6.

No Matlab é possível fatorar estas matrizes em passos de *lifting* usando o comando LIFTWAVE. Os detalhes teóricos para realizar esta fatoração estão completamente descritos no trabalho de Ingrid Daubechies [83].

## 4.4 Esquema *Lifting* para a Fatoração Espectral

Gostaríamos de usar a fatoração espectral na dimensão temporal incorporando a estimação de movimento. Para isso, inicialmente, teríamos que implementar a *wavelet* não-decimada usando LS. Em [84], existe a proposta de se implementar os passos de *lifting* segundo a Figura 4.7.

Nesta Figura, observa-se que, como ela necessita dos operadores  $P(z)$ ,  $P(z^2)$ ,  $U(z)$  e  $U(z^2)$ , ela está restrita ao uso de filtros lineares, e portanto não é adequada para a incorporação nos operadores  $P(\cdot)$  e  $U(\cdot)$  de uma etapa de estimação de movimento (alinhamento temporal). Desta forma, não será possível a implementação da fatoração espectral incorporando a estimação de movimento pois ela necessitaria desta implementação de uma transformada *wavelet* não-decimada usando passos de *lifting*.

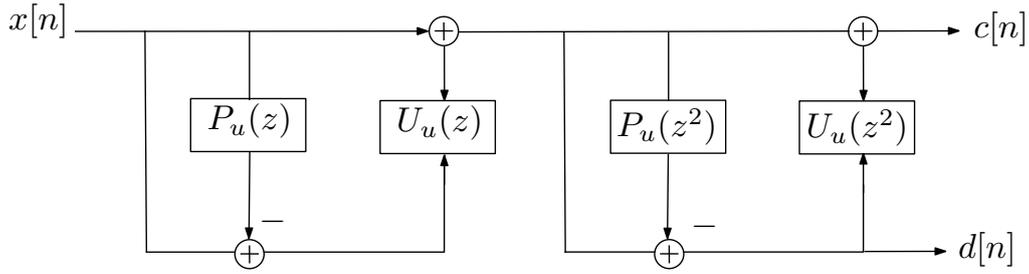


Figura 4.7: Estrutura *Lifting* não-decimada usando os passos provenientes da estrutura decimada.

## 4.5 Esquema *Lifting* Haar usando Compensação de Movimento

Seguindo a ideia das duas últimas Seções, pode-se sugerir um filtro para todo desenvolvimento teórico. Em geral, os usos de MCTF/LS iniciam com o banco de filtros Haar. Esse banco é composto pelos filtros  $h(n) = [1 \ 1]/\sqrt{2}$  e  $g(n) = [1 \ -1]/\sqrt{2}$ . Modificando a Figura 4.4 e adicionando um deslocamento (representado pelo bloco  $\mathbf{d}$ ), a estrutura *lifting* usando o banco de filtros Haar e MCTF é apresentada na Figura 4.8.

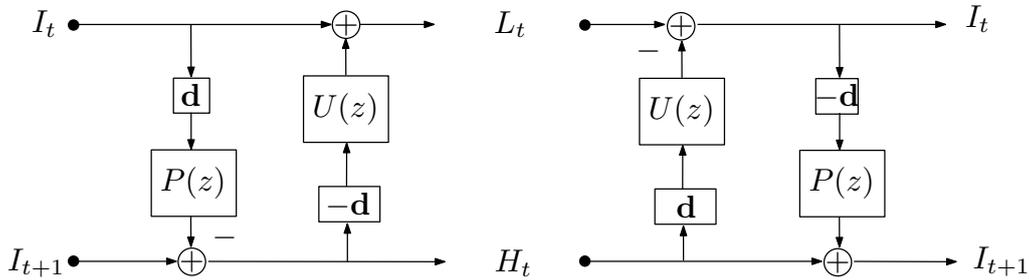


Figura 4.8: Esquemas *lifting* usando compensação de movimento e o banco de filtros Haar. Neste caso, o filtro de previsão  $P(z)$  é igual a  $1/\sqrt{2}$  e o filtro de atualização  $U(z)$  é igual a  $\sqrt{2}$ . A esquerda está o esquema responsável pela decomposição. A direita, o esquema que realiza a síntese das sub-bandas.

Analicamente, da Figura 4.8, se  $P(z) = 1/\sqrt{2}$  e  $U(z) = 1/2$ , podemos escrever que os pixels conectados serão filtrados ao longo da trajetória por [85]:

$$H_t[m, n] = I_{t+1}[m, n] - (I_t[m - d_m, n - d_n]/\sqrt{2}). \quad (4.15)$$

$$L_t[m - d_m, n - d_n] = I_t[m, n] + \sqrt{2}H_t[m + d_m, n + d_n], \quad (4.16)$$

onde  $\mathbf{d} = [d_m, d_n]$  é o vetor de movimento *forward*, ou seja, a referência é o quadro anterior.

Os quadros serão reconstruídos usando:

$$I_t[m - d_m, n - d_n] = L_t[m - d_m, n - d_n] - \sqrt{2}H_t[m, n], \quad (4.17)$$

$$I_{t+1}[m, n] = H_t[m, n] + (L_t[m + d_m, n + d_n]/\sqrt{2}). \quad (4.18)$$

Para os pixels não-conectados, ou seja, os pixels que não são definidos após a compensação de movimento, o LS pode ser usado para contornar este problema. Por garantir a invertibilidade de quaisquer operações  $P(\cdot)$  e  $U(\cdot)$ , pode-se tratar de maneira diferente cada tipo de pixel. Seguindo a sugestão de tratamento dada por Peisong *et al.* [78], tem-se o seguinte resultado:

$$L_t[m, n] = \frac{2I_t}{\sqrt{2}}, \quad (4.19)$$

$$H_t[m, n] = (I_{t+1}[m, n] - I_t[m, n])/\sqrt{2}. \quad (4.20)$$

As Equações 4.19 e 4.20 são aproximações da filtragem Haar quando não é possível realizá-la através do MCTF, porém garantindo assim a reconstrução perfeita.

Note que nas Equações 4.15– 4.18,  $\mathbf{d}$  não precisa ser inteiro. Neste caso, as mesmas Equações 4.15– 4.18 podem ser usadas em conjunto com a compensação de movimento usando acurácia de subpixel. Porém, é importante atentar aos passos em que é necessário realizar o processo de interpolação antes de prosseguir.

## 4.6 Esquema *Lifting* Le Gall 5-3 usando Compensação de Movimento

De maneira equivalente ao banco de filtros Haar, o banco Le Gall 5-3 pode ser decomposto em passos de *lifting* como [86]:

$$P_1(z) = -0.200000; \quad (4.21)$$

$$U_1(z) = -0.208333 - 0.625000z^{-1}; \quad (4.22)$$

$$P_2(z) = +0.900000z + 1.500000; \quad (4.23)$$

$$U_2(z) = -0.069444z + 0.333333 - 0.069444z^{-1}. \quad (4.24)$$

Essa decomposição utiliza dois passos de decomposição, o que torna esta solução mais complexa de ser implementada. Tendo isto em vista, a literatura [85] apresenta uma maneira mais simples de computar esta decomposição usando apenas um passo de *lifting*. Usando apenas um passo, o banco de filtros Le Gall 5-3 pode ser escrito pelos filtros de predição  $P(z)$  e atualização  $U(z)$  como:

$$P(z) = 0.5(1 + z); \quad (4.25)$$

$$U(z) = 0.25(z^{-1} + 1). \quad (4.26)$$

Usando esta decomposição, a sub-banda passa-altas e passa-baixas ( $H_t$  e  $L_t$ ) são descritas por:

$$H_t[m, n] = I_{2t-1}[m, n] - 0.5(I_{2t}[m - d_m, n - d_n] + I_{2t-2}[m - b_m, n - b_n]), \quad (4.27)$$

$$L_t[m, n] = I_{2t}[m, n] + 0.25(H_{t+1}[m - d_m, n - d_n] + H_t[m - b_m, n - b_n]). \quad (4.28)$$

onde  $\mathbf{b} = [b_n, b_m]$  é o vetor de movimento *backward*, ou seja, a referência é o quadro seguinte ao atual. Em alguns casos, pode-se simplificar a estimação de movimentos fazendo  $\mathbf{d} = -\mathbf{b}$ .

Neste ponto é importante perceber que podemos ter dois tipos de pixels não-conectados. Os pixels parcialmente conectados são aqueles cujos uma das direções de compensação ainda é válida. Por exemplo, o pixel  $[i - d_m, j - d_n]$  existe e o pixel  $[i - b_m, j - b_n]$  não. O outro caso é quando ambos os pixels do exemplo não existem.

Os pixels parcialmente não-conectados (ou seja, os vetores de movimento *forward* ou *backward* não representam uma posição válida para um pixel, porém ao menos uma ainda é válida) são tratados por:

$$H_t[m, n] = I_{2t-1}[m, n] - I_{2t}[m - d_m, n - d_n], \quad (4.29)$$

$$L_t[m, n] = \begin{cases} I_{2t}[m, n] + 0.5(H_t[m - d_m, n - d_n]) \\ I_{2t}[m, n] + 0.5(H_t[m - b_m, n - b_n]). \end{cases} \quad (4.30)$$

Para os pixels não-conectados (para ambos os tipos de vetores de movimento), as sub-bandas são escritas como:

$$H_t[m, n] = I_{2t-1}[m, n] - I_{2t}[m, n], \quad (4.31)$$

$$L_t[m, n] = I_{2t}[m, n]. \quad (4.32)$$

$$(4.33)$$

O mesmo processo descrito na Seção 4.5 para a adequação ao aumento da acurácia da compensação de movimento deve ser utilizado aqui.

## 4.7 Esquema *Lifting* Cohen-Daubechies-Feauveau 9-7 usando Compensação de Movimento

O banco de filtros Cohen-Daubechies-Feauveau 9-7 é decomposto em passos de *lifting* como [86]:

$$P_1(z) = +1.586134(1 + z^{-1}); \quad (4.34)$$

$$U_1(z) = -0.052980(1 + z); \quad (4.35)$$

$$P_2(z) = -0.882911(1 + z^{-1}); \quad (4.36)$$

$$U_2(z) = -0.052980(1 + z). \quad (4.37)$$

Novamente, este filtro é realizado através de dois passos de *lifting* e não há implementação deste banco em apenas um passo disponível na literatura. Porém, é possível [77, 85] usar duas vezes a descrição dos passos do filtro Le Gall 5-3 a fim de obter os mesmos resultados do banco de filtros Cohen-Daubechies-Feauveau 9-7. A única modificação importante é o escalamento entre cada aplicação do filtro Le Gall 5-3. O escalamento entre as aplicações do filtro são as constantes  $\mu_1$  e  $\mu_2$  que necessitamos multiplicar as sub-bandas provenientes da primeira aplicação do LS com o filtro Le Gall 5-3 tal que, ao aplicarmos novamente este esquema, obtenhamos o resultado de um LS com o filtro Cohen-Daubechies-Feauveau 9-7.

Em [77], as escalas para as sub-bandas passa-baixas e passa-altas ( $\mu_1$  e  $\mu_2$  respectivamente) são dadas por:

$$\mu_1 = 1 + 2\beta(1 + 2\alpha), \quad (4.38)$$

$$\mu_2 = 1 + 2\delta \left( \frac{1 + 2\alpha}{\mu_1} + 2\gamma \right), \quad (4.39)$$

onde  $\alpha = -1.586134$ ,  $\beta = -0.052980$ ,  $\delta = 0.443506$  e  $\gamma = 0.882911$ .

## 4.8 Considerações Finais

No início deste Capítulo, vimos alguns trabalhos de fusão de vídeos que tentam adicionar a informação temporal em seu método de decomposição. Usando ferramentas presentes nos padrões de codificação de vídeo existentes, propomos a utilização do MCTF, ou seja, da filtragem temporal usando compensação de movimento, como uma forma de incluir a informação temporal no processo de fusão.

Para adicionar a etapa de compensação de movimento à filtragem, usamos a

implementação de banco de filtros em passos de *lifting*. Com esse esquema, além de a compensação de movimento ser incorporada mantendo a reconstrução perfeita, obtemos vantagens computacionais e conseguimos tratar exceções, como no caso dos pixels não-conectados, visto nas Seções 4.5, 4.6 e 4.7.

Nas Seções 4.5, 4.6 e 4.7 foram mostradas maneiras de realizar a decomposição temporal utilizando 3 bancos de filtros populares nesta área: Haar, Cohen-Daubechies-Feauveau 9-7 e Le Gall 5-3.

No próximo Capítulo, iremos aplicar as decomposições apresentadas aos métodos de fusão para a inclusão da informação temporal.

# Capítulo 5

## Algoritmos de Fusão

### Espaço-temporal

Os métodos espaciais de fusão de imagens são muito disseminados na literatura, em especial os métodos multirresolução (Capítulo 3). Pode-se estender o uso destes métodos em vídeos, aplicando-os em cada quadro. No entanto, esta maneira de aplicação desconsidera a informação temporal presente na sequência de vídeo.

Tendo em vista os métodos multirresolução, a extensão tridimensional dos métodos é a solução direta para este problema [24]. Porém há restrições em considerar as dimensões espaciais e temporal da mesma maneira.

No Capítulo anterior, o desenvolvimento da filtragem temporal usando compensação de movimentos (MCTF) promove condições para a agregação da informação temporal de maneira diferenciada. O MCTF pode ser usado para realçar saliências temporais, diferente dos métodos multirresolução tridimensionais.

Neste Capítulo, serão abordadas as extensões tridimensionais das transformadas *wavelet* discretas e não-decimadas com fatoração espectral. Após isso, serão propostos alguns métodos espaço-temporais tais como a decomposição temporal em sub-bandas, a decomposição espaço-temporal em sub-bandas e a transformada *wavelet* com movimento compensado.

#### 5.1 Decomposição Tridimensional

No Capítulo 3, foram descritos os métodos DWT e UWT para uma dimensão. Como dito nesta descrição, esses métodos foram estendidos para duas dimensões aplicando as transformadas nas linhas e nas colunas. Nesta Seção, será feita a extensão para a dimensão temporal.

Usando os mesmos filtros para linhas/colunas/eixo temporal, serão testadas as versões tridimensionais 3D-DWT e 3D-UWT.

Começando pela DWT, na Tabela 5.1, foram selecionados alguns filtros que obtiveram os melhores resultados para a comparação com a versão 3D. Os resultados para as métricas definidas no Capítulo 2 estão na Tabela 5.1.

<i>Camouflage, bior2.2, level 2</i>				
	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
2D	<b>0.501457</b>	<b>0.157272</b>	<b>0.4224137</b>	0.735748
3D	0.495783	0.156885	0.420978	<b>0.900603</b>
<i>Guanabara Bay, haar, level 3</i>				
	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
2D	<b>0.465891</b>	0.222232	<b>0.416806</b>	0.666392
3D	0.448982	<b>0.266659</b>	0.412872	<b>0.774091</b>
<i>Trees, haar, level 4</i>				
	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
2D	<b>0.454091</b>	0.333061	<b>0.426256</b>	0.746475
3D	0.443040	<b>0.348343</b>	0.421361	<b>0.840910</b>
<i>Hangar, haar, level 4</i>				
	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
2D	0.526536	0.262136	0.420055	0.984964
3D	<b>0.527718</b>	<b>0.390944</b>	<b>0.432343</b>	<b>1.244758</b>
<i>Lab, haar, level 3</i>				
	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
2D	<b>0.543947</b>	0.314095	<b>0.425516</b>	0.778577
3D	0.522047	<b>0.346328</b>	0.421797	<b>0.840596</b>

Tabela 5.1: Comparação entre a DWT-2D e a DWT-3D, ou seja, usando a dimensão temporal na decomposição ou não.

Com os resultados da Tabela 5.1, é difícil a comprovação de que a inclusão do tempo na transformada melhora os resultados objetivos. Analisando separadamente, podemos ver que:

- $Q_{AB/F}$ :  
Apenas a sequência *Hangar* atingiu um resultado superior na versão tridimensional (ganho de aprox. 0.22%). No geral, o acréscimo do tempo não favoreceu as bordas de cada quadro individualmente.
- $M_{AB/F}$ :  
Com esta métrica é possível ver que o acréscimo do eixo temporal na transformada acrescentou mais informação em cada quadro do vídeo. A exceção foi a sequência *Camouflage* que não atingiu valores melhores usando esta abordagem.
- $DQ$ :  
Igualmente a métrica  $Q_{AB/F}$ , esta métrica só obteve resultados melhores para a sequência *Hangar*.

Camouflage				
2D	$Q_{AB/F}$ <b>0.515953</b>	$M_{AB/F}$ <b>0.145741</b>	$DQ$ <b>0.423470</b>	$MI - IFD$ <b>0.995856</b>
3D	0.337847	0.051903	0.390981	0.053608
Guanabara Bay				
2D	$Q_{AB/F}$ <b>0.460692</b>	$M_{AB/F}$ <b>0.234711</b>	$DQ$ <b>0.416152</b>	$MI - IFD$ <b>0.866403</b>
3D	0.306106	0.196600	0.393604	0.084161
Trees				
2D	$Q_{AB/F}$ <b>0.469360</b>	$M_{AB/F}$ <b>0.344809</b>	$DQ$ <b>0.429091</b>	$MI - IFD$ <b>0.971692</b>
3D	0.312232	0.084050	0.370404	0.015836
Hangar				
2D	$Q_{AB/F}$ <b>0.539683</b>	$M_{AB/F}$ <b>0.265776</b>	$DQ$ <b>0.424366</b>	$MI - IFD$ <b>1.169460</b>
3D	0.264529	0.108488	0.283289	0.041566
Lab				
2D	$Q_{AB/F}$ <b>0.548274</b>	$M_{AB/F}$ <b>0.294486</b>	$DQ$ <b>0.426164</b>	$MI - IFD$ <b>0.992761</b>
3D	0.286474	0.087992	0.378743	0.067320

Tabela 5.2: Comparação da UWT-SF aplicada a cada quadro com a UWT-SF tridimensional (usando o banco de filtros  $B_3$ -spline e 4 níveis de decomposição).

- MI-IFD:

Todos os resultados provenientes da transformada tridimensional se mostraram melhores usando esta métrica.

Da mesma maneira, foram computados os resultados para a transformada 3D-UWT com fatoração espectral. Os resultados estão disponíveis na Tabela 5.2.

De uma maneira geral, os resultados obtidos para a versão tridimensional da UWT-SF obteve resultados inferiores a UWT-SF bidimensional.

Nas Figuras 5.1 e 5.2 estão presentes amostras dos resultados de cada método para duas sequências. Destas imagens podemos constatar alguns fatos. O primeiro deles é a presença marcante de artefatos de fusão nos vídeos fundidos usando a 3D-DWT (Figura 5.1(a)). Há presença de artefatos também nas bordas que participam do movimento da cena, como por exemplo, na Figura 5.1(b).

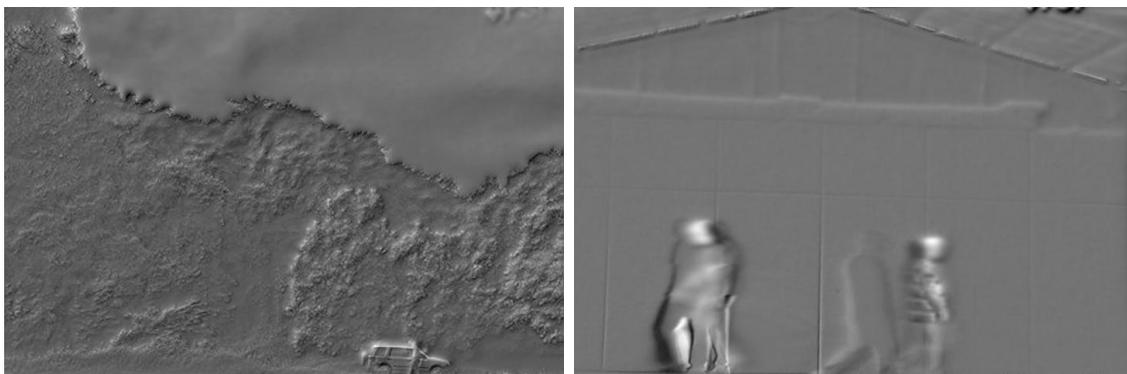
Os resultados ruins da 3D-UWT são corroborados ao analisar a Figura 5.2. Os filtros de suporte mais longos em conjunto com a UWT fazem com que os quadros fundidos tenham baixa qualidade e presença de informações provenientes de vários quadros. Note que, nesta abordagem, não foi utilizada a técnica de filtragem usando compensação de movimento (MCTF).



(a) Quadro 100 da sequência *Trees*

(b) Quadro 5 da sequência *Hangar*

Figura 5.1: Exemplos de quadros fundidos usando a técnica 3D-DWT.



(a) Quadro 100 da sequência *Trees*

(b) Quadro 5 da sequência *Hangar*

Figura 5.2: Exemplos de quadros fundidos usando a técnica 3D-UWT.

## 5.2 Algoritmo de Fusão usando Esquema *Lifting* e Compensação de movimento

O arcabouço proveniente das ferramentas apresentadas no Capítulo 4 será usado nesta Seção para incluir a informação temporal nos algoritmos de fusão espaço-temporais. Vimos que com o esquema *lifting* é possível acrescentar alguns tipos de processamento assegurando a invertibilidade do processo.

Nesta Seção, será utilizada a sequência *Camouflage* como exemplo para os métodos propostos. Além disso, os resultados para as demais sequências de teste estão disponíveis no Apêndice B.

### 5.2.1 Decomposição em Sub-bandas Temporais

O domínio temporal tem diferentes características em comparação ao espacial em um vídeo. Por exemplo, transições abruptas nos valores entre os pixels de um dado quadro representam bordas, ou detalhes que podem ser importantes. No

domínio temporal, estas transições podem representar algum movimento da cena, uma oclusão/aparecimento de algum objeto, ou ruído. Em relação ao último ponto, no domínio temporal, apenas uma leve alteração na média entre os quadros pode trazer uma transição abrupta no valor dos pixels. As situações citadas implicam em um dilema: a preservação ou não destas transições pode trazer resultados melhores ou piores.

Uma maneira de contornar este problema é a ferramenta que foi proposta no Capítulo 4. A compensação de movimento é capaz de alinhar as bordas espaciais da imagem e reduzir as transições na dinâmica de cada pixel dos quadros ao longo do tempo. Com isso, as transições temporais dos valores de cada pixel podem ser classificadas como provenientes do movimento ou não. E isto será utilizado no algoritmo de fusão de vídeos. Para utilizar a informação presente no movimento, iremos computar a transformada *wavelet* no eixo temporal em conjunto com a compensação de movimento usando o esquema *lifting* mostrado no Capítulo 4.

O modelo proposto ao qual nomeamos como *Decomposição em Sub-bandas temporais* pode ser dividido em três etapas principais:

1. Estimação/Compensação de movimento: calcular os vetores de movimento entre o quadro de referência e os outros; calcular os quadros compensados.
2. Filtragem: computar a filtragem de cada posição do quadro ao longo do tempo. Ou seja, realizar o MCTF.
3. Regra de fusão temporal: Após as sub-bandas temporais calculadas, tem-se que aplicar uma regra com o intuito de reunir ambas as entradas em apenas um conjunto de sub-bandas.

Em primeiro lugar, o processo de estimação de movimento deve levar em conta o tipo de filtragem que será realizado. Nos bancos de filtros descritos nas Seções 4.5, 4.6 e 4.7 foram usadas tanto estimação de movimento unidirecional (*forward* ou *backward*) quanto bidirecional. As formas de realizar a estimação e a compensação de movimento são amplamente abordadas na literatura [72]. O algoritmo usado nesta dissertação usará a técnica presente em [87] no intuito de reduzir a complexidade computacional.

A não-linearidade no processo de compensação pode tornar difícil o processo de reconstrução (transformada inversa) da imagem fundida. Por isso, serão usados os métodos de MCTF/LS estudados. Além disso, alguns pontos devem ser definidos para facilitar este processo. Um deles é a escolha dos vetores de movimento. Tendo em vista o par de vídeos infravermelho e de luz visível, tem-se dois conjuntos de vetores de movimento possíveis nesta decomposição que são o conjunto *vis*, proveniente da sequência visível e o conjunto *ir*, da sequência infravermelha. Porém, na

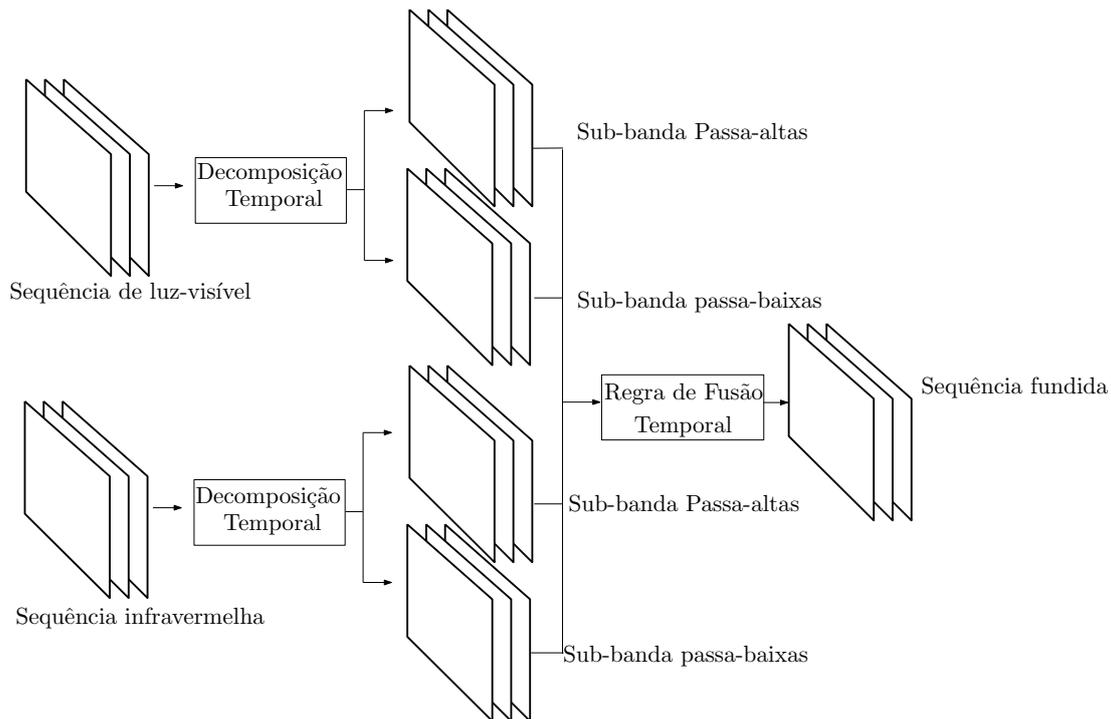


Figura 5.3: Diagrama de fusão usando Decomposição Temporal em Sub-bandas. O primeiro passo é a decomposição temporal usando MCTF. Após a decomposição, as sub-bandas são combinadas através da regra de fusão temporal.

reconstrução, apenas um conjunto deve ser escolhido. Na prática, será definida uma regra para os vetores de movimento. O conjunto resultante desta regra será aplicado na decomposição de ambas as entradas e na reconstrução.

Descrevendo um pouco mais estes processos, é preciso definir alguns parâmetros. Em relação à estimação de movimento, é preciso definir o tamanho do bloco e a acurácia. Nos nossos experimentos serão utilizados blocos de tamanho 4 e 8 e serão usadas acurácias unitária,  $1/2$  e  $1/4$  de pixel.

Após isto, uma regra para fusão dos vetores de movimento (*MVR – Motion Vector Rule*) deve ser definida. Definimos 5 regras simples:

1. *Mean*: Média entre os MVs de ambas as entradas.
2. *Max*: Máximo entre cada componente do MV das entradas.
3. *Min*: Mínimo entre cada componente do MV das entradas.
4. *Ir*: Apenas os MV da sequência infravermelho.
5. *Vis*: Apenas os MV da sequência de luz visível.

Após a escolha dos vetores de movimento, deve-se realizar a filtragem. Esta etapa será feita conforme descrito nas Seções 4.5, 4.6 e 4.7.

Por fim, a regra de fusão temporal tem que ser definida. Foram criadas 5 regras simples para a fusão temporal, são elas:

<b>Regra de fusão temporal</b>	<b>Sub-banda Passa-baixas</b>	<b>Sub-banda Passa-altas</b>
<i>minmax</i>	Valor mínimo	Valor máximo
<i>meanmax</i>	Valor médio	Valor máximo
<i>maxmean</i>	Valor máximo	Valor médio
<i>meanmean</i>	Valor médio	Valor médio

Tabela 5.3: Regras de fusão temporal que serão usadas nos experimentos.

Foram definidas regras simples com o intuito de testar/exemplificar a eficácia da decomposição propriamente dita. Sumarizando o método, tem-se o diagrama da Figura 5.3 e o Algoritmo 1.

No primeiro experimento será usado o tamanho de bloco igual a 8 para a compensação de movimentos e acurácia unitária. Será usado também o banco de filtros Haar, de acordo como descrito na Seção 4.5.

Este primeiro experimento apresenta alguns pontos interessantes a serem destacados, como pode-se observar na Tabela 5.4. No que diz respeito às regras temporais, há diferentes sensibilidades a elas em cada métrica. Por exemplo, para as métricas  $Q_{AB/F}$  e  $M_{AB/F}$ , as regras temporais *meanmean* e *maxmean* se destacam. A métrica DQ também mensura a manutenção das bordas espaciais, tal como o  $Q_{AB/F}$ , e inclui as bordas temporais. Porém não há tanta discrepância entre as regras para esta métrica. A métrica MI-IFD apresenta valores muito inferiores aos vistos no Capítulo 3 e na Seção 5.1. A respeito desta métrica, discutiremos mais adiante. Em relação às regras para a escolha dos vetores de movimento, não há padrão eminente que destaque melhores resultados.

Seguindo os experimentos, usaremos tamanho de bloco igual a 4 e acurácia unitária para a mesma estrutura do experimento anterior. Os resultados estão dispostos na Tabela 5.5.

Nos resultados deste experimento, percebe-se que as regras temporais resultam no mesmo comportamento do experimento anterior. Além disso, este experimento corrobora que as regras para a escolha do vetor de movimento não interferem tanto no resultado.

Seguindo adiante, pode-se aumentar a acurácia na estimação/compensação de movimento. Nas Tabelas 5.6 e 5.7 estão os resultados obtidos para a acurácia de  $1/2$  e  $1/4$  de pixel. As regras de fusão temporal *minmax* e *meanmax* serão omitidas por não trazerem nenhum ganho.

Nestes resultados têm-se uma ligeira melhoria aos obtidos anteriormente. Ainda nestes resultados, não há padrão visível em relação as regras de fusão dos vetores de movimento.

---

**Algoritmo 1:** DECOMPOSIÇÃO EM SUB-BANDAS TEMPORAIS USANDO HAAR

---

```
Entrada:  $I_A, I_B, blk, pel$   
for ( $t=1$  até  $t = \text{número de quadros}-1$ ) do  
  --Estimar movimentos  
   $\mathbf{d}_A(t) \leftarrow \text{estimarMovimento}(I_A(t), I_A(t+1), blk, pel);$   
   $\mathbf{d}_B(t) \leftarrow \text{estimarMovimento}(I_B(t), I_B(t+1), blk, pel);$   
  --Definir vetor de movimentos que será usado  
   $\mathbf{d}(t) \leftarrow MVR(\mathbf{d}_A(t), \mathbf{d}_B(t));$   
  --Inicialização das variáveis  
   $H_A, L_A, H_B, L_B \leftarrow NULL;$   
  --Filtragem dos pixels conectados  
  for ( $m=1$  até  $m= \text{número de linhas}$ ) do  
    for ( $n=1$  até  $n= \text{número de colunas}$ ) do  
      --Sub-bandas da sequência A  
       $H_A[m, n] \leftarrow I_A[m, n](t+1) - (I_A[m - d_m(t), n - d_n(t)](t)/\sqrt{2});$   
       $L_A[m - d_m(t), n - d_n(t)] \leftarrow$   
         $I_A[m, n](t) + \sqrt{2}H_A[m + d_m(t), n + d_n(t)](t);$   
      --Sub-bandas da sequência B  
       $H_B[m, n] \leftarrow I_B[m, n](t+1) - (I_B[m - d_m(t), n - d_n(t)](t)/\sqrt{2});$   
       $L_B[m - d_m(t), n - d_n(t)] \leftarrow$   
         $I_B[m, n](t) + \sqrt{2}H_B[m + d_m(t), n + d_n(t)](t);$   
    end  
  end  
  --Filtragem dos pixels não-conectados  
  for ( $m=1$  até  $m= \text{número de linhas}$ ) do  
    for ( $n=1$  até  $n= \text{número de colunas}$ ) do  
      if ( $H_A[m, n] == NULL \vee L_A[m, n] == NULL$ ) then  
         $L_A[m, n] \leftarrow 2I_A[m, n](t)/\sqrt{2};$   
         $H_A[m, n] \leftarrow (I_A[m, n](t+1) - I_A[m, n](t))/\sqrt{2};$   
      end  
      if ( $H_B[m, n] == NULL \vee L_B[m, n] == NULL$ ) then  
         $L_B[m, n] \leftarrow 2I_B[m, n](t)/\sqrt{2};$   
         $H_B[m, n] \leftarrow (I_B[m, n](t+1) - I_B[m, n](t))/\sqrt{2};$   
      end  
    end  
  end  
  --Aplicar regra de fusão temporal  
   $H_F \leftarrow \text{regraDeFusao.1}(H_A, H_B);$   
   $L_F \leftarrow \text{regraDeFusao.2}(L_A, L_B);$   
  --Reconstrução  
  for ( $m=1$  até  $m= \text{número de linhas}$ ) do  
    for ( $n=1$  até  $n= \text{número de colunas}$ ) do  
       $I_F[m - d_m(t), n - d_n(t)](t) \leftarrow$   
         $L_F[m - d_m(t), n - d_n(t)](t) - \sqrt{2}H_F[m, n](t);$   
       $I_F[m, n](t+1) \leftarrow H_F[m, n](t) + (L_F[m + d_m(t), n + d_n(t)](t)/\sqrt{2});$   
    end  
  end  
end  
retorna  $I_F$ 
```

---

Filtro Haar - Tamanho de bloco 8 e acurácia unitária.					
MVR	Regra temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	minmax	0.201741	0.057032	0.341277	0.069802
max		0.206172	0.055258	0.345326	0.065786
min		0.201321	0.058361	0.339255	0.079685
ir		0.203974	0.055829	0.341430	0.069816
vis		0.206819	0.057120	0.344607	0.076578
mean	meanmax	0.194986	0.086923	0.332982	0.086222
max		0.198664	0.082549	0.333212	0.066156
min		0.198917	0.089753	0.332543	0.099788
ir		0.198002	0.086707	0.332064	0.081120
vis		0.199949	0.085199	0.333594	0.076963
mean	maxmean	0.485077	<b>0.284353</b>	0.387027	0.119455
max		0.485185	0.280612	0.387926	0.063003
min		0.479693	0.282693	0.386060	0.155468
ir		0.482803	0.282216	0.385589	0.098221
vis		0.482785	0.281227	0.387978	0.071080
mean	meanmean	<b>0.517359</b>	0.162229	<b>0.423124</b>	0.483879
max		<b>0.517359</b>	0.162229	<b>0.423124</b>	0.483868
min		0.517356	0.162229	0.423123	<b>0.483892</b>
ir		0.517357	0.162229	0.423123	0.483888
vis		0.517358	0.162229	0.423123	0.483869

Tabela 5.4: Fusão da cena *Camouflage* usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 8 e acurácia unitária.

### Outros Banco de Filtros

Após estes testes em que foram testados os parâmetros da compensação de movimento, pode-se alterar também o banco de filtros. No contexto do uso do MCTF, a literatura [77, 80, 85] apresenta comumente os bancos de filtros Le Gall 5-3 e Cohen-Daubechies-Feauveau 9-7 [63], conforme definidos no Capítulo anterior.

Usando a metodologia apresentada nas Seções 4.6 e 4.7, obtivemos os resultados presentes na Tabela . Nesta estão apresentados os resultados usando a compensação de movimentos mais fina testada (bloco tamanho 4 e acurácia de 1/4 de pixel). Aqui foi apresentada também apenas a regra de fusão temporal *meanmean*. Esta foi escolhida mediante os testes anteriores.

Os resultados obtidos por estes filtros foram inferiores (para todas as métricas) comparativamente aos obtidos para o banco Haar. Conforme obtido na Seção 5.1, os filtros mais longos aplicados no eixo temporal deterioram a sequência de vídeo. Pelos resultados mostrados aqui, a compensação de movimento reduz esta deterioração, porém não o suficiente para ultrapassar os resultados obtidos para o banco Haar.

Filtro Haar - Tamanho de bloco 4 e acurácia unitária.					
MVR	Regra temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	minmax	0.215004	0.054169	0.344490	0.049804
max		0.223826	0.051101	0.349021	0.036062
min		0.205930	0.055090	0.340408	0.061896
ir		0.214337	0.052653	0.343794	0.044582
vis		0.222147	0.052926	0.347348	0.043905
mean	meanmax	0.210144	0.078300	0.334221	0.054111
max		0.218723	0.071290	0.342793	0.040618
min		0.199679	0.084214	0.329090	0.073352
ir		0.207966	0.077787	0.329584	0.050995
vis		0.217854	0.074987	0.342147	0.049378
mean	maxmean	0.482146	0.268728	0.394562	0.032165
max		0.472851	0.254971	0.398947	0.034704
min		0.485773	<b>0.278734</b>	0.388069	0.067886
ir		0.483730	0.269019	0.396457	0.030707
vis		0.475391	0.262325	0.396576	0.032457
mean	meanmean	<b>0.517362</b>	0.162229	<b>0.423124</b>	0.483887
max		0.517359	0.162229	0.423123	0.483880
min		0.517358	0.162229	0.423123	0.483890
ir		0.517362	0.162229	<b>0.423124</b>	0.483872
vis		0.517357	0.162229	0.423123	<b>0.483898</b>

Tabela 5.5: Fusão da cena *Camouflage* usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 4 e acurácia unitária.

Filtro Haar					
MVR	Regra temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	maxmean	0.500393	0.280727	0.393452	0.061849
max		0.487583	0.264614	0.395008	0.038488
min		0.507369	<b>0.293293</b>	0.391151	0.127655
ir		0.502353	0.281182	0.392498	0.042062
vis		0.492268	0.273820	0.395830	0.051936
mean	meanmean	<b>0.517369</b>	0.162231	<b>0.423125</b>	<b>0.483940</b>
max		0.517363	0.162230	0.423124	0.483902
min		0.517365	0.162231	0.423124	0.483851
ir		0.517365	0.162231	0.423124	0.483895
vis		0.517364	0.162230	0.423124	0.483860

Tabela 5.6: Fusão da cena *Camouflage* usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 4 e acurácia de 1/2 pixel.

Filtro Haar					
MVR	Regra temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	maxmean	0.500393	0.280727	0.393452	0.061849
max		0.487583	0.264614	0.395008	0.038488
min		0.507369	<b>0.293293</b>	0.391151	0.127655
ir		0.502353	0.281182	0.392498	0.042062
vis		0.492268	0.273820	0.395830	0.051936
mean	meanmean	0.517366	0.162230	0.423124	0.483880
max		0.517367	0.162229	0.423124	0.483909
min		0.517363	0.162229	0.423124	0.483831
ir		<b>0.517372</b>	0.162230	<b>0.423125</b>	0.483838
vis		0.517367	0.162229	0.423124	<b>0.483991</b>

Tabela 5.7: Fusão da cena *Camouflage* usando decomposição em sub-bandas temporais e banco de filtros Haar. Os vetores de movimento foram estimados usando blocos de tamanho 4 e acurácia de 1/4 pixel.

5-3 Lifting Scheme						
Cena	MVR	Regra temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Camouflage	vis	meanmean	0.422651	0.148564	0.402371	0.049448
Bay			0.359412	0.228817	0.365252	0.033764
Trees			0.390775	0.341079	0.406808	0.132723
Hangar			0.312503	0.234414	0.344651	0.048991
Lab			0.328484	0.308087	0.388607	0.045908
9-7 Lifting Scheme						
Camouflage	vis	meanmean	0.410325	0.143541	0.391097	0.023256
Bay			0.349497	0.216315	0.348272	0.009260
Trees			0.384103	0.335516	0.401372	0.025459
Hangar			0.300826	0.215836	0.329593	0.023064
Lab			0.319526	0.289261	0.373492	0.017840

Tabela 5.8: Cenas fundidas usando decomposição em sub-bandas temporais aplicando os filtros 5-3 e 9-7. Em todas as cenas, os vetores de movimento foram calculados usando tamanho de bloco igual a 4 e acurácia unitária.

## Outras sequências

Nesta Seção foram mostrados apenas os resultados para a sequência *Camouflage* para facilitar a variação dos parâmetros e seus efeitos. Entretanto, as outras sequências escolhidas também foram fundidas seguindo os algoritmos propostos. É importante salientar que mesmo para as outras sequências, nenhuma regra de fusão dos vetores de movimento pode ser definida como melhor. Ainda a respeito da compensação de movimento, é notado que o bloco de tamanho 4 tem resultados um pouco melhores do que os resultados obtidos para o tamanho 8. O aumento da

acurácia da estimação também tem efeito positivo na fusão. Porém, a acurácia 1/4 não fornece resultados tão superiores aos resultados obtidos pela acurácia 1/2 de pixel. Tabelas com os resultados completos estão presentes no Apêndice B.

### 5.2.2 Decomposição em Sub-bandas Espaço-Temporais

Após o estudo da decomposição temporal, pode-se incluir também a decomposição espacial. Na última Subseção foi possível concluir que as regras temporais *mean* e *maxmean* obtiveram melhores resultados. Porém, como afirmado na Seção 5.2.1, as regras testadas não cobrem todas as possibilidades.

No Capítulo 3, um *benchmark* foi criado a partir de métodos de fusão de imagens (métodos espaciais). Com estes resultados, foi notado que a UWT-SF obtém os melhores resultados. E também, na maioria destes resultados, os filtros “Haar starck 2”, “ $B_3$ -spline” e “Le Gall 5-3” foram usados.

A ideia da decomposição em sub-bandas espaço-temporais proposta aqui é mostrada na Figura 5.4. Primeiramente, a decomposição temporal em sub-bandas é realizada. Com as sub-bandas calculadas, é realizada a fusão espacial conforme mostrado no Capítulo 3. Com esse posicionamento de operações, não é necessário incluir nenhuma regra de fusão para as sub-bandas temporais.

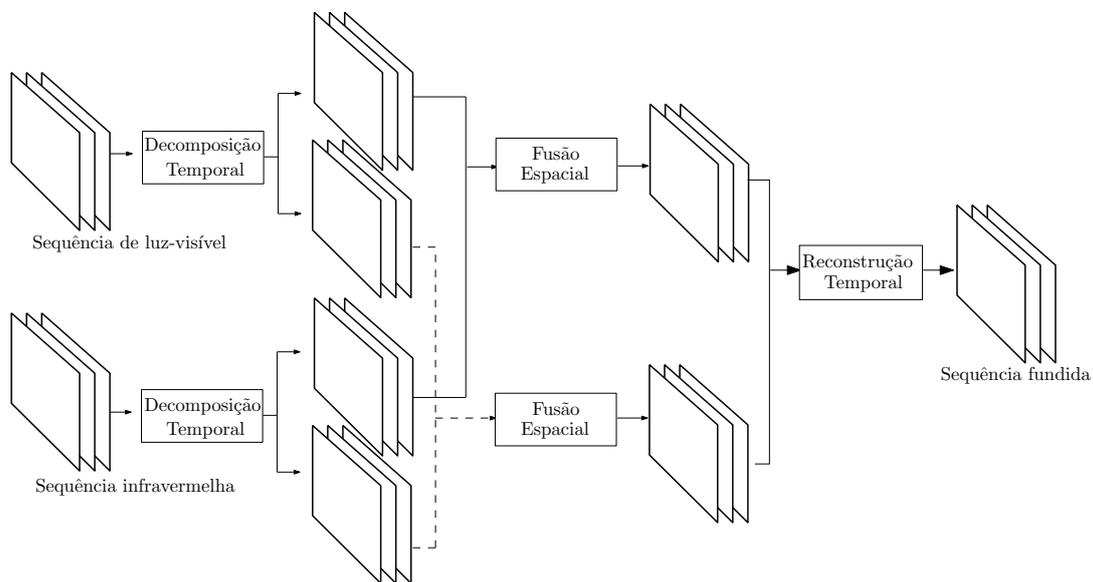


Figura 5.4: Diagrama de fusão usando decomposição em sub-bandas espaço-temporais. Nesta abordagem, primeiramente é aplicada a decomposição em sub-bandas temporais. Após isto, cada sub-banda temporal é fundida através de alguma técnica de fusão de imagens.

Seguindo esta ideia, foi realizado o experimento utilizando a UWT-SF e os três filtros citados anteriormente. Além disso, acompanhando os resultados da decom-

posição temporal em sub-bandas, os parâmetros da compensação de movimentos foram escolhidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

Utilizando a sequência *Camouflage* foram obtidos os resultados presentes na Tabela 5.9

Cena <i>Camouflage</i> - Resultados da fusão espaço-temporal proposta					
Filtro	MVR	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	mean	0.514792	0.142605	0.424459	0.368580
	max	0.514137	0.142397	0.424298	0.354047
	min	0.515021	0.142653	0.424507	0.409397
	ir	0.514733	0.142628	0.424365	0.382517
	vis	0.514629	0.142373	<b>0.424521</b>	0.383467
5-3	mean	0.509873	0.144046	0.424055	0.360374
	max	0.508956	0.143705	0.423652	0.341781
	min	0.510212	<b>0.144083</b>	0.424031	0.373035
	ir	0.509898	0.144043	0.423962	0.357813
	vis	0.509616	0.143731	0.424001	0.354739
$B_3$ -spline	mean	0.513980	0.139534	0.424053	0.383740
	max	0.513193	0.139358	0.423797	0.361953
	min	<b>0.514297</b>	0.139635	0.424102	<b>0.417475</b>
	ir	0.513980	0.139602	0.424000	0.387418
	vis	0.513725	0.139338	0.424041	0.383693

Tabela 5.9: Nesta abordagem, foi usado a decomposição em sub-bandas temporais e então a UWT-SF para a decomposição em sub-bandas espaciais. Na primeira etapa, foi usado o banco de filtros Haar, por obter os melhores resultados na última Seção. Na etapa de compensação de movimento, foi escolhido tamanho de bloco igual a 4 e 1/2 pixel de acurácia. A regra temporal escolhida foi a *meanmean*.

Novamente, os resultados para as outras sequências estão disponíveis no Apêndice B.

### 5.3 Transformada *Wavelet* com Movimento Compensado

Recentemente, Liang Xu *et al.* publicaram um novo método de fusão de vídeos utilizando a informação temporal [59]. Este método contempla tanto a informação temporal quanto a informação espacial de maneira diferente do tratamento utilizado até agora.

As diferenças começam na filtragem temporal. Nos dois métodos propostos até aqui, a filtragem temporal compensada foi a forma utilizada para incluir a informação temporal. A Transformada *Wavelet* com Movimento Compensado -

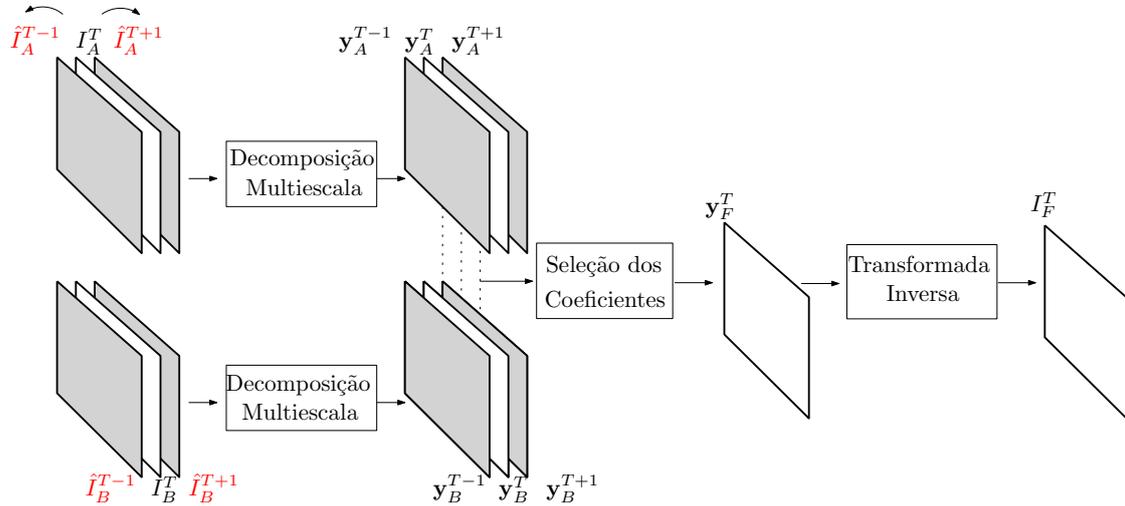
MCWT (*Motion Compensated Wavelet Transform*) utiliza a compensação de movimento para a escolha dos coeficientes multirresolução de diferentes quadros para reconstruir o quadro fundido.

Descrevendo o algoritmo, tem-se as entradas  $I_A(T)$  e  $I_B(T)$ , tomadas no instante de referência  $T$ . Para reduzir a notação, será usado  $I_A^T$  e  $I_B^T$  com o mesmo fim no diagrama da Figura 5.5. Seguindo o diagrama, os quadros  $I_A^{T-1}$  e  $I_B^{T-1}$  são os quadros imediatamente anteriores aos quadros  $I_A^T$  e  $I_B^T$ , respectivamente. Da mesma maneira,  $I_A^{T+1}$  e  $I_B^{T+1}$  são os quadros imediatamente posteriores aos quadros de referência. Os quadros compensados são denotados por  $\hat{I}_A^{t-1}, \hat{I}_B^{t-1} / \hat{I}_A^{t+1}, \hat{I}_B^{t+1}$ .

No diagrama da Figura 5.5, está denotado o algoritmo usando apenas os quadros anterior e posterior à referência. Porém, o número de quadros que pode ser usado é um dos parâmetros deste algoritmo. Após o armazenamento dos quadros compensados e da referência de ambas as entradas, faz-se a decomposição multirresolução (conforme Capítulo 3) para obter os coeficientes multirresolução.

A integração da informação temporal se dá no passo da seleção dos coeficientes. Neste passo, faz-se a escolha de quais coeficientes serão utilizados no quadro  $I_F^T$ . Ou seja, são escolhidos coeficientes para formar o quadro fundido dentre os coeficientes dos quadros compensados seguinte (s) ou anterior (es). Após os coeficientes do quadro fundido serem obtidos, a transformada inversa pode ser tomada e a imagem fundida reconstruída.

Compensação de movimento



Compensação de movimento

Figura 5.5: Diagrama da Transformada *Wavelet* com Movimento Compensado usando um quadro passado e um futuro.

O método originalmente utiliza a DWT como a decomposição multirresolução. Este trabalho também propõe uma regra específica para a escolha dos coeficientes. No intuito comparativo aos métodos propostos neste Capítulo, serão usadas regras

de fusão simples tais quais usadas antes. Como regra de fusão dos coeficientes, será usado o valor máximo dentre os coeficientes de detalhe e o valor médio dentre os coeficientes de aproximação (em menção à regra utilizada nos métodos espaciais).

Na Tabela 5.10 são mostrados os melhores resultados obtidos para a sequência *Camouflage* com diferentes parâmetros na compensação de movimento.

Transformada <i>Wavelet</i> com Movimento Compensado					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
bior2.2	4	0.402161	0.138556	0.402161	0.035865
	8	0.464986	0.150090	0.413282	0.104175
	16	<b>0.475442</b>	0.152380	0.416040	0.137073
bior2.4	4	0.408370	0.137583	0.399348	0.031412
	8	0.462511	0.150496	0.412781	0.093317
	16	0.474727	0.153043	0.416379	0.179530
bior6.8	4	0.416547	0.139530	0.400783	0.028563
	8	0.462089	0.150695	0.412641	0.101432
	16	0.474241	<b>0.15323</b>	<b>0.416772</b>	<b>0.197061</b>

Tabela 5.10: Melhores resultados usando a MCWT. Neste experimento, foi usado tamanhos de bloco iguais a 4,8 e 16; acurácia unitária e 2 níveis de decomposição.

Comparando aos resultados obtidos pela 2D-DWT, pode-se ver a degradação dos resultados causadas por esta técnica. Porém, no trabalho de Liang Xu *et al.*[59], são mostrados resultados satisfatórios. Pode-se concluir disso a eficácia da regra de escolha dos coeficientes utilizada no trabalho e não utilizada aqui.

Para prosseguir com os testes, também foi estudada a possibilidade de uso da UWT no lugar da DWT neste algoritmo. Para a sequência *Camouflage*, apresentamos os melhores resultados obtidos na Tabela 5.11.

Transformada <i>Wavelet</i> com Movimento Compensado usando UWT.					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Le Gall 5-3	4	0.498676	0.137701	0.418945	0.097485
	8	0.510179	0.140149	0.422146	0.226486
	16	0.513273	0.141201	0.423142	0.252690
near_sym_a	4	0.493733	0.137113	0.418590	0.080316
	8	0.504701	0.139166	0.422531	0.257583
	16	0.507848	<b>0.140065</b>	0.423263	<b>0.332224</b>
$B_3$ -spline	4	0.502622	0.132724	0.418282	0.077539
	8	0.504701	0.139166	0.422531	0.257583
	16	<b>0.517908</b>	0.136317	<b>0.423476</b>	0.279874

Tabela 5.11: Melhores resultados usando MCWT e UWT. Neste experimento, foi usado tamanhos de bloco iguais a 4,8 e 16; acurácia unitária e 4 níveis de decomposição.

Novamente, os resultados para esta sequência são inferiores aos obtidos pelo equivalente método espacial (UWT-SF). Porém, para as outras sequências testadas, esta modificação traz alguns ganhos, como pode-se ver no Apêndice B, ao qual discutiremos mais adiante.

### 5.3.1 Transformada *Wavelet* Temporal sem Compensação de Movimento

Se acompanharmos o efeito do tamanho do bloco em cada resultado anterior, pode-se notar que há melhoras ao diminuir o tamanho do bloco. Para cobrir todas as possibilidades, foi realizado um experimento sem compensação de movimentos. Na Tabela 5.12 estão os resultados para este experimento.

Transformada <i>Wavelet</i> Temporal sem Compensação de Movimento usando DWT				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
bior2.2	<b>0.491229</b>	0.154767	<b>0.420613</b>	0.307007
bior2.4	0.489158	0.155249	0.419844	<b>0.333642</b>
bior6.8	0.487661	<b>0.155294</b>	0.419565	0.300543

Tabela 5.12: Melhores resultados usando a Transformada *Wavelet* Temporal sem Compensação de Movimento e DWT.

Este teste também foi realizado para o algoritmo usando a UWT. O resultado está disposto na Tabela 5.13.

Transformada <i>Wavelet</i> Temporal sem Compensação de Movimento usando UWT				
Filtro	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Le Gall 5-3	0.470316	<b>0.134298</b>	0.417834	0.075721
near_sym_a	0.468023	0.134144	0.417939	0.084526
$B_3$ -spline	<b>0.477535</b>	0.130391	<b>0.418900</b>	<b>0.106844</b>

Tabela 5.13: Melhores resultados usando a Transformada *Wavelet* Temporal sem Compensação de Movimento e UWT.

Os efeitos da retirada da etapa de compensação de movimento é adverso para cada decomposição multirresolução. Para a DWT, a Tabela 5.12 mostra melhores resultados retirando a compensação de movimento. Entretanto, para a UWT, a Tabela 5.13 mostra piores resultados.

Concluindo este tópico, estes experimentos demonstraram a necessidade de uma regra de combinação dos coeficientes bem elaborada, como descrita em [59], para superar os métodos de fusão bidimensionais.

## 5.4 Análise dos resultados

Neste Capítulo, foram testadas diversas formas de incluir a informação temporal nos métodos de fusão espacial.

Para finalizar, podemos comparar os resultados obtidos neste Capítulo com os demais resultados. A Tabela 5.14 apresenta a análise destes resultados.

Camouflage			
$Q_{AB/F}$	<b>0.517908</b>	MCWT usando UWT / <i>LiangXu</i> , $B_3$ -spline Bloco 16	p. 73
$M_{AB/F}$	<b>0.293293</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/4, Regra Temp. 'maxmean', MVR 'min'.	p. 69
DQ	<b>0.424521</b>	Decomposição espaço-temporal usando Haar and Haar Starck 2 / Bloco 4, pel 1/2., MVR 'mean'.	p. 71
MI-IFD	<b>0.483991</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/4, Regra Temp. 'meanmean', MVR 'vis'.	p. 69
Guanabara Bay			
$Q_{AB/F}$	<b>0.491924</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'maxmean', MVR 'vis'.	p. 98
$M_{AB/F}$	<b>0.426030</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'maxmean', MVR 'vis'.	p. 98
DQ	<b>0.407423</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'maxmean', MVR 'vis'. MV	p. 98
MI-IFD	<b>0.280173</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'meanmean', MVR 'ir'. MV	p. 98
Trees			
$Q_{AB/F}$	<b>0.531188</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'maxmean', MVR 'vis'.	p. 99
$M_{AB/F}$	<b>0.439540</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'maxmean', MVR 'min'.	p. 99
DQ	<b>0.427762</b>	Decomposição espaço-temporal usando Haar and $B_3$ -spline / Bloco 4, pel 1/2., MVR 'max'. MV	p. 102
MI-IFD	<b>0.500549</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'meanmean', MVR 'ir'.	p. 99
Hangar			
$Q_{AB/F}$	<b>0.533689</b>	Decomposição espaço-temporal usando Haar and Haar Starck 2 / Bloco 4, pel 1/2, MVR 'mean'.	p. 101
$M_{AB/F}$	<b>0.422714</b>	Decomposição espaço-temporal usando Haar and Haar Starck 2 / Bloco 4, pel 1/2, MVR 'mean'. MV.	p. 101
DQ	<b>0.427665</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'meanmean', MVR 'mean'.	p. 99
MI-IFD	<b>0.901050</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'meanmean', MVR 'ir'.	p. 99
Lab			
$Q_{AB/F}$	<b>0.548033</b>	Decomposição espaço-temporal usando Haar and Haar Starck 2 / Bloco 4, pel 1/2, MVR 'mean'.	p. 102
$M_{AB/F}$	<b>0.362498</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'meanmean', MVR 'mean'.	p. 100
DQ	<b>0.425134</b>	Decomposição espaço-temporal usando Haar and $B_3$ -spline / Bloco 4, pel 1/2, MVR 'mean'.	p. 102
MI-IFD	<b>0.574343</b>	Decomposição temporal usando Haar / Bloco 4, pel 1/2, Regra Temp. 'meanmean', MVR 'min'.	p. 100

Tabela 5.14: Melhores resultados usando as técnicas espaço-temporais propostas. Na descrição dos métodos, a abreviação *pel* indica a acurácia da compensação de movimento.

Uma observação importante sobre os resultados da Tabela 5.14, é que a sequência *Camouflage* foi exaustivamente testada neste Capítulo. A compensação de movi-

mento mais fina testada para esta sequência foi para tamanho de bloco igual a 4 e acurácia de 1/4 de pixel. Como foi visto que o ganho entre a acurácias 1/4 e 1/2 é bem pequeno, as demais sequências foram testadas até a acurácia 1/2.

No geral, pode-se notar bons resultados comparativamente aos mostrados na Tabela 3.19. *A priori*, a Decomposição Temporal em Sub-bandas obteve melhor desempenho. Porém, tendo em vista as 4 métricas escolhidas como guia deste trabalho, não é possível notar um método único capaz de maximizar todas elas.

Usando os resultados obtidos no Capítulo 3, especificamente os presentes na Tabela 3.19, pode-se comparar o desempenho segundo as métricas objetivas usadas aos métodos espaciais.

Camouflage			
	Métodos Espaciais	Métodos Espaço-temporais	Ganho relativo
$Q_{AB/F}$	<b>0.518077</b>	0.517908	-0.03 %
$M_{AB/F}$	0.160552	<b>0.293293</b>	82.67%
DQ	0.424082	<b>0.424521</b>	0.10 %
MI-IFD	<b>0.995856</b>	0.483991	-51.39%
Guanabara Bay			
$Q_{AB/F}$	0.465892	<b>0.491924</b>	5.58%
$M_{AB/F}$	0.270686	<b>0.426030</b>	57.38%
DQ	<b>0.416807</b>	0.407423	-2.25%
MI-IFD	<b>0.866403</b>	0.280173	-67.66%
Trees			
$Q_{AB/F}$	0.477809	<b>0.531188</b>	11.17%
$M_{AB/F}$	0.368957	<b>0.439540</b>	19.13%
DQ	<b>0.429091</b>	0.427762	-0.30%
MI-IFD	<b>0.971692</b>	0.500549	-48.48%
Hangar			
$Q_{AB/F}$	<b>0.539683</b>	0.533689	-1.11%
$M_{AB/F}$	0.363756	<b>0.422714</b>	16.20%
DQ	<b>0.427706</b>	0.427665	-0.009%
MI-IFD	<b>1.169460</b>	0.901050	-22.95 %
Lab			
$Q_{AB/F}$	<b>0.552258</b>	0.548033	-0.76%
$M_{AB/F}$	0.356818	<b>0.362498</b>	1.59%
DQ	<b>0.426866</b>	0.425134	-1.29%
MI-IFD	<b>0.992761</b>	0.574343	-42.14%

Tabela 5.15: Resultado comparativo entre os métodos espaciais e os métodos espaço-temporais.

Na Tabela 5.15, os melhores resultados para cada métrica foram utilizadas na comparação. Para facilitar a conclusão da análise, iremos separar cada sequência de teste e analisar as métricas  $Q_{AB/F}$ ,  $M_{AB/F}$  e DQ. A métrica MI-IFD será analisada separadamente.

### 1. *Camouflage*

Esta cena pode ser dividida em três diferentes planos: o fundo (montanha, oceano e céu), a cena propriamente dita (composta pelos indivíduos caminhando) e a fumaça (ligeiramente à frente do plano anterior). Pode-se notar esta divisão através da Figura 5.6.

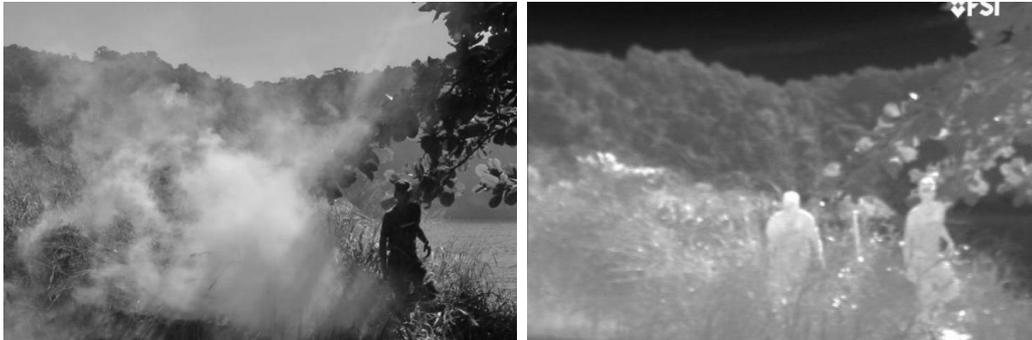


Figura 5.6: Cena *Camouflage*:quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda). Pode-se notar diferentes planos na cena.

Dadas estas características da cena, pode-se acompanhar através das amostras presentes no Apêndice C.1, que existem poucas diferenças em relação a estrutura de cada plano. Ou seja, as melhorias/pioras das bordas dos objetos dos planos referidos não são aparentes. Entretanto, é notável a redução dos artefatos de fusão causados pelo erro de registro. Estes erros são claramente notados ao sobrepor os quadros infravermelho e de luz visível. A utilização da compensação de movimento resulta na diminuição do erro nesta sobreposição. Nota-se isto ao comparar as Figuras 5.7(a) e 5.7(b).



(a) Fusão espacial usando UWT-SF

(b) Decomposição em Sub-bandas Temporais

Figura 5.7: Cena *Camouflage*: amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)).

Das métricas objetivas, a métrica  $M_{AB/F}$  apresentou um ganho notável de 82.67%, denotando uma transferência de informação maior ao usar o método

espaço-temporal de fusão. As métricas  $Q_{AB/F}$  e DQ apresentaram ganhos/perdas pequenos.

## 2. Guanabara Bay

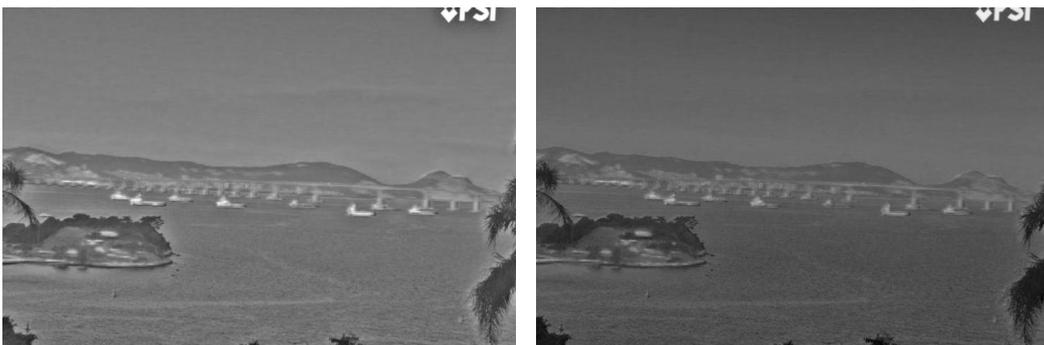
Ao contrário da cena anterior, o plano da cena (Figura 5.8) é distante o suficiente para mitigar os erros de registro.



Figura 5.8: Cena *Guanabara Bay*: quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda).

As características dessa sequência resultam em um ganho de 6.77% na métrica  $Q_{AB/F}$ . Outro ganho expressivo é visto para a métrica  $M_{AB/F}$ . Apesar do ganho da métrica de Xydeas e Petrović, a métrica que avalia a qualidade dinâmica demonstra perdas na qualidade.

Nota-se nesta cena o ganho de qualidade nas bordas da ilha a esquerda (Figuras 5.9(b)) e nas folhas mais próximas a câmera em relação a fusão espacial usando a UWT-SF (Figura 5.9(a)).



(a) Fusão espacial usando UWT-SF

(b) Decomposição em Sub-bandas Temporais

Figura 5.9: Cena *Guanabara Bay*: amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)).

## 3. Trees

Nesta sequência, pode-se corroborar o efeito da diminuição do erro de registro em função da distância do plano da cena (Figura 5.10).



Figura 5.10: Cena *Trees*: quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda).

Da mesma maneira, pode-se ver um comportamento nos ganhos/perdas semelhantes a cena anterior. Porém, não pode-se notar ganhos significativos na qualidade da cena, como podemos ver nas Figuras 5.11(a) e 5.11(b).



(a) Fusão espacial usando UWT-SF      (b) Decomposição em Sub-bandas Temporais

Figura 5.11: Cena *Trees*: amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)).

#### 4. *Hangar*

De maneira similar à cena *Camouflage*, a sequência de teste *Hangar* tem um plano de cena mais próximo (Figura 5.12).

Pelas características parecidas com a sequência citada, essa sequência apresenta um comportamento parecido em relação a avaliação das métricas objetivas. O ganho mais expressivo é ainda em relação ao  $M_{AB/F}$ . Nas Figuras 5.13(a) e 5.13(b) podemos notar apenas algumas mudanças no contraste, porém não há ganhos visuais expressivos.

#### 5. *Lab*

Pela proximidade do plano de cena, esta sequência apresenta características desafiadoras no que diz respeito ao processo de registro (Figura 5.14).



Figura 5.12: Cena *Hangar*: quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda).



(a) Fusão espacial usando UWT-SF      (b) Decomposição em Sub-bandas Temporais

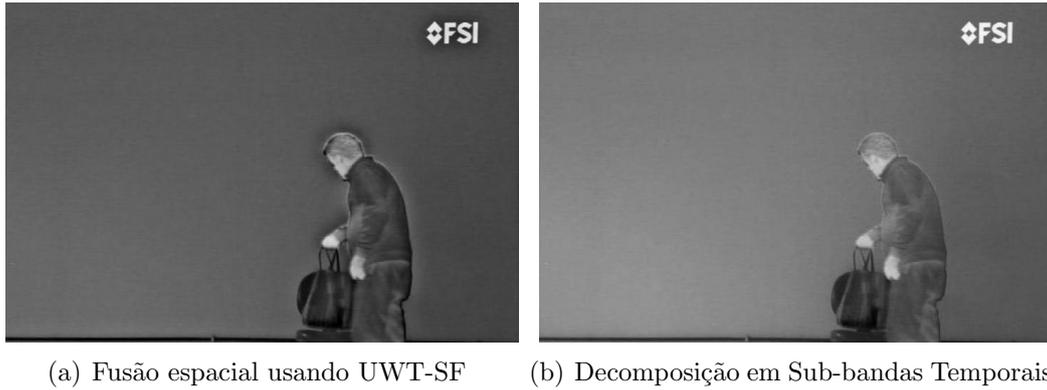
Figura 5.13: Cena *Hangar*: amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)).



Figura 5.14: Cena *Lab*: quadro 100 da sequência de luz-visível (direita) e infravermelha (esquerda).

Como pode ser visto nas Figuras da Seção C.5, a adição da informação temporal não foi eficaz na melhoria nas métricas objetivas nesta sequência. Entretanto, pode-se notar que o efeito do erro de registro foi mitigado de maneira eficaz. Este fato pode ser notado na comparação das Figuras 5.15(a) e 5.15(b).

Apesar dos resultados para as métricas  $Q_{AB/F}$ ,  $M_{AB/F}$  e  $DQ$  estarem próximos ou superiores aos obtidos para os métodos espaciais de fusão, a métrica MI-IFD obteve avaliações bem inferiores ao incluir o tempo. Este fato pode causar estra-

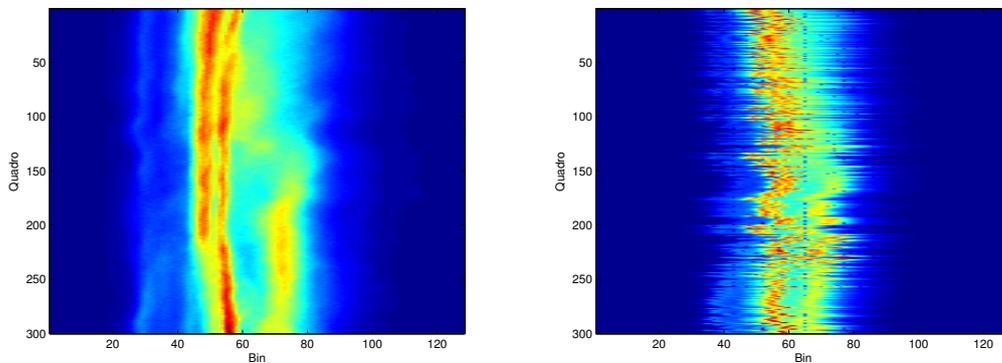


(a) Fusão espacial usando UWT-SF      (b) Decomposição em Sub-bandas Temporais

Figura 5.15: Cena *Lab*: amostra dos resultados de fusão usando a UWT-SF (Figura (a)) e a Decomposição em Sub-bandas Temporais (Figura (b)).

nhamento pois esta métrica é um desdobramento da métrica  $M_{AB/F}$  que sempre denotou melhorias nos métodos propostos.

Como foi dito, esta métrica usa a informação mútua, que por sua vez é dependente da estatística da sequência. Para analisar o efeito da decomposição temporal sobre a métrica MI-IFD, podemos comparar a variação dos histogramas dos valores dos pixels dos vídeos fundidos ao longo do tempo. Nota-se que o valor médio dos quadros algumas vezes varia abruptamente. Como pode ser visto na Figura 5.16.



(a) Método Espacial: UWT-SF      (b) Método Espaço-temporal: Decomposição Temporal

Figura 5.16: Histograma comparativo do valor médio do quadro avaliado ao longo da cena *Camouflage*. Na imagem da esquerda, é mostrado um histograma do melhor resultado segundo a métrica MI-IFD do Capítulo 3. Na direita, o melhor resultado mostrado neste Capítulo.

Esta variação causa um efeito subjetivo indesejado, conhecido como ruído *Flicker* ou cintilação. Para reduzir este efeito sem pós-processamento, devemos impor aos algoritmos usados alguma forma de controle do valor da média de cada quadro.

Como os algoritmos utilizados neste Capítulo usaram compensação de movimento na filtragem, não foi garantida a estabilidade da média, como na Fi-

gura 5.16(a). Por isso, o resultado para esta métrica foi tão inferior aos métodos espaciais.

Testes subjetivos informais mostraram que o efeito produzido por esta variação da média costuma incomodar os observadores. Este fato sugere que as métricas  $Q_{AB/F}$ ,  $M_{AB/F}$  e  $DQ$ , nestes casos, não possuem uma boa correspondência com os testes subjetivos. Além disso, como a métrica MI-IFD falha em outros casos (veja as Tabelas 5.14 e 5.15), concluimos que estes fatos proporcionam uma boa justificativa para que se continue investigando métricas adequadas para a avaliação de técnicas de fusão de vídeo.

## 5.5 Considerações Finais

Neste Capítulo foi introduzida a informação temporal nos algoritmos de fusão. Primeiramente, estendemos a ideia das transformadas *wavelet* discreta e *wavelet* não-decimada para a dimensão temporal (Seção 5.1). Vimos que os resultados deixaram a desejar na maioria dos casos. O principal motivo para isto ter acontecido é termos tratado o eixo temporal igualmente ao espacial.

Usando o arcabouço teórico mostrado no Capítulo 4, iniciamos o estudo dos algoritmos utilizando decomposição temporal. Primeiro, usamos apenas a Decomposição em Sub-bandas temporais. Após isso, acrescentamos também a decomposição espacial. Além dessas duas propostas, analisamos o estado da arte dos algoritmos de fusão de vídeos (com algumas simplificações), a MCWT.

Os vídeos testados no Capítulo 3 foram fundidos as técnicas espaço-temporais estudadas. Os resultados objetivos foram discutidos para cada cena. De maneira geral, podemos ver que sempre conseguimos ganhos para a métrica  $M_{AB/F}$ .

Alguns quadros das sequências utilizadas estão disponíveis no Apêndice C para a avaliação visual.

# Capítulo 6

## Conclusão

Ao longo desta dissertação, foram testadas diversas técnicas de fusão (espacial, temporal ou espaço-temporal) com o intuito de gerar a melhor sequência de vídeo fundida. O objetivo destas técnicas é aglomerar os diferentes tipos de conteúdo das entradas em um único vídeo. Para testar estes diferentes algoritmos, foi usado o banco de vídeos VLIRVDIF [20]. Com este banco, foi possível testar cada algoritmo em diferentes situações e com isso encontrar qual método resulta nos melhores vídeos fundidos.

Em relação à qualidade de fusão, foi visto no Capítulo 2 que, idealmente, esta deve ser mensurada subjetivamente. Tendo em vista as dificuldades deste tipo de análise, métricas objetivas de qualidade foram utilizadas.

Como base de comparação, no Capítulo 3, foram analisados resultados para algoritmos de fusão que utilizavam as transformadas *wavelet* discreta, *Contourlet* (e também a versão sem decimação) e a *wavelet* não-decimada usando fatoração espectral. Além destas técnicas multirresolução, foram testadas a média e a análise de componente principais como método de fusão.

Os estudos das técnicas espaciais de fusão iniciam os experimentos para adicionar a informação temporal aos vídeos fundidos. No Capítulo 4, a teoria sobre filtragem temporal com compensação de movimento foi apresentada. Com este desenvolvimento, no Capítulo 5, foi possível incluir esta informação através da fusão dos vetores de movimento em diferentes configurações.

Através das diversas métricas objetivas, os métodos espaço-temporais foram comparados aos métodos espaciais. Os métodos que usam as versões tridimensionais das transformadas *wavelet* discreta e *wavelet* não-decimada com fatoração espectral, no geral, não apresentaram resultados superiores às versões bidimensionais. Usando a outra abordagem, a qual denotamos como Decomposição em Sub-bandas Temporais, foram alcançados valores superiores aos dos métodos espaciais (Tabela 5.15) para a métrica  $M_{AB/F}$ . As outras métricas ficaram ligeiramente piores que os melhores resultados dos apresentados nos Capítulo 3.

Adicionando a informação espacial a este método, apresentamos a Decomposição Espaço-temporal em sub-bandas, onde a informação espacial é extraída através da aplicação das transformadas bidimensionais nas sub-bandas temporais. Este método apresentou desempenho um pouco inferior ao da Decomposição Temporal em sub-bandas. Entretanto, este método mostrou-se como o melhor para a métrica objetiva DQ.

Ainda nos métodos espaço-temporais, apresentamos uma simplificação de um método presente na literatura [59]. Na decomposição espacial, foi incluído a UWT-SF pelos resultados obtidos ao longo desta dissertação. Em geral, este método não foi superior aos anteriores.

Por fim, pode-se concluir que a adição da informação temporal nos métodos de fusão pode melhorar a qualidade do vídeo fundido. Como etapa preliminar, os estudos introduzidos nesta dissertação proporcionaram uma investigação sobre maneiras eficientes de incluir a informação temporal através do uso da compensação de movimentos. Os resultados obtidos ao longo dessa dissertação comprovaram que os métodos propostos apresentam eficácia superior, na maioria dos casos, aos melhores métodos espaciais presentes na literatura.

## 6.1 Trabalhos Futuros

Neste trabalho foram propostos dois métodos espaço-temporais que melhoram os resultados objetivos em alguns casos (de acordo com as métricas escolhidas). Ainda foi incluída uma modificação ao algoritmo proposto por Liang Xu [59] tendo em vista o método de extração dos coeficientes multirresolução.

O foco deste trabalho foi o método de decomposição em sub-bandas e também a maneira de agregar as diferentes informações de interesse. Entretanto, alguns pontos não foram abordados e podem melhorar os resultados.

Um ponto interessante é o estudo das regras de fusão (tanto das sub-bandas temporais quanto dos vetores de movimento). Foram testadas regras simples, que não incluíram nenhuma informação estrutural ou temporal das sequências de vídeo. Regras mais complexas podem implicar em melhores resultados.

Além disso, a informação presente nos vetores de movimento pode ser acrescida usando compensação bidirecional. Na maioria dos resultados presentes neste trabalho, foi usada compensação *forward*. Este fato faz com que a quantidade de pixels não-conectados seja maior.

Ainda na compensação de movimento, um ponto que pode melhorar a representação das sub-bandas temporais é o tamanho fixo dos blocos da estimação de movimento. Permitindo diferentes tamanhos de blocos, pode-se inferir de maneira mais precisa o movimento e com isso, melhorar a fusão.

Outro ponto a ser tratado em trabalhos futuros é quanto a consistência das métricas objetivas para avaliação de sequência de vídeo fundidas. Como este campo é recente, não há muitos trabalhos que as validem. Novas métricas podem ser propostas para a avaliação de vídeos de maneira diferente das usadas para imagens.

# Referências Bibliográficas

- [1] OMAR, Z., STATHAKI, T. “Image Fusion: An Overview”. In: *Intelligent Systems, Modelling and Simulation (ISMS), 2014 5th International Conference on*, pp. 306–310, Jan 2014. doi: 10.1109/ISMS.2014.58.
- [2] PETROVIĆ, V., XYDEAS, C. “Objective image fusion performance characterisation”. In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. IEEE, 2005.
- [3] ZHANG, Q., HUA, S., BLUM, R. S., et al. “Video fusion performance assessment based on spatial-temporal phase congruency”, *Signal Processing*, 2014.
- [4] “QuickBird Satellite Images”. <http://www.satimagingcorp.com/gallery/quickbird/>. Online, acesso em: 26/08/2016.
- [5] LI, S., YANG, B., HU, J. “Performance comparison of different multi-resolution transforms for image fusion”, *Information Fusion*, v. 12, n. 2, pp. 74–84, 2011.
- [6] DO, M., VETTERLI, M. “The contourlet transform: an efficient directional multiresolution image representation”, *Image Processing, IEEE Transactions on*, v. 14, n. 12, pp. 2091–2106, Dez 2005. ISSN: 1057-7149. doi: 10.1109/TIP.2005.859376.
- [7] NIKOLOV, S., HILL, P., BULL, D., et al. “Wavelets for image fusion”. In: *Wavelets in signal and image analysis*, Springer, pp. 213–241, 2001.
- [8] DA CUNHA, A., ZHOU, J., DO, M. “The Nonsampled Contourlet Transform: Theory, Design, and Applications”, *Image Processing, IEEE Transactions on*, v. 15, n. 10, pp. 3089–3101, Out 2006. ISSN: 1057-7149. doi: 10.1109/TIP.2006.877507.
- [9] PETROVIĆ, V. S., XYDEAS, C. S. “Gradient-based multiresolution image fusion”, *Image Processing, IEEE Transactions on*, v. 13, n. 2, pp. 228–237, 2004.

- [10] ZHANG, Q., WANG, L., LI, H., et al. “Similarity-based multimodality image fusion with shiftable complex directional pyramid”, *Pattern recognition letters*, v. 32, n. 13, pp. 1544–1553, 2011.
- [11] LIANG, J., HE, Y., LIU, D., et al. “Image fusion using higher order singular value decomposition”, *Image Processing, IEEE Transactions on*, v. 21, n. 5, pp. 2898–2909, 2012.
- [12] ELLMAUTHALER, A., DA SILVA, E. A., PAGLIARI, C. L., et al. “Infrared-visible image fusion using the undecimated wavelet transform with spectral factorization and target extraction”. In: *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pp. 2661–2664. IEEE, 2012.
- [13] YANG, B., LI, S. “Pixel-level image fusion with simultaneous orthogonal matching pursuit”, *Information Fusion*, v. 13, n. 1, pp. 10–19, 2012.
- [14] ZHANG, Z., BLUM, R. S. “A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application”, *Proceedings of the IEEE*, v. 87, n. 8, pp. 1315–1326, 1999.
- [15] ZHANG, Q., CHEN, Y., WANG, L. “Multisensor video fusion based on spatial-temporal salience detection”, *Signal Processing*, 2013.
- [16] TORRESAN, H., TURGEON, B., IBARRA-CASTANEDO, C., et al. “Advanced surveillance systems: combining video and thermal imagery for pedestrian detection”. In: *Defense and Security*, pp. 506–515. International Society for Optics and Photonics, 2004.
- [17] HAQ, A.-U., GONDAL, I., MURSHED, M. “Automated multi-sensor color video fusion for nighttime video surveillance”. In: *Computers and Communications (ISCC), 2010 IEEE Symposium on*, pp. 529–534. IEEE, 2010.
- [18] MEYER, C. R., BOES, J. L., KIM, B., et al. “Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations”, *Medical image analysis*, v. 1, n. 3, pp. 195–206, 1997.
- [19] HIGGINS, W. E., HELFERTY, J. P., LU, K., et al. “3D CT-video fusion for image-guided bronchoscopy”, *Computerized Medical Imaging and Graphics*, v. 32, n. 3, pp. 159–173, 2008.
- [20] ELLMAUTHALER, A., DA SILVA, E. A., PAGLIARI, C. L., et al. “VLIRV-DIF: Visible-Light and Infrared Video Database for Image Fusion”. [www.smt.ufrj.br/~fusion](http://www.smt.ufrj.br/~fusion). Online, acesso em: 26/08/2016.

- [21] TOET, A., FRANKEN, E. M. “Perceptual evaluation of different image fusion schemes”, *Displays*, 2003.
- [22] ZHANG, Q., WANG, L., LI, H., et al. “Video fusion performance evaluation based on structural similarity and human visual perception”, *Signal Processing*, 2012.
- [23] PETROVIĆ, V., COOTES, T., PAVLOVIC, R. “Dynamic image fusion performance evaluation”. In: *Information Fusion, 2007 10th International Conference on*. IEEE, Abr 2007.
- [24] ZHANG, Q., WANG, L., MA, Z., et al. “A novel video fusion framework using surfacelet transform”, *Optics Communications*, 2012.
- [25] WANG, Z., BOVIK, A. C. “A universal image quality index”, *Signal Processing Letters, IEEE*, 2002.
- [26] JAIN, A. K. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.
- [27] WANG, Q., YU, D., SHEN, Y. “An overview of image fusion metrics”. In: *Instrumentation and Measurement Technology Conference, 2009. I2MTC'09. IEEE*. IEEE, 2009.
- [28] PIELLA, G. “New quality measures for image fusion”. In: *Proceedings of the Seventh International Conference on Information Fusion*, pp. 542–546, 2004.
- [29] QU, G., ZHANG, D., YAN, P. “Information measure for performance of image fusion”, *Electronics letters*, 2002.
- [30] VIOLA, P., WELLS III, W. M. “Alignment by maximization of mutual information”, *International journal of computer vision*, 1997.
- [31] HOSSNY, M., NAHAVANDI, S., CREIGHTON, D., et al. “Image fusion metrics: evolution in a nutshell”. In: *Computer Modelling and Simulation (UKSim), 2013 UKSim 15th International Conference on*. IEEE, 2013.
- [32] GUPTA, S., RAJKUMAR, S., VIJAYARAJAN, V., et al. “Quantitative Analysis of various Image Fusion techniques based on various metrics using different Multimodality Medical Images”, *International Journal of Engineering Science & Technology*, 2013.
- [33] XYDEAS, C., PETROVIĆ, V. “Objective image fusion performance measure”, *Electronics Letters*, 2000.

- [34] SOBEL, I., FELDMAN, G. “A 3x3 isotropic gradient operator for image processing”, *a talk at the Stanford Artificial Project in*, pp. 271–272, 1968.
- [35] PETROVIĆ, V. “Subjective tests for image fusion evaluation and objective metric validation”, *Information Fusion*, 2007.
- [36] VAN DEN BRANDEN LAMBRECHT, C. J. “A working spatio-temporal model of the human visual system for image restoration and quality assessment applications”. In: *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, v. 4, pp. 2291–2294. IEEE, 1996.
- [37] ROCKINGER, O., FECHNER, T. “Pixel-level image fusion: the case of image sequences”. In: *Aerospace/Defense Sensing and Controls. International Society for Optics and Photonics*, Abr 1998.
- [38] ZHANG, X., LI, X., FENG, Y., et al. “The use of ROC and AUC in the validation of objective image fusion evaluation metrics”, *Signal Processing*, v. 115, pp. 38–48, 2015.
- [39] DAUBECHIES, I., OTHERS. *Ten lectures on wavelets*, v. 61. SIAM, 1992.
- [40] STARCK, J.-L., CANDÈS, E. J., DONOHO, D. L. “The curvelet transform for image denoising”. v. 11, pp. 670–684. IEEE, 2002.
- [41] FOWLER, J. E. “The redundant discrete wavelet transform and additive noise”, *Signal Processing Letters, IEEE*, v. 12, n. 9, pp. 629–632, 2005.
- [42] ELLMAUTHALER, A. *Multiscale Image Fusion*. Tese de Doutorado, Programa de Engenharia Elétrica, COPPE, Universidade Federal do Rio de Janeiro, 2013.
- [43] MALLAT, S. *A wavelet tour of signal processing*. Academic press, 1999.
- [44] OLSHAUSEN, B. A., OTHERS. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, *Nature*, 1996.
- [45] MALLAT, S. “A theory for multiresolution signal decomposition: the wavelet representation”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Jul 1989.
- [46] RIOUL, O., VETTERLI, M. “Wavelets and signal processing”, *Signal Processing Magazine, IEEE*, Out 1991.

- [47] PIELLA, G. *Adaptive Wavelets and their Applications to Image Fusion and Compression*. Tese de Doutorado, Department of Mathematics, University of Amsterdam, 2003., 2003.
- [48] DAUBECHIES, I. “The wavelet transform, time-frequency localization and signal analysis”, *IEEE Transactions on Information Theory*, v. 36, n. 5, pp. 961–1005, Set 1990. ISSN: 0018-9448. doi: 10.1109/18.57199.
- [49] DINIZ, P. S., DA SILVA, E. A., NETTO, S. L. *Digital signal processing: system analysis and design*. Cambridge University Press, 2010.
- [50] BURT, P. J., ADELSON, E. H. “The Laplacian pyramid as a compact image code”, *Communications, IEEE Transactions on*, v. 31, n. 4, pp. 532–540, 1983.
- [51] DO, M. “Contourlet Toolbox on Matlab”. 2005. Available on: <<http://www.mathworks.com/matlabcentral/fileexchange/8837-contourlet-toolbox>>. Online, acesso em: 16/11/2015.
- [52] CUNHA, A. “Nonsubsampled Contourlet Transform”. 2006. Available on: <<http://www.mathworks.com/matlabcentral/fileexchange/10049-nonsubsampled-contourlet-toolbox>>. Online, acesso em: 16/11/2015.
- [53] STARCK, J.-L., FADILI, J., MURTAGH, F. “The undecimated wavelet decomposition and its reconstruction”, *Image Processing, IEEE Transactions on*, v. 16, n. 2, pp. 297–309, 2007.
- [54] ELLMAUTHALER, A., PAGLIARI, C. L., DA SILVA, E. A. “Multiscale image fusion using the undecimated wavelet transform with spectral factorization and nonorthogonal filter banks”, *Image Processing, IEEE Transactions on*, v. 22, n. 3, pp. 1005–1017, 2013.
- [55] <http://vcipl-okstate.org/pbvs/bench/>, . Online, acesso em: 26/08/2016.
- [56] <http://www.cis.rit.edu/pelz/scanpaths/data/bristol-eden.htm>, . Online, acesso em: 09/11/2015.
- [57] <http://www.polymtl.ca/litiv/vid/>, . Online, acesso em: 09/11/2015.
- [58] ELLMAUTHALER, A., DA SILVA, E. A., PAGLIARI, C. L., et al. “A database of registered visible-light and infrared videos built using a novel spatiotemporal registration technique”. IEEE, 2016. Não publicado.

- [59] XU, L., DU, J., ZHANG, Z. “Infrared-visible video fusion based on motion-compensated wavelet transforms”, *IET Image Processing*, v. 9, n. 4, pp. 318–328, 2015.
- [60] PAJARES, G., DE LA CRUZ, J. M. “A wavelet-based image fusion tutorial”, *Pattern recognition*, v. 37, n. 9, pp. 1855–1872, 2004.
- [61] ZHANG, Y. “Understanding image fusion”, *Photogrammetric engineering and remote sensing*, v. 70, n. 6, pp. 657–661, 2004.
- [62] MATHWORKS, M. “Wavelet Filters”. 2016. Available on: <<http://www.mathworks.com/help/wavelet/ref/wfilters.html>>. Online, accessed: 11/07/2016.
- [63] COHEN, A., DAUBECHIES, I., FEAUVEAU, J.-C. “Biorthogonal bases of compactly supported wavelets”, *Communications on pure and applied mathematics*, v. 45, n. 5, pp. 485–560, 1992.
- [64] HAAR, A. “Zur theorie der orthogonalen funktionensysteme”, *Mathematische Annalen*, v. 69, n. 3, pp. 331–371, 1910.
- [65] PHOONG, S.-M., KIM, C. W., VAIDYANATHAN, P., et al. “A new class of two-channel biorthogonal filter banks and wavelet bases”, *Signal Processing, IEEE Transactions on*, v. 43, n. 3, pp. 649–665, 1995.
- [66] ABDELNOUR, A., SELESNICK, I. “Design of 2-band orthogonal near-symmetric CQF”. In: *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, v. 6, pp. 3693–3696 vol.6, 2001. doi: 10.1109/ICASSP.2001.940644.
- [67] LI, J., NIKOLOV, S. G., BENTON, C. P., et al. “Motion-based video fusion using optical flow information”. In: *Information Fusion, 2006 9th International Conference on*, pp. 1–8. IEEE, 2006.
- [68] LIU, C., JING, Z., XIAO, G., et al. “Feature-based fusion of infrared and visible dynamic images using target detection”, *Chinese optics letters*, v. 5, n. 5, pp. 274–277, 2007.
- [69] ZHANG, Q., WANG, Y., LEVINE, M. D., et al. “Multisensor video fusion based on higher order singular value decomposition”, *Information Fusion*, v. 24, pp. 54–71, 2015.
- [70] XU, L., DU, J., ZHANG, Z. “Image sequence fusion and denoising based on 3D shearlet transform”, *Journal of Applied Mathematics*, v. 2014, 2014.

- [71] OF ITU., T. S. S. *ITU-T Recommendation H. 261: Line Transmission of Non-Telephone Signals: Video Codec for Audiovisual Services at PX 64 Kbits*. International Telecommunication Union, 1993.
- [72] FLIERL, M., GIROD, B. *Video coding with superimposed motion-compensated signals: applications to H. 264 and beyond*. Springer Science & Business Media, 2004.
- [73] OHM, J.-R. “Three-dimensional subband coding with motion compensation”, *Image Processing, IEEE Transactions on*, v. 3, n. 5, pp. 559–571, Set 1994. ISSN: 1057-7149. doi: 10.1109/83.334985.
- [74] NAVEEN, T., WOODS, J. “Motion compensated multiresolution transmission of high definition video”, *Circuits and Systems for Video Technology, IEEE Transactions on*, v. 4, n. 1, pp. 29–41, Fev 1994. ISSN: 1051-8215. doi: 10.1109/76.276170.
- [75] CHOI, S.-J., WOODS, J. “Motion-compensated 3-D subband coding of video”, *Image Processing, IEEE Transactions on*, v. 8, n. 2, pp. 155–167, Fev 1999. ISSN: 1057-7149. doi: 10.1109/83.743851.
- [76] HSIANG, S.-T., WOODS, J., OHM, J.-R. “Invertible temporal subband/wavelet filter banks with half-pixel-accurate motion compensation”, *Image Processing, IEEE Transactions on*, v. 13, n. 8, pp. 1018–1028, Ago 2004. ISSN: 1057-7149. doi: 10.1109/TIP.2004.828421.
- [77] GOLWELKAR, A., WOODS, J. W. “Motion-compensated temporal filtering and motion vector coding using biorthogonal filters”, *Circuits and Systems for Video Technology, IEEE Transactions on*, v. 17, n. 4, pp. 417–428, 2007.
- [78] CHEN, P., WOODS, J. W. “Bidirectional MC-EZBC with lifting implementation”, *Circuits and Systems for Video Technology, IEEE Transactions on*, v. 14, n. 10, pp. 1183–1194, 2004.
- [79] WOODS, J. W. *Multidimensional signal, image, and video processing and coding*. Academic press, 2011.
- [80] DING, W., WU, F., LI, S. “Lifting-based wavelet transform with directionally spatial prediction”. In: *Picture Coding Symposium*, v. 62, pp. 291–294, 2004.

- [81] SWELDENS, W. “Lifting scheme: a new philosophy in biorthogonal wavelet constructions”. In: *SPIE’s 1995 International Symposium on Optical Science, Engineering, and Instrumentation*, pp. 68–79. International Society for Optics and Photonics, 1995.
- [82] SOLÉ ROJALS, J., OTHERS. *Optimization and generalization of lifting Schemes: application to lossless image compression*. Universitat Politècnica de Catalunya, 2006.
- [83] DAUBECHIES, I., SWELDENS, W. “Factoring wavelet transforms into lifting steps”, *Journal of Fourier analysis and applications*, v. 4, n. 3, pp. 247–269, 1998.
- [84] LEE, C. S., LEE, C. K., YOO, K. Y. “New lifting based structure for undecimated wavelet transform”, *Electronics Letters*, v. 36, n. 22, pp. 1894–1895, 2000.
- [85] GOLWELKAR, A. *Motion compensated temporal filtering and motion vector coding using longer filters*. Tese de Doutorado, Rensselaer Polytechnic Institute, 2004.
- [86] MATHWORKS, M. “Lifting schemes on Matlab”. 2015. Available on: <<http://www.mathworks.com/help/wavelet/ref/liftwave.html>>. Online, acesso em: 15/10/2015.
- [87] CHAN, S. H., VÕ, D. T., NGUYEN, T. Q. “Subpixel motion estimation without interpolation”. In: *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 722–725. IEEE, 2010.

# Apêndice A

## Validação das métricas espaço-temporais

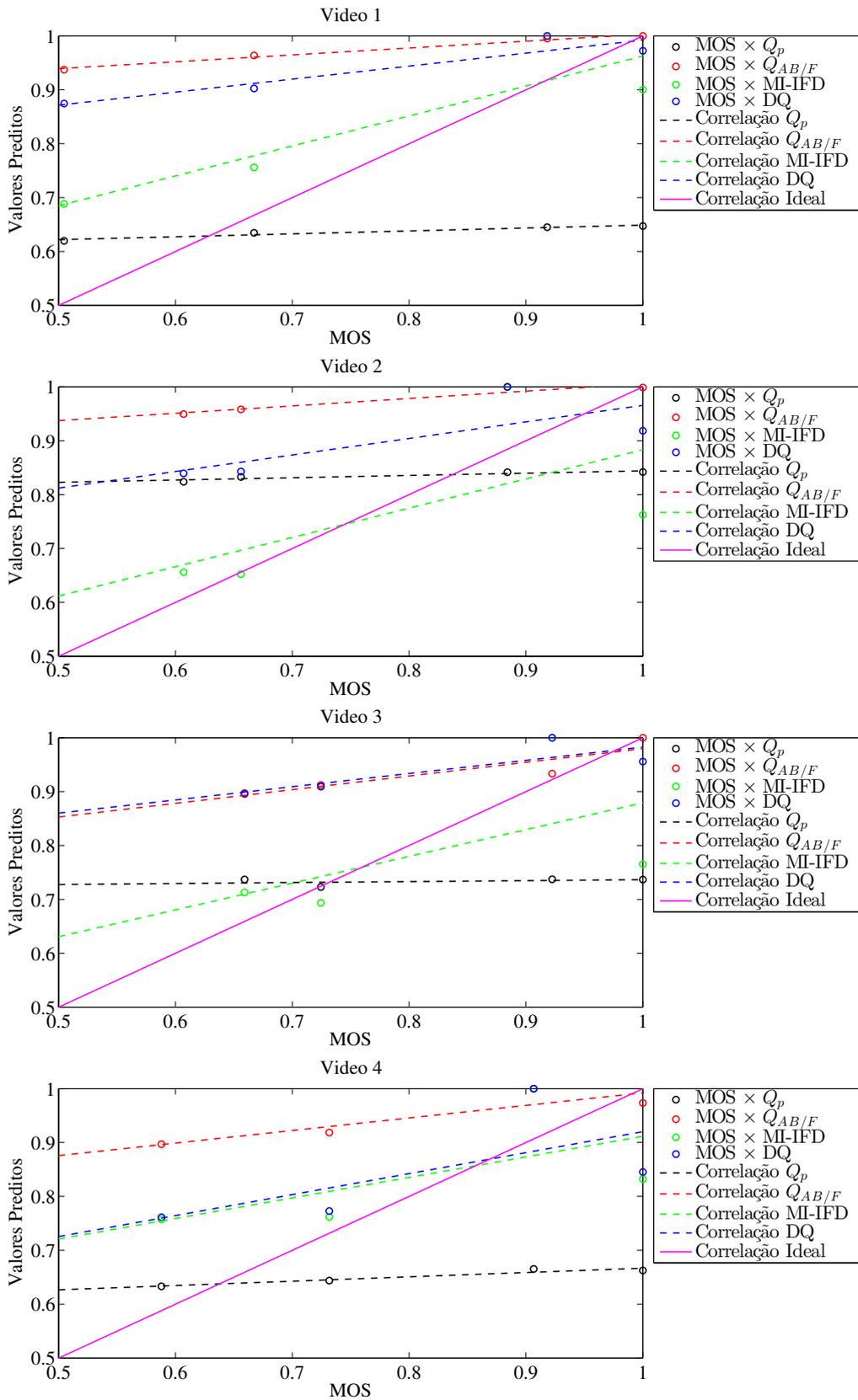
No trabalho de Zhang *et al.* [22], foram realizados testes subjetivos com 12 avaliadores que julgaram 4 vídeos fundidos proveniente de 4 diferentes algoritmos. Os avaliadores foram questionados sobre 3 pontos: qualidade espacial, temporal e espaço-temporal.

Como este teste obteve poucos dados para uma análise mais profunda, a correlação entre os resultados subjetivos e os objetivos foi obtida através da inclinação da reta formada por ambas as medidas após ajuste linear (método dos mínimos quadrados).

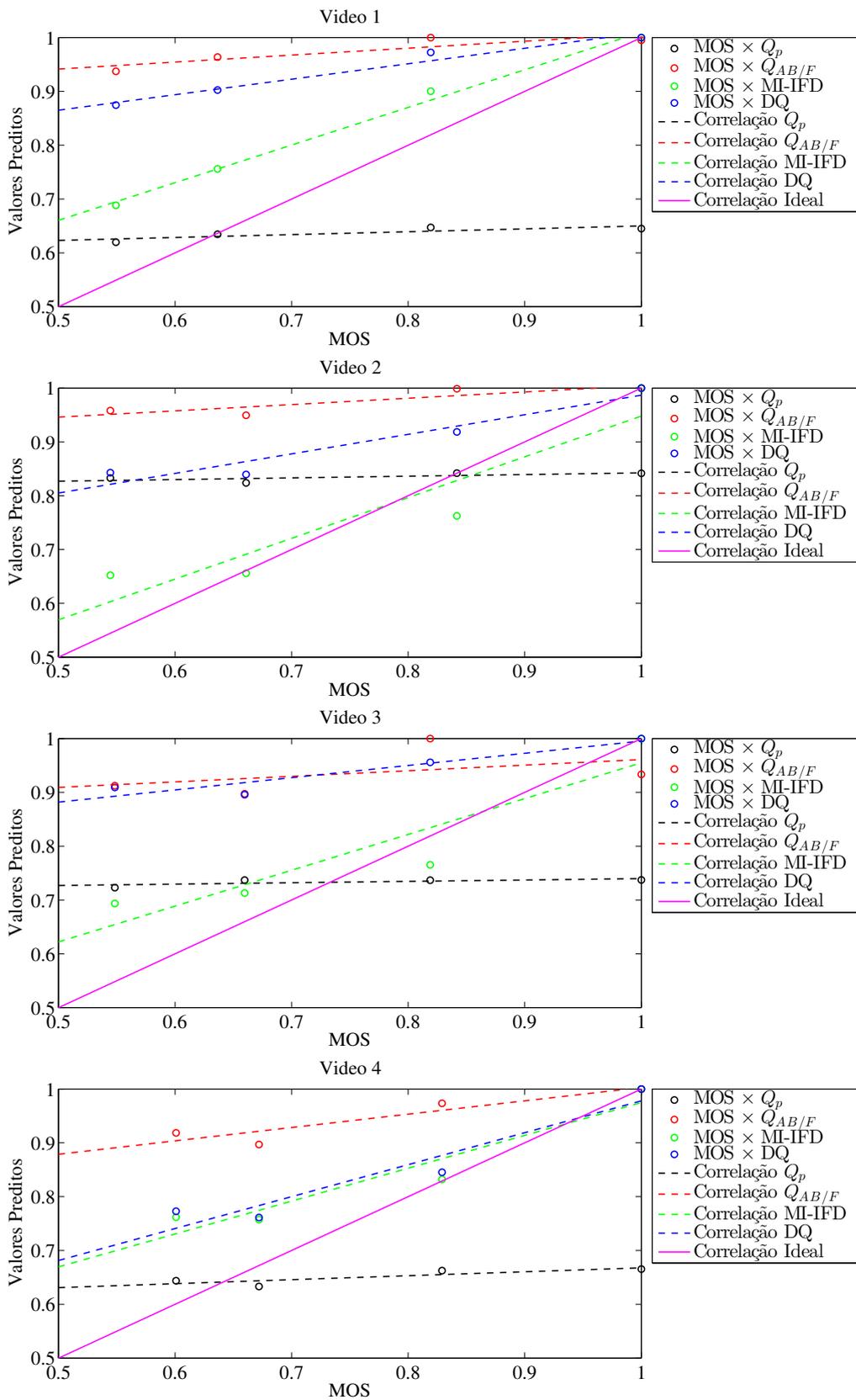
No eixo horizontal dos gráficos abaixo estão os resultados subjetivos. No eixo vertical, estão os resultados objetivos. Os pares ordenados foram normalizados de maneira que a máxima correlação seja equivalente a inclinação unitária.

Os gráficos foram divididos aqui pela pergunta direcionada aos avaliadores. Na Seção A.1 estão os resultados da validação feita para a qualidade espacial; na Seção A.2, para a qualidade temporal; e na Seção A.3, para a qualidade espaço-temporal.

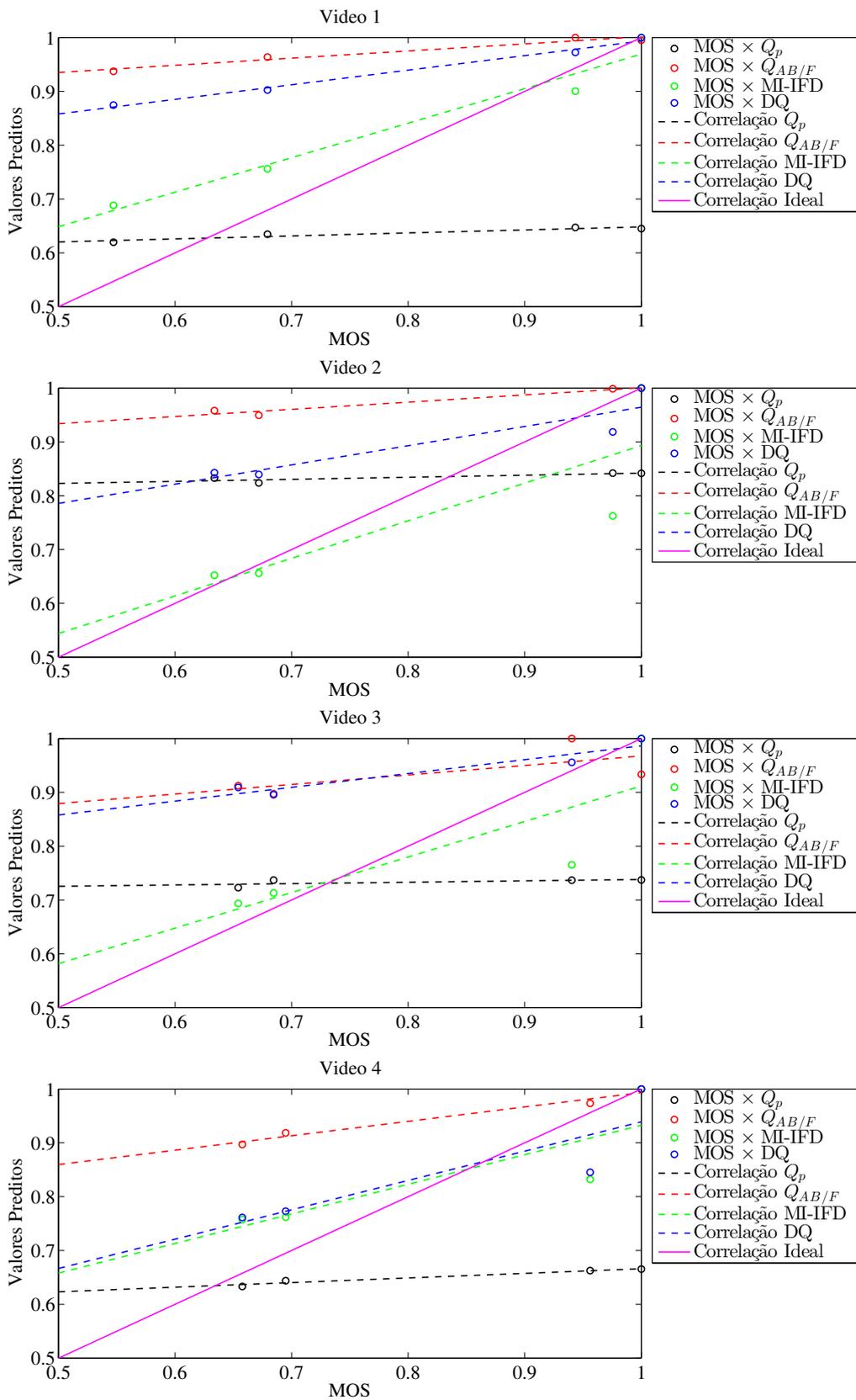
## A.1 Avaliação Espacial



## A.2 Avaliação Temporal



## A.3 Avaliação Espaço-temporal



# Apêndice B

## Resultados de Avaliação Objetiva dos Vídeos Fundidos

### B.1 Decomposição em Sub-bandas Temporais

<i>Guanabara Bay</i> - LS usando banco de filtros Haar					
MVR	Regra Temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	maxmean	0.481494	0.418961	0.387398	0.016308
max	maxmean	0.476494	0.399862	0.406028	0.028992
min	maxmean	0.476037	0.435569	0.378065	0.058531
ir	maxmean	0.467669	0.409007	0.375912	0.015375
vis	maxmean	0.491924	0.426030	0.407423	0.038077
mean	meanmean	0.442950	0.273087	0.396825	0.280064
max	meanmean	0.442965	0.273079	0.396829	0.280128
min	meanmean	0.442944	0.273083	0.396826	0.280090
ir	meanmean	0.442951	0.273083	0.396826	0.280173
vis	meanmean	0.442945	0.273085	0.396827	0.280082

Tabela B.1: Fusão da cena *Guanabara Bay* computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

<i>Trees</i> - LS usando banco de filtros Haar					
MVR	Regra Temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	maxmean	0.512631	0.413970	0.406866	0.062125
max	maxmean	0.493351	0.392111	0.409801	0.063637
min	maxmean	0.542852	0.439540	0.406071	0.114150
ir	maxmean	0.503953	0.405804	0.411976	0.068200
vis	maxmean	0.531188	0.424848	0.403868	0.090874
mean	meanmean	0.475477	0.369555	0.423555	0.500539
max	meanmean	0.475478	0.369555	0.423555	0.500483
min	meanmean	0.475478	0.369556	0.423555	0.500541
ir	meanmean	0.475479	0.369555	0.423556	0.500549
vis	meanmean	0.475475	0.369556	0.423555	0.500515

Tabela B.2: Fusão da cena *Trees* computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

<i>Hangar</i> - LS usando banco de filtros Haar					
MVR	Regra Temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	maxmean	0.311247	0.179175	0.292041	0.055320
max	maxmean	0.297012	0.173012	0.291066	0.045362
min	maxmean	0.297699	0.191785	0.279525	0.056312
ir	maxmean	0.296962	0.187119	0.279599	0.063002
vis	maxmean	0.302817	0.174676	0.293412	0.037948
mean	meanmean	0.524773	0.368890	0.427665	0.900837
max	meanmean	0.524754	0.368883	0.427663	0.900884
min	meanmean	0.524754	0.368886	0.427663	0.900947
ir	meanmean	0.524758	0.368887	0.427664	0.901050
vis	meanmean	0.524753	0.368880	0.427663	0.900791

Tabela B.3: Fusão da cena *Hangar* computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

<i>Lab</i> - LS usando banco de filtros Haar					
MVR	Regra Temporal	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
mean	maxmean	0.457653	0.255802	0.388045	0.046221
max	maxmean	0.471227	0.251602	0.386059	0.056355
min	maxmean	0.411398	0.266210	0.365633	0.039700
ir	maxmean	0.412676	0.256854	0.368547	0.038508
vis	maxmean	0.470349	0.256254	0.385724	0.049655
mean	meanmean	0.537901	0.362498	0.418668	0.573925
max	meanmean	0.537917	0.362492	0.418671	0.574084
min	meanmean	0.537911	0.362495	0.418670	0.574343
ir	meanmean	0.537910	0.362496	0.418670	0.574043
vis	meanmean	0.537914	0.362496	0.418671	0.574133

Tabela B.4: Fusão da cena *Lab* computada através da decomposição temporal usando lifting e o banco de filtros Haar. Os vetores de movimento foram obtidos com tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

## B.2 Decomposição em Sub-bandas Espaço-temporais

<i>Guanabara Bay</i> - Resultados da fusão espaço-temporal					
Filtro	MVR	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	mean	0.456369	0.229031	0.401050	0.194261
	max	0.456013	0.228382	0.400058	0.152918
	min	0.456019	0.228213	0.400925	0.208771
	ir	0.456585	0.228999	0.400862	0.157783
	vis	0.455479	0.227791	0.400105	0.206264
5-3	mean	0.449366	0.224742	0.398975	0.192941
	max	0.448630	0.223382	0.396216	0.133276
	min	0.449518	0.224117	0.399763	0.215765
	ir	0.449675	0.224501	0.399637	0.177322
	vis	0.448258	0.223231	0.396306	0.186509
$B_3$ -spline	mean	0.459261	0.229955	0.402419	0.215608
	max	0.459210	0.229649	0.401857	0.161744
	min	0.458968	0.229205	0.402351	0.232454
	ir	0.459627	0.230022	0.402397	0.169826
	vis	0.458487	0.229063	0.401841	0.217489

Tabela B.5: Fusão da cena *Guanabara Bay* computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

<i>Hangar</i> - Resultados da fusão espaço-temporal					
Filtro	MVR	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	mean	0.530892	0.269444	0.420781	0.397645
	max	0.528421	0.267830	0.419374	0.342325
	min	0.530109	0.269399	0.420335	0.417167
	ir	0.529564	0.268907	0.420325	0.430878
	vis	0.529824	0.268159	0.419017	0.325630
5-3	mean	0.522561	0.272048	0.413076	0.332782
	max	0.518367	0.268675	0.410610	0.252570
	min	0.521687	0.272033	0.413487	0.382222
	ir	0.520857	0.271226	0.413007	0.391960
	vis	0.520606	0.269670	0.411544	0.251798
$B_3$ -spline	mean	0.533689	0.265937	0.422714	0.430963
	max	0.531843	0.264930	0.421333	0.386090
	min	0.532984	0.265734	0.421800	0.441091
	ir	0.532491	0.265418	0.422188	0.433596
	vis	0.532951	0.265066	0.421004	0.360116

Tabela B.6: Fusão da cena *Hangar* computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

<i>Trees</i> - Resultados da fusão espaço-temporal					
Filtro	MVR	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	mean	0.468535	0.347295	0.427667	0.424962
	max	0.468227	0.346427	0.427562	0.402893
	min	0.468138	0.347066	0.427510	0.436139
	ir	0.468104	0.347450	0.427548	0.420733
	vis	0.468341	0.346031	0.427581	0.431459
5-3	mean	0.465498	0.349831	0.427407	0.433353
	max	0.464522	0.348676	0.427150	0.420867
	min	0.465140	0.349486	0.427290	0.453357
	ir	0.465105	0.350027	0.427331	0.437584
	vis	0.464686	0.348178	0.427136	0.445484
$B_3$ -spline	mean	0.469639	0.342502	0.427716	0.422299
	max	0.469567	0.341952	0.427762	0.417569
	min	0.469243	0.342205	0.427567	0.439495
	ir	0.469193	0.342635	0.427688	0.426594
	vis	0.469600	0.341524	0.427650	0.427226

Tabela B.7: Fusão da cena *Trees* computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

<i>Lab</i> - Resultados da fusão espaço-temporal					
Filtro	MVR	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
Haar Starck 2	mean	0.548033	0.299942	0.425077	0.269581
	max	0.545854	0.298010	0.423893	0.177382
	min	0.546887	0.300234	0.424739	0.290313
	ir	0.546454	0.299611	0.424806	0.273947
	vis	0.547048	0.298475	0.424385	0.194929
5-3	mean	0.543811	0.302060	0.423578	0.180414
	max	0.540787	0.298902	0.421343	0.110339
	min	0.542861	0.301985	0.423891	0.239352
	ir	0.542209	0.301455	0.423627	0.217884
	vis	0.542603	0.299403	0.422127	0.123519
$B_3$ -spline	mean	0.547554	0.291778	0.425134	0.294758
	max	0.545359	0.290576	0.423832	0.195996
	min	0.546512	0.291911	0.424912	0.308970
	ir	0.546121	0.291538	0.424734	0.278636
	vis	0.546235	0.290873	0.424469	0.197992

Tabela B.8: Fusão da cena *Lab* computada através da decomposição espaço-temporal, tamanho de bloco igual a 4 e acurácia de 1/2 pixel.

## B.3 Extensão Temporal dos Coeficientes Multirresolução

<i>Camouflage</i>					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
9-7	4	0.489602	0.135104	0.417313	0.084623
	8	0.500601	0.137297	0.421235	0.230042
	16	0.503902	0.138284	0.422085	0.284434
Haar starck1	4	0.361148	0.088415	0.394230	0.037390
	8	0.360561	0.089889	0.395230	0.049987
	16	0.359971	0.090942	0.394988	0.046652
Haar starck2	4	0.497643	0.133267	0.414556	0.067969
	8	0.509797	0.136390	0.418541	0.135896
	16	0.512745	0.137750	0.419708	0.162221
isotropic 1	4	0.488834	0.133124	0.415505	0.072835
	8	0.499987	0.135386	0.420013	0.223124
	16	0.503133	0.136492	0.421372	0.307750
isotropic 2	4	0.498618	0.131212	0.414126	0.049607
	8	0.511949	0.134171	0.419871	0.163346
	16	0.515282	0.135470	0.421420	0.199552
Le Gall 5-3	4	0.498682	0.137703	0.418946	0.097500
	8	0.510179	0.140149	0.422146	0.226486
	16	0.513273	0.141201	0.423142	0.252690
near_sym_a	4	0.493733	0.137113	0.418590	0.080316
	8	0.504701	0.139166	0.422531	0.257583
	16	0.507848	0.140065	0.423263	0.332224
near_sym_b	4	0.488403	0.134822	0.417078	0.082499
	8	0.499602	0.136871	0.421042	0.216794
	16	0.503017	0.137824	0.421824	0.271652
$B_1$ -spline	4	0.500846	0.134838	0.417814	0.072667
	8	0.512664	0.137437	0.422021	0.196575
	16	0.515869	0.138594	0.423195	0.262104
$B_3$ -spline	4	0.502622	0.132724	0.418282	0.077539
	8	0.514536	0.135192	0.421970	0.205042
	16	0.517908	0.136317	0.423476	0.279874

Tabela B.9: Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária.

<i>Guanabara Bay</i>					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
9-7	4	0.430139	0.198835	0.391733	0.048168
	8	0.437259	0.206338	0.396001	0.137424
	16	0.440191	0.208910	0.398352	0.136840
Haar starck1	4	0.354475	0.150608	0.385428	0.007836
	8	0.350890	0.156099	0.387259	0.008869
	16	0.347640	0.160209	0.385936	0.011206
Haar starck2	4	0.432122	0.200398	0.386471	0.033549
	8	0.440807	0.211338	0.387661	0.091914
	16	0.446393	0.216198	0.391230	0.128911
isotropic 1	4	0.429521	0.202363	0.388905	0.040220
	8	0.437612	0.210957	0.397900	0.124910
	16	0.442102	0.214797	0.400934	0.167737
isotropic 2	4	0.436247	0.203688	0.384485	0.045202
	8	0.444606	0.213532	0.389959	0.111796
	16	0.449921	0.218149	0.393341	0.142115
Le Gall 5-3	4	0.434499	0.203362	0.388050	0.045215
	8	0.443060	0.211648	0.392211	0.118936
	16	0.446358	0.214601	0.395439	0.154642
near_sym_a	4	0.432510	0.202590	0.388594	0.043488
	8	0.439171	0.210451	0.392499	0.140749
	16	0.441910	0.213234	0.395348	0.160734
near_sym_b	4	0.430458	0.198299	0.393962	0.061691
	8	0.437511	0.205151	0.397973	0.140200
	16	0.440093	0.207330	0.400129	0.140122
$B_1$ -spline	4	0.436842	0.204077	0.384576	0.035528
	8	0.444134	0.213637	0.389901	0.112553
	16	0.448594	0.217667	0.392987	0.136020
$B_3$ -spline	4	0.440103	0.206537	0.383453	0.041811
	8	0.447650	0.215515	0.391547	0.125347
	16	0.451904	0.219511	0.394624	0.146717

Tabela B.10: Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária.

<i>Trees</i>					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
9-7	4	0.446060	0.337980	0.405348	0.087891
	8	0.456659	0.344573	0.420627	0.199587
	16	0.463369	0.347735	0.425979	0.348869
Haar starck1	4	0.322743	0.199653	0.391177	0.007594
	8	0.323129	0.202207	0.392361	0.007673
	16	0.324139	0.203927	0.393089	0.006944
Haar starck2	4	0.445828	0.330508	0.397029	0.071417
	8	0.457447	0.339521	0.415976	0.127099
	16	0.466418	0.345128	0.424364	0.323978
isotropic 1	4	0.447598	0.335793	0.406604	0.087989
	8	0.458830	0.342696	0.419917	0.168498
	16	0.466360	0.346549	0.425592	0.354731
isotropic 2	4	0.447308	0.328177	0.397602	0.072042
	8	0.459005	0.336109	0.416283	0.137629
	16	0.467715	0.340634	0.424248	0.329010
Le Gall 5-3	4	0.446124	0.337339	0.402236	0.081610
	8	0.456974	0.344661	0.418855	0.165122
	16	0.464145	0.348427	0.425475	0.337507
near_sym_a	4	0.447777	0.340158	0.410191	0.093718
	8	0.458210	0.346272	0.421678	0.216768
	16	0.464696	0.349229	0.426127	0.356305
near_sym_b	4	0.446167	0.338356	0.406988	0.090437
	8	0.456691	0.344559	0.421128	0.214783
	16	0.463066	0.347445	0.426103	0.348166
$B_1$ -spline	4	0.449076	0.334643	0.402876	0.081851
	8	0.459910	0.342175	0.418209	0.148787
	16	0.467392	0.346467	0.425030	0.334131
$B_3$ -spline	4	0.450215	0.330462	0.403344	0.079777
	8	0.461095	0.337356	0.417756	0.148767
	16	0.468417	0.341091	0.424843	0.328840

Tabela B.11: Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária.

<i>Hangar</i>					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
9-7	4	0.421025	0.226704	0.373661	0.104361
	8	0.439360	0.235198	0.383980	0.183466
	16	0.449421	0.238752	0.383459	0.202911
Haar starck1	4	0.297467	0.142708	0.324093	0.031190
	8	0.300471	0.144654	0.327621	0.035463
	16	0.297371	0.146689	0.323118	0.032293
Haar starck2	4	0.413924	0.215867	0.379787	0.118890
	8	0.434961	0.224794	0.389235	0.189085
	16	0.448952	0.229695	0.388793	0.152688
isotropic 1	4	0.421906	0.224902	0.381016	0.135719
	8	0.440578	0.232987	0.388546	0.202005
	16	0.452789	0.236809	0.388388	0.173435
isotropic 2	4	0.419046	0.217019	0.381846	0.129913
	8	0.440170	0.225216	0.391876	0.213492
	16	0.454255	0.229454	0.391106	0.152598
Le Gall 5-3	4	0.419826	0.225105	0.376479	0.115625
	8	0.438679	0.234107	0.388044	0.187284
	16	0.450199	0.238439	0.386180	0.185068
near_sym_a	4	0.425986	0.231729	0.378034	0.115760
	8	0.443850	0.239941	0.388035	0.202598
	16	0.453263	0.243229	0.385967	0.221357
near_sym_b	4	0.421623	0.227334	0.373924	0.100849
	8	0.440157	0.235420	0.383609	0.187806
	16	0.449486	0.238478	0.383095	0.206635
$B_1$ -spline	4	0.425018	0.223700	0.381939	0.143541
	8	0.444397	0.232079	0.391888	0.194358
	16	0.456710	0.236096	0.390955	0.163854
$B_3$ -spline	4	0.427929	0.222213	0.385491	0.143425
	8	0.448478	0.230061	0.393703	0.209244
	16	0.460260	0.233444	0.392947	0.158390

Tabela B.12: Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária.

<i>Lab</i>					
Filtro	Bloco	$Q_{AB/F}$	$M_{AB/F}$	$DQ$	$MI - IFD$
9-7	4	0.455830	0.268463	0.405457	0.080127
	8	0.465568	0.272041	0.406022	0.076888
	16	0.461872	0.271398	0.406736	0.07078
Antonini	4	0.489602	0.135104	0.417313	0.084623
	8	0.500599	0.137297	0.421235	0.230042
	16	0.503902	0.138283	0.422085	0.284433
Haar starck1	4	0.347404	0.212987	0.385723	0.030584
	8	0.348668	0.215496	0.388396	0.030708
	16	0.344278	0.218284	0.387208	0.028426
Haar starck2	4	0.454934	0.260140	0.404931	0.080886
	8	0.468517	0.264730	0.402201	0.071762
	16	0.468573	0.266215	0.402752	0.079513
isotropic 2	4	0.456516	0.256569	0.406806	0.088106
	8	0.469800	0.260620	0.404303	0.074934
	16	0.469293	0.261794	0.405598	0.083409
isotropic 1	4	0.458866	0.268827	0.406168	0.090652
	8	0.469523	0.272419	0.405707	0.074435
	16	0.467181	0.272301	0.406267	0.075685
Le Gall 5-3	4	0.449733	0.270867	0.405650	0.088107
	8	0.461163	0.275116	0.406134	0.082960
	16	0.461835	0.274969	0.406324	0.080371
near_sym_a	4	0.458690	0.275454	0.406343	0.086809
	8	0.468335	0.278896	0.407517	0.082905
	16	0.464441	0.278092	0.407487	0.077517
near_sym_b	4	0.457195	0.268022	0.405829	0.080775
	8	0.466863	0.271297	0.406684	0.073993
	16	0.461800	0.270386	0.406861	0.072481
$B_1$ -spline	4	0.459997	0.267865	0.407427	0.099675
	8	0.470745	0.271864	0.405295	0.081826
	16	0.469716	0.272288	0.406898	0.079978
$B_3$ -spline	4	0.461455	0.261919	0.408193	0.106888
	8	0.473507	0.265674	0.406147	0.084676
	16	0.471768	0.265988	0.407853	0.084210

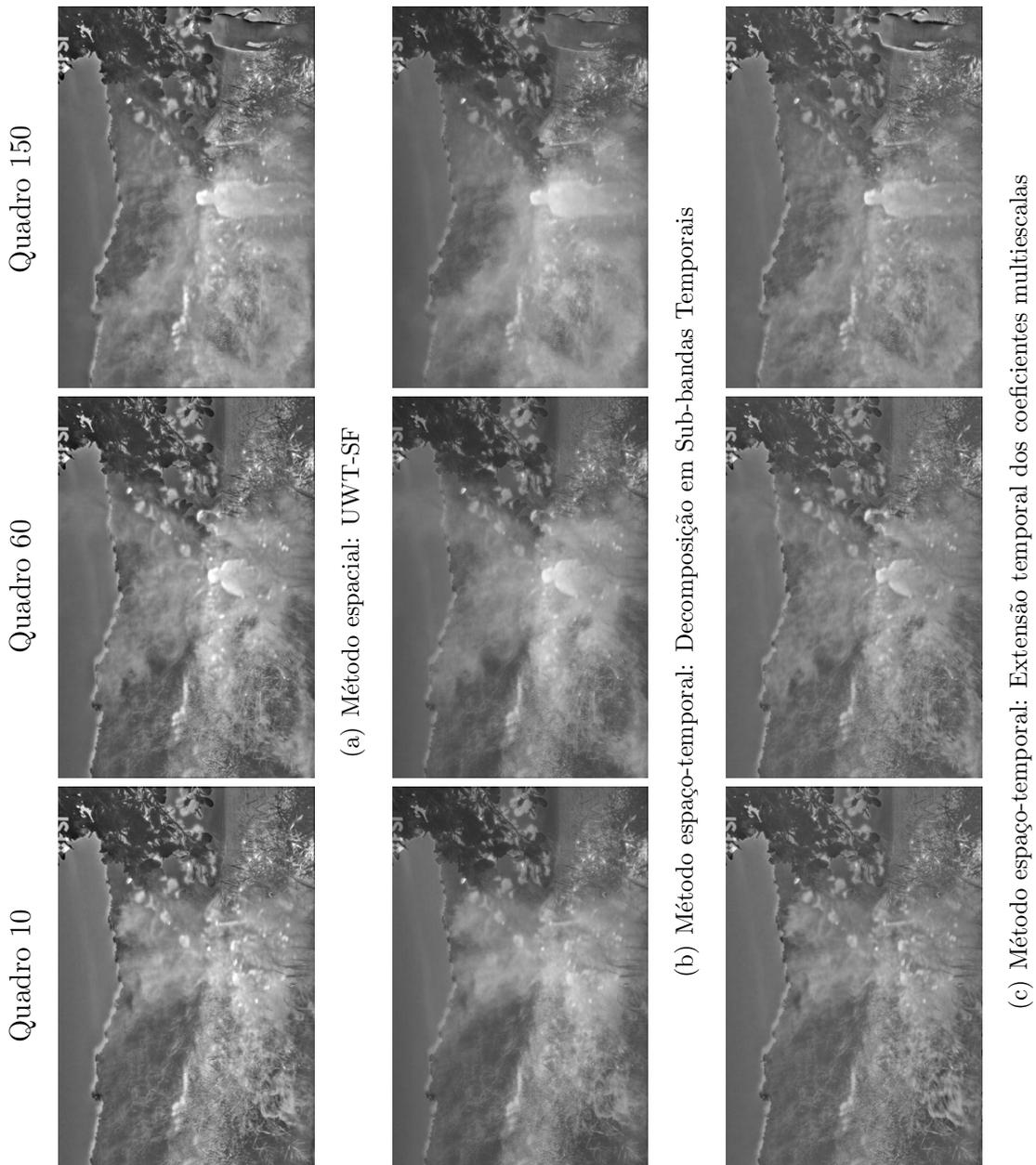
Tabela B.13: Todos os resultados usando a extensão temporal dos coeficientes da UWT. Neste experimento foram usados os tamanhos de bloco 4, 8 e 16 com acurácia unitária.



# Apêndice C

## Amostras Fundidas

### C.1 *Camouflage*



## C.2 Guanabara Bay

Quadro 10



Quadro 60



Quadro 150



(d) Método espacial: UWT-SF



(e) Método espaço-temporal: Decomposição em Sub-bandas Temporais

### C.3 *Trees*

Quadro 10



Quadro 60



Quadro 150



(f) Método espacial: UWT-SF



(g) Método espaço-temporal: Decomposição em Sub-bandas Temporais

## C.4 *Hangar*

Quadro 10

Quadro 60

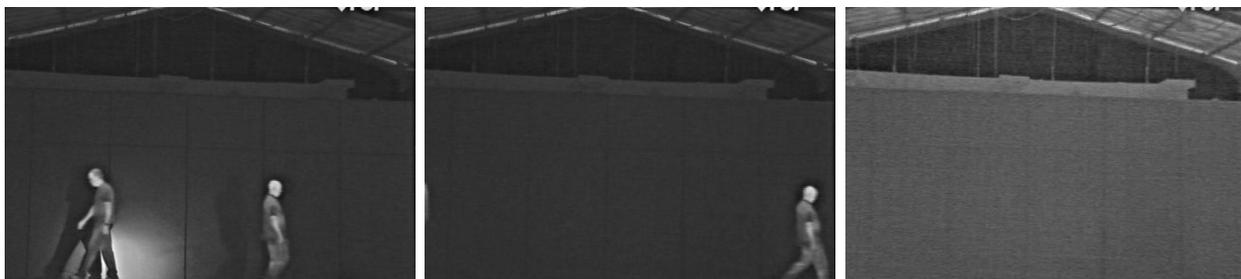
Quadro 150



(h) Método espacial: UWT-SF



(i) Método espaço-temporal: Decomposição em Sub-bandas Temporais



(j) Método espaço-temporal: Decomposição em Sub-bandas Espaço-temporais

## C.5 Lab

Quadro 10



Quadro 60



Quadro 150



(k) Método espacial: UWT-SF



(l) Método espaço-temporal: Decomposição em Sub-bandas Temporais



(m) Método espaço-temporal: Decomposição em Sub-bandas Espaço-temporais