

MARCELO VICTOR WÜST ZIBETTI

**SUPER-RESOLUÇÃO SIMULTÂNEA PARA
SEQÜÊNCIA DE IMAGENS**

**FLORIANÓPOLIS
2007**

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA**

**SUPER-RESOLUÇÃO SIMULTÂNEA PARA
SEQÜÊNCIA DE IMAGENS**

Tese submetida à
Universidade Federal de Santa Catarina
como requisito parcial à obtenção do grau de
Doutor em Engenharia Elétrica

MARCELO VICTOR WÜST ZIBETTI

Florianópolis, Março de 2007.

SUPER-RESOLUÇÃO SIMULTÂNEA PARA SEQUÊNCIA DE IMAGENS

Marcelo Victor Wüst Zibetti

Esta tese foi julgada adequada para a obtenção do título de *Doutor em Engenharia Elétrica*, área de concentração *Comunicações e Processamento de Sinais*, sendo aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.

Prof. Joceli Mayer, Ph.D.
Orientador

Prof. Nelson Sadowski, Dr.
Coordenador do Curso de Pós-Graduação em Engenharia Elétrica
da Universidade Federal de Santa Catarina.

Banca Examinadora:

Prof. Joceli Mayer, Ph.D.
Presidente

Prof. José Carlos Moreira Bermudez, Ph.D.

Prof. Fermín Sinforiano Viloche Bazán, Dr.

Prof. Vitor Heloiz Nascimento, Ph.D.

Prof. Alvaro Rodolfo De Pierro, Dr.
Relator

Agradecimentos

Agradeço a minha Família, minha Esposa Juliane, meu Pai, minha Mãe e meus irmãos, pelo carinho, apoio e suporte emocional, que me ajudaram a cumprir esta grande e difícil tarefa.

Também agradeço aos meus mestres: ao Prof. Joceli Mayer, pela orientação dedicada e pela amizade; ao Prof. José Carlos M. Bermudez, pelos sábios conselhos e significativo apoio técnico, e ao Prof. Fermín S. V. Bazán, pelas profundas discussões matemáticas e grande suporte nas pesquisas, principalmente, na segunda contribuição desta tese.

Finalmente, gostaria também de agradecer as meus colegas, mais do que isso, amigos, do Laboratório de Pesquisa em Processamento Digital de Sinais (LPDS), com os quais pude contar sempre, fosse para desabafar, fosse discutir assuntos técnicos, ou para rir.

A todos vocês que, de uma forma ou de outra, contribuíram para que este trabalho se realizasse. Muito obrigado!

Resumo da Tese apresentada à UFSC como parte dos requisitos para obtenção do grau de Doutor em Engenharia Elétrica.

SUPER-RESOLUÇÃO SIMULTÂNEA PARA SEQÜÊNCIA DE IMAGENS

Marcelo Victor Wüst Zibetti

Março/2007

Orientador: Joceli Mayer, Ph.D.

Área de concentração: comunicações e processamento de sinais.

Palavras-chave: processamento de imagens, super-resolução, seqüência de imagens, aquisição de imagem, regularização, inferência Bayesiana, otimização.

Número de páginas: 167.

RESUMO:

Esta tese apresenta duas contribuições: uma nova classe de algoritmos de super-resolução (SR) simultânea e um novo método de determinação dos coeficientes de regularização. Os algoritmos SR buscam produzir imagens com resolução superior à fornecida pelo dispositivo de aquisição. Os algoritmos de super-resolução simultânea, por sua vez, produzem toda uma seqüência de imagens de alta resolução num único processo. O algoritmo SR simultâneo existente [1] é capaz de produzir imagens de qualidade superior à dos métodos tradicionais [2], porque explora a similaridade entre as imagens em alta resolução. Contudo, o custo computacional deste método é alto e ele apresenta baixa robustez aos grandes erros de movimento. A primeira contribuição desta tese foi o aprimoramento dos métodos SR simultâneos. As expressões que exploram similaridade entre as imagens em alta resolução foram preservadas e generalizadas, ao passo que algumas expressões relacionadas aos dados, que são sensíveis aos grandes erros de movimento e são redundantes para a solução do problema, foram removidas. Este aprimoramento reduziu o custo computacional e possibilitou um maior controle sobre os grandes erros de movimento, aumentando a robustez do método e mantendo a qualidade superior das estimativas. A determinação dos coeficientes de regularização é uma etapa necessária nos algoritmos SR estudados nesta tese. Os métodos clássicos de determinação dos coeficientes, que possuem boa qualidade de estimativa e grande estabilidade, têm alto custo computacional no problema de super-resolução; por outro lado, métodos rápidos de determinação dos coeficientes, ou têm pouca qualidade na estimativa, ou são instáveis. Logo, a segunda contribuição proposta nesta tese é um novo método de determinação dos coeficientes de regularização. Este método, baseado na teoria estatística Bayesiana JMAP e nos métodos baseados na Curva-L, alcança boa qualidade e grande estabilidade de estimativa, além de obter um baixo custo computacional.

Abstract of Thesis presented to UFSC as a partial fulfillment of the requirements for the degree of
Doctor in Electrical Engineering.

SIMULTANEOUS SUPER-RESOLUTION FOR IMAGE SEQUENCES

Marcelo Victor Wüst Zibetti

March/2007

Advisor: Joceli Mayer, Ph.D.

Area of Concentration: communication and signal processing.

Keywords: image processing, super-resolution, image sequences, image acquisition, regularization, Bayesian inference, optimization.

Number of Pages: 167.

ABSTRACT:

This Thesis presents two contributions: a new class of simultaneous super-resolution (SR) algorithms, and a new method for determining the regularization parameters. The SR algorithms produces images with a superior resolution compared to the resolution of the captured images. The simultaneous SR methods produces all the high-resolution images in the sequence together, in one process. The existing simultaneous SR algorithms [1] is capable to produce images with superior quality when compared with the traditional methods [2]. This improvement is achieved because the simultaneous method exploits the similarity of the frames in the high-resolution space. However, the computational cost of the existing method is very high, and it presents low robustness to large motion errors. The first contribution of this Thesis is the improvement of the simultaneous methods. The expressions used to exploit the similarity of the images in the high-resolution space are preserved and generalized, while the expressions related to the data were removed, which are sensitive to the large motion errors and are redundant to the estimation. These changes reduced the computational cost of the simultaneous methods and allowed an increased control over the large motion errors, improving the robustness of the method and maintaining the superior quality of the estimative. A necessary task in the regularized SR methods is the determination of the regularization parameters. The classical methods for determining the regularization parameters provide good quality of the estimative and are stable, however, they have high computational cost for SR problems. On the other hand, fast methods for determining the regularization parameters, either provide low quality of the estimative, or are instable. Thus, the second contribution proposed in this Thesis is a new method for choosing the regularization parameters. This method is based on the JMAP estimation and on the L-curve methods. The proposed method achieves good quality and great stability the estimative, besides, it provides low computational cost.

Sumário

1	Introdução	1
1.1	Super-Resolução	2
1.2	Tipos de Aplicações para a Super-Resolução	3
1.3	Super-Resolução de Sequência de Imagens	4
1.4	Contribuições desta Tese	5
1.5	Organização da Tese	5
2	Revisão Bibliográfica e Fundamentação Teórica	7
2.1	Modelos Usados em Super-Resolução	8
2.1.1	Aquisição de Imagem	8
2.1.2	Movimento entre Imagens	9
2.1.3	Modelo de Aquisição e de Movimento Combinados	11
2.1.4	Modelos de Imagem	12
2.2	Métodos de Super-Resolução	13
2.2.1	Interpolação Não-Uniforme e Restauração	14
2.2.2	<i>IBP-Iterative Back Projection</i>	15
2.2.3	Métodos Baseados em Algoritmos Adaptativos	16
2.2.4	<i>POCS - Projection Onto Convex Sets</i>	17
2.2.5	Métodos Regularizados	18
2.3	Fundamentos Teóricos dos Métodos Regularizados	20
2.3.1	Métodos Determinísticos Regularizados	20
2.3.2	Métodos Estatísticos Bayesianos	22
3	Super-Resolução Simultânea	26
3.1	Problemas Identificados no Método Simultâneo Existente	26
3.1.1	Alto Custo Computacional	27
3.1.2	Baixa Robustez do Método na Presença de <i>Outliers</i>	28
3.2	Solução Proposta de Super-Resolução Simultânea	29
3.2.1	Idéia Principal	29
3.2.2	Modelo Geral Proposto	31
3.2.3	Implementação Proposta	35
3.3	Comparações Teóricas	36
3.3.1	Comparações entre as Equações	36

3.3.2	Qualidade Obtida pelos Métodos Simultâneos	37
3.3.3	Redução de Custo Computacional	39
3.3.4	Aumento da Robustez aos <i>Outliers</i>	41
4	Experimentos Comparativos dos Métodos Regularizados de Super-Resolução	43
4.1	Descrição do Experimento	43
4.1.1	Procedimento	43
4.1.2	Algoritmos Implementados	44
4.1.3	Seqüências de Imagens Utilizadas	49
4.2	Resultados Obtidos	51
4.2.1	Qualidade e Robustez das Estimativas	51
4.2.2	Custo Computacional	59
5	Determinação dos Coeficientes de Regularização	62
5.1	Métodos Existentes de Determinação dos Coeficientes de Regularização para os Algoritmos de Super-Resolução	63
5.1.1	Objetivo dos Métodos de Determinação dos Coeficientes de Regularização	63
5.1.2	Revisão dos Métodos Existentes	65
5.1.3	Classificação dos Métodos Revisados	69
5.2	Método Proposto de Determinação dos Coeficientes	70
5.2.1	Revisão do JMAP Existente	70
5.2.2	Método JMAP Proposto	73
5.2.3	Métodos de Otimização Propostos	77
5.2.4	Extensão para Super-Resolução com Múltiplos Coeficientes	78
6	Experimentos Comparativos dos Métodos de Determinação dos Coeficientes	81
6.1	Descrição dos Experimentos	82
6.1.1	Procedimento	82
6.1.2	Métodos de Determinação dos Coeficientes Implementados	83
6.2	Experimentos com Super-Resolução Tradicional de um Coeficiente	85
6.2.1	Equacionamento do Algoritmo SR e Objetivo do Coeficiente	85
6.2.2	Resultados	86
6.3	Experimentos com Super-Resolução Tradicional de Múltiplos Coeficientes	92
6.3.1	Equacionamento do Algoritmo SR e Objetivo dos Coeficientes	92
6.3.2	Resultados	92
6.4	Experimentos com Super-resolução Simultânea de Dois Coeficientes	96
6.4.1	Equacionamento do Algoritmo SR e Objetivo dos Coeficientes	96
6.4.2	Resultados	96
7	Conclusões e Linhas de Trabalhos Futuros	100
7.1	Conclusões	100
7.1.1	Conclusões Sobre o Método de Super-Resolução Simultânea	100
7.1.2	Conclusões Sobre o Método de Determinação dos Coeficientes de Regularização	101

7.2	Linhas de Trabalhos Futuros	102
7.2.1	Linhas de Trabalhos Futuros em Super-Resolução Simultânea	102
7.2.2	Linhas de Trabalhos Futuros na Determinação dos Coeficientes	102
7.2.3	Outras Linhas de Trabalhos Futuros	103
A	Complementos do Capítulo 2	104
A.1	Discretização do Sistema de Aquisição	104
A.1.1	Sistemas Invariantes	106
A.1.2	Exemplo do Sistema de Aquisição	108
A.2	Modelo Contínuo e Discreto de Movimento	109
A.2.1	Exemplo do Transformação de Movimento	110
A.3	Formas Comuns da Matriz \mathbf{R}	111
A.4	Funções de Discrepância mais Conhecidas	113
B	Complementos do Capítulo 3	115
B.1	Termos de Dados Considerados	115
B.2	Exemplos Construtivos de \mathbf{M}	117
B.3	Métodos de Otimização do Gradiente Conjugado	118
B.3.1	Gradiente Conjugado Linear	118
B.3.2	Gradiente Conjugado Não-Linear	119
B.3.3	Cálculo do Gradiente para o Gradiente Conjugado Não-Linear	120
B.4	Gradientes das Funções de Discrepância mais Conhecidas	121
C	Complementos do Capítulo 5	123
C.1	Método de Gradiente Conjugado Não Linear (GC-NL)	123
C.2	Análise de Convexidade do Método Proposto com Minimização Direta	124
C.2.1	Cálculo do Gradiente	125
C.2.2	Cálculo da Hessiana	126
C.3	Análise de Convergência do Critério Iterativo Alternado de Aproximação Sucessiva	128
C.3.1	GSVD - Decomposição Generalizada de Valores Singulares	128
C.3.2	Decomposição dos Termos	129
C.3.3	Análise da Convergência da Implementação por Aproximações Sucessivas	132
C.4	Escolha Alternativa para o μ	138
C.4.1	Decomposição dos Termos Através da GSVD	139
C.4.2	Análise da Similaridade dos Termos	140
C.4.3	Escolha do μ	144

Lista de Figuras

1.1	Efeitos visuais da baixa resolução espacial e da baixa resolução da intensidade.	1
1.2	Ilustração dos casos mais comuns de obtenção de diferentes informações entre as imagens capturadas.	3
2.1	Ilustração da degradação sofrida na aquisição.	9
2.2	Ilustração da nova informação temporal $e_{j,k}$	11
3.1	Desvio-padrão da diferença entre quadros adjacentes ao longo da seqüência.	38
3.2	Observação da redução de custo computacional do método proposto.	40
3.3	Histograma do erro de movimento de uma seqüência com a presença de <i>outliers</i> e com os <i>outliers</i> removidos.	42
4.1	Função discreta de aquisição utilizada.	44
4.2	Seqüência de Imagens <i>Barco</i>	49
4.3	Seqüência de Imagens <i>Árvore</i>	49
4.4	Seqüência de Imagens <i>Sala</i>	50
4.5	Seqüência de Imagens <i>Estrada</i>	50
4.6	Seqüência de Imagens <i>Jato</i>	50
4.7	Seqüência de Imagens <i>Jardim</i>	51
4.8	Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma ℓ_2 ao quadrado, modelos de primeira ordem e com detecção de <i>outliers</i> ativada	53
4.9	Exemplo visual mostrando a imagem 2 da seqüência Estrada, usando norma ℓ_2 ao quadrado, modelos de segunda ordem e com detecção de <i>outliers</i> ativada	53
4.10	Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma ℓ_2 ao quadrado, modelos de primeira ordem e com detecção de <i>outliers</i> desativada	54
4.11	Exemplo visual mostrando a imagem 2 da seqüência Estrada, usando norma ℓ_2 ao quadrado, modelo de segunda ordem e com detecção de <i>outliers</i> desativada	54
4.12	Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma Huber/ ℓ_1 , com modelos de primeira ordem e com detecção de <i>outliers</i> ativada	56
4.13	Exemplo visual mostrando a imagem 2 da seqüência Jato, usando norma Huber/ ℓ_1 , com modelos de segunda ordem e com detecção de <i>outliers</i> ativada	57
4.14	Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma Huber/ ℓ_1 , usando modelos de primeira ordem e com detecção de <i>outliers</i> desativada	57

4.15	Exemplo visual mostrando a imagem 2 da seqüência Jato, usando norma Huber/ ℓ_1 , modelos de segunda ordem e com detecção de <i>outliers</i> desativada	58
4.16	Exemplo visual mostrando a imagem 6 da seqüência Jardim, usando norma ℓ_2 ao quadrado, modelos de primeira ordem e com detecção de <i>outliers</i> ativada	59
4.17	Exemplo visual mostrando a imagem 6 da seqüência Jardim, usando norma Huber/ ℓ_1 , modelos de primeira ordem e com detecção de <i>outliers</i> ativada	59
5.1	Comportamento esperado da Norma ℓ_2 do erro de estimativa.	64
5.2	Resultados ilustrativos com λ muito pequeno, λ_{opt} e λ muito grande.	64
5.3	Curva-L em diversas escalas.	65
5.4	Curva da superfície de desempenho do método GCV.	67
5.5	Densidade Gama, sugerida para os hiperparâmetros, com média 1 e variância 0.1.	74
6.1	Função discreta de aquisição utilizada.	82
6.2	Ilustração do nível de ruído na aquisição.	82
6.3	Exemplo visual mostrando a imagem 2 da seqüência Barco, com alto ruído.	88
6.4	Exemplo visual mostrando a imagem 2 da seqüência Barco em outra realização, com alto ruído.	89
6.5	Exemplo visual mostrando a imagem 2 da seqüência Jardim, com baixo ruído.	90
6.6	Exemplo visual mostrando a imagem 2 da seqüência Sala, com baixo ruído.	94
6.7	Exemplo visual mostrando a imagem 2 da seqüência Jardim, com baixo ruído.	94
6.8	Exemplo visual mostrando a imagem 3 da seqüência Jardim, com baixo ruído.	95
6.9	Exemplo visual mostrando a imagem 2 da seqüência Sala, com baixo ruído.	98
6.10	Exemplo visual mostrando a imagem 2 da seqüência Jardim, com baixo ruído.	98
A.1	Ilustração da posição dos pixels nas imagens e da função de aquisição.	108
A.2	Ilustração da posição dos pixels nas imagens e da mudança de posição ocorrida com o movimento.	111
C.1	Comportamento das curvas $\frac{1}{b(\lambda)}$, $c(\lambda)$ e $NS(\lambda)$	142
C.2	Comportamento das curvas $\frac{1}{a(\lambda)}$, $r(\lambda)$ e $DA(\lambda)$	144
C.3	Erro de estimativa, critério com $\mu = 1$ e critério corrigido.	145

Lista de Tabelas

2.1	Algoritmo LMS aplicado em super-resolução	17
3.1	Modelos utilizados nos algoritmos de super-resolução	37
3.2	Custo por iteração do GC linear em operações (adição ou multiplicação)	39
3.3	Exemplo de custo por iteração do GC com valores típicos encontrados em SR	40
4.1	Qualidade dos métodos SR com norma ℓ_2 ao quadrado e detecção de <i>outliers</i> ativada , em SNR (dB)	52
4.2	Qualidade dos métodos SR com norma ℓ_2 ao quadrado e detecção de <i>outliers</i> desativada , em SNR (dB)	52
4.3	Qualidade dos métodos SR com norma Huber ou ℓ_1 e detecção de <i>outliers</i> ativada , em SNR (dB)	55
4.4	Qualidade dos métodos SR com norma Huber ou ℓ_1 e detecção de <i>outliers</i> desativada , em SNR (dB)	55
4.5	Custo Computacional dos métodos SR com norma ℓ_2 ao quadrado, em segundos	60
4.6	Custo Computacional dos métodos SR com norma Huber ou ℓ_1 , em segundos	60
5.1	Classificação dos Métodos de Determinação dos Coeficientes em Super-Resolução	70
6.1	Listagem dos Métodos utilizados em cada algoritmo SR	85
6.2	Qualidade, em SNR (dB), dos Métodos para SR Tradicional de um Coeficiente com Alto Ruído	86
6.3	Qualidade, em SNR (dB), dos Métodos para SR Tradicional de um Coeficiente com Baixo Ruído	87
6.4	Custo computacional, em segundos, dos Métodos para SR Tradicional de Um Coeficiente com Alto Ruído	91
6.5	Custo computacional, em segundos, dos Métodos para SR Tradicional de Um Coeficiente com Baixo Ruído	91
6.6	Qualidade, em SNR (dB), dos Métodos para SR Tradicional de Múltiplos Coeficientes com Alto Ruído	93
6.7	Qualidade, em SNR (dB), dos Métodos para SR Tradicional de Múltiplos Coeficientes com Baixo Ruído	93
6.8	Custo computacional, em segundos, dos Métodos para SR Tradicional de Múltiplos Coeficientes com Alto Ruído	95

6.9	Custo computacional, em segundos, dos Métodos para SR Tradicional de Múltiplos Coeficiente com Baixo Ruído	96
6.10	Qualidade, em SNR (dB), dos Métodos para SR Simultânea com Dois Coeficientes com Alto Ruído	97
6.11	Qualidade, em SNR (dB), dos Métodos para SR Simultânea com Dois Coeficientes com Baixo Ruído	97
6.12	Custo computacional, em segundos, dos Métodos para SR Simultânea com Dois Coeficientes com Alto Ruído	99
6.13	Custo computacional, em segundos, dos Métodos para SR Simultânea com Dois Coeficientes com Baixo Ruído	99
B.1	Gradiente Conjugado Linear	119
B.2	Gradiente Conjugado Não-Linear	120
C.1	Gradiente Conjugado Não-Linear para Solução da Imagem HR e do Coeficiente	124

Lista de Símbolos e Siglas

Símbolos

$f(x_1, x_2)$	— Imagem de alta resolução na forma contínua.
$f[m_1, m_2]$	— Imagem de alta resolução na forma discreta.
$f_k[m_1, m_2]$	— Imagem discreta de alta resolução no instante k de uma seqüência de imagens.
\mathbf{f}_k	— Imagem de alta resolução discreta, no instante k , representada na forma de vetor com ordenação lexicográfica.
$M = M_1 M_2$	— Número de pixels da imagem de alta resolução, na direção horizontal, M_1 e na direção vertical M_2 .
$\mathbf{f} = [\mathbf{f}_1^T \dots \mathbf{f}_L^T]^T$	— Seqüência de imagens de alta resolução, representada na forma de vetor.
\mathbf{D}	— Matriz de degradação.
\mathbf{D}_k	— Matriz de degradação do instante k .
$g[m_1, m_2]$	— Imagem discreta de baixa resolução.
$g_k[m_1, m_2]$	— Imagem discreta de baixa resolução no instante k de uma seqüência de imagens.
\mathbf{g}_k	— Imagem de baixa resolução discreta, no instante k , representada na forma de vetor com ordenação lexicográfica.
$N = N_1 N_2$	— Número de pixels da imagem de baixa resolução, na direção horizontal, N_1 e na direção vertical N_2 .
R	— Fator de redução da resolução.
$\mathbf{g} = [\mathbf{g}_1^T \dots \mathbf{g}_L^T]^T$	— Seqüência de Imagens de baixa resolução, representada na forma de vetor.
$\boldsymbol{\eta}_k$	— Ruído de aquisição, na forma de vetor, aplicado na imagem de baixa resolução no instante k .
$\boldsymbol{\eta}$	— Ruído de aquisição, na forma de vetor, aplicado em toda a seqüência de imagens.
\mathbf{B}_k	— Matriz de <i>Blurring</i> ou distorção óptica do instante k .
\mathbf{S}_k	— Matriz de subamostragem do instante k .

$\mathbf{M}_{k,j}$	—	Matriz de transformação de movimento, entre as imagens no instante j e k .
$\mathbf{e}_{k,j}$	—	Erro de movimento entre a imagem no instante k e a imagem do instante j com movimento compensado.
$\mathbf{I}_{k,j}$	—	Matriz diagonal de preservação dos <i>inliers</i> do movimento entre o instante j e k .
$\bar{\mathbf{M}}_{k,j}$	—	Matriz de transformação de movimento dos <i>inliers</i> das imagens no instante j e k .
$\bar{\mathbf{e}}_{k,j}$	—	Erro de movimento entre os <i>inliers</i> da imagem no instante k e no instante j , com movimento compensado.
$\mathbf{C}_{k,j}$	—	Matriz de transformação conjunta de aquisição e movimento.
$\boldsymbol{\epsilon}_{k,j}$	—	Erro de movimento em baixa resolução
$\boldsymbol{\varepsilon}_{k,j} = \boldsymbol{\epsilon}_{k,j} + \boldsymbol{\eta}_k$	—	Erro do modelo combinado, ruído de aquisição mais o erro de movimento em baixa resolução
\mathbf{R}_k	—	Matriz de regularização espacial para a imagen no instante k .
$\rho(\mathbf{x})$	—	Densidade de probabilidade conjunta do vetor \mathbf{x} .
$Q(\mathbf{f})$	—	Penalidade de regularização.
β_i	—	Hiperparâmetros do modelo estatístico das imagens ou seqüência de imagens.
\mathbf{R}	—	Matriz de regularização espacial para a seqüência de imagens.
\mathbf{M}	—	Matriz de regularização na trajetória de movimento da seqüência de imagens.

Siglas

SR	—	super-resolução
HR	—	<i>High-Resolution</i> alta resolução
LR	—	<i>Low-Resolution</i> baixa resolução
i.i.d.	—	independentes e identicamente distribuídas
i.e.	—	<i>id est</i> (Latim) - isto é
MAP	—	<i>maximum a posteriori</i>
JMAP	—	<i>joint maximum a posteriori</i>
POCS	—	<i>projection onto convex sets</i>
IBP	—	<i>iterative back projection</i>

Capítulo 1

Introdução

O uso de imagens digitais tem se tornado cada vez mais freqüente em diversas áreas. Um dos fatores que vêm motivando essa utilização é o baixo custo para obtenção de imagens fornecidas por dispositivos digitais quando comparados aos dispositivos analógicos. A qualidade de uma imagem digital está diretamente relacionada à sua resolução. A palavra “resolução”, neste contexto, é definida como o nível de detalhe de uma representação visual [3, 4]. Quanto maior a resolução de uma imagem digital, mais precisa é sua representação da imagem real.

A resolução de uma imagem digital é representada por dois parâmetros: a *resolução espacial*, que especifica o número de amostras, elementos ou *pixels*, em uma imagem, e a *resolução da intensidade*, ou quantização, que especifica o número de níveis possíveis para representar a intensidade em determinado pixel. A Figura 1.1 ilustra o efeito da baixa resolução espacial e da baixa resolução da intensidade. O foco principal deste trabalho é a melhoria da resolução espacial, embora a melhoria da resolução da intensidade também aconteça.

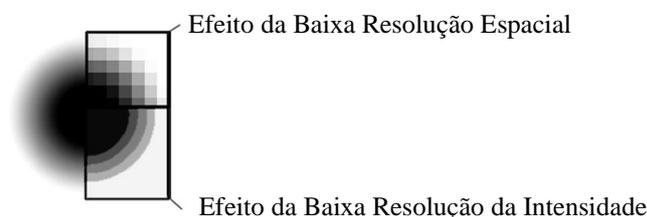


Figura 1.1: Efeitos visuais da baixa resolução espacial e da baixa resolução da intensidade.

Imagens com alta resolução (HR - *high-resolution*) são fundamentais em várias aplicações. Alguns exemplos práticos são: o diagnóstico por imagem pode detectar doenças ainda em estágios iniciais [5, 6]; a astronomia pode revelar novas estrelas e galáxias nunca antes visíveis [7], e o sensoriamento remoto

pode ser feito de distâncias ainda maiores.

Entretanto, nem sempre é vantajoso, ou mesmo possível, melhorar a resolução de um dispositivo de aquisição de imagens. Utilizar lentes de melhor qualidade, sensores mais sofisticados e mecanismos mais precisos pode elevar significativamente o custo de um dispositivo. Os fabricantes de sensores têm demonstrado que o aumento da densidade de células detectoras num sensor é proporcional ao aumento do ruído de aquisição e leva, conseqüentemente, ao decréscimo na qualidade da imagem capturada [8, 3]. Além disso, deve-se considerar a grande quantidade de imagens já capturadas de cenários que podem não se repetir mais; nestes casos, a aplicação de processamento digital para melhorar a resolução da imagem é fundamental.

Neste trabalho são consideradas as técnicas de processamento digital chamadas de “super-resolução” (SR) [3, 2], cujo objetivo é melhorar a resolução das imagens capturadas, aprimorando os detalhes nas imagens originalmente obtidas.

1.1 Super-Resolução

As técnicas de super-resolução proporcionam uma forma de obter imagens com resolução superior à fornecida pelo dispositivo de aquisição [3, 2, 9, 10, 11]. Estas técnicas removem distorções causadas durante o processo de aquisição, realçam características importantes nas imagens, como as bordas dos objetos, e, ainda, recuperam detalhes pela utilização da informação contida em um grande conjunto de imagens capturadas.

A super-resolução teve origem na óptica, na qual se busca recriar as frequências além do limite de difração, empregando extrapolação das altas frequências espaciais [9]. De acordo com [12], ainda na física, pesquisadores demonstraram que era possível recuperar a resolução espacial utilizando a informação distribuída no tempo, ou seja, a informação contida em um conjunto de imagens capturadas ao longo do tempo [13, 14, 15].

A obtenção da super-resolução através da utilização de um conjunto de imagens é uma das formas mais exploradas atualmente [10, 11, 3, 2]. Assume-se que as imagens capturadas, chamadas de imagens de baixa resolução (LR - *low-resolution*), são altamente correlacionadas. Porém, uma pequena diferença de informação presente em cada imagem capturada proporciona uma nova informação para a recuperação dos detalhes na imagem HR. Os exemplos mais comuns de obtenção dessa diferença de informação são:

- o movimento de subpixel [2], que ocorre quando o movimento entre as imagens é fracionário com respeito à grade de baixa resolução;

- as diferentes distorções na aquisição [16, 17], que ocorrem quando, por exemplo, a aquisição é feita por sensores diferentes ou por câmeras diferentes.

A Figura 1.2 ilustra como a posição relativa do arranjo de células durante o processo de aquisição proporciona uma observação única da imagem HR. Neste trabalho será enfatizada a super-resolução obtida por movimento de subpixel, embora a formulação inclua as duas possibilidades.

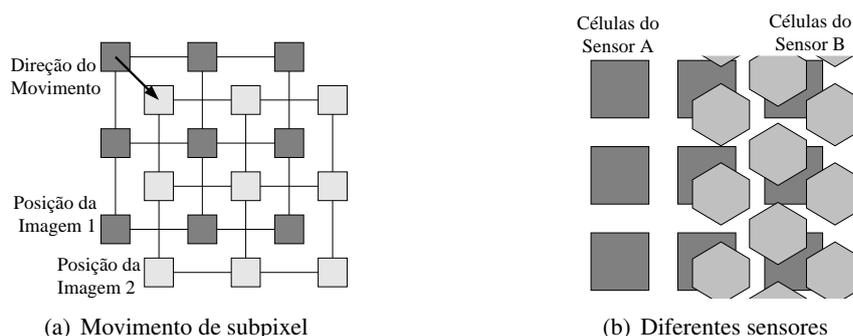


Figura 1.2: Ilustração dos casos mais comuns de obtenção de diferentes informações entre as imagens capturadas.

As técnicas de super-resolução diferenciam-se de técnicas simples de ampliação digital de imagem, como as técnicas de interpolação [18, 19], principalmente porque estas não tentam remover distorções da aquisição. As técnicas de super-resolução têm uma formulação semelhante à das técnicas de restauração de imagem [20], porém são diferentes quanto aos objetivos. A restauração de imagem não busca aumentar a resolução, mas, apenas, remover as distorções que degradaram a imagem. A super-resolução pode ser vista como uma técnica mista de restauração e interpolação, pois proporciona ampliação e remoção de distorções.

1.2 Tipos de Aplicações para a Super-Resolução

As aplicações que se beneficiam com a super-resolução podem ser classificadas quanto ao seu objetivo final. Assim, têm-se as aplicações cujo objetivo é a obtenção de uma *única imagem* de alta resolução a partir do conjunto de imagens capturadas e as aplicações cujo objetivo é a obtenção de uma *seqüência de imagens* de alta resolução a partir da seqüência de imagens de baixa resolução.

No primeiro grupo, em que o objetivo é obter uma única imagem HR, podem-se citar os seguintes exemplos: imagens estáticas a partir de vídeo [21]; fotografias panorâmicas e mosaicos [22]; imagens aéreas e imagens de diagnóstico médico.

No segundo grupo, no qual o objetivo é obter uma seqüência de imagens HR, podem-se citar exem-

plos como: vídeo de alta definição; recuperação de vídeos antigos; remoção de distorções causadas pela compressão de vídeo [23]; seqüência de imagens de ultra-som, vigilância e sensoriamento remoto.

Este trabalho enfatiza as aplicações pertencentes ao segundo grupo, ou seja, as aplicações nas quais o objetivo é a obtenção de uma seqüência de imagens de alta resolução.

1.3 Super-Resolução de Seqüência de Imagens

Considerando aplicações cujo objetivo é recuperar uma seqüência de imagens, tem-se que os algoritmos de super-resolução se diferenciam em três classes de acordo com a forma de aplicação da super-resolução:

- Métodos Tradicionais: são métodos que aplicam as técnicas para a obtenção de uma única imagem HR. O procedimento é repetido, buscando recuperar as demais imagens da seqüência. Porém, esses métodos não aproveitam os resultados previamente obtidos na super-resolução de uma outra imagem da seqüência [3, 2, 10].
- Métodos Seqüenciais: são métodos que aplicam a super-resolução nas imagens da seqüência, uma a uma, utilizando os resultados previamente obtidos [24, 25];
- Métodos Simultâneos: são métodos que estimam toda a seqüência de imagens em um único processo [1, 26].

Os métodos tradicionais de super-resolução foram os primeiros a serem aplicados em seqüências de imagens. As publicações anteriores do presente autor, provenientes do trabalho de mestrado, são baseadas nos métodos tradicionais [27, 28, 29, 30]. Porém, esses métodos estimam uma imagem da seqüência de cada vez, sem aproveitar as imagens HR previamente estimadas. O algoritmo seqüencial de [24] foi introduzido com o objetivo de reduzir o custo computacional total, aproveitando as imagens previamente estimadas, contudo a qualidade das estimativas foi comprometida, sendo inferior à obtida pelos métodos tradicionais.

Por outro lado, os métodos simultâneos foram desenvolvidos para buscar uma qualidade superior nos resultados. A melhora na qualidade é obtida através da melhor utilização da informação sobre o movimento presente na seqüência de imagens. Todavia, o custo computacional para a obtenção dessa melhoria no primeiro algoritmo simultâneo proposto [1] era muito maior do que o dos métodos tradicionais. Nesta tese, o método simultâneo existente, proposto em [1], que configura o atual estado-da-arte na busca por super-resolução de alta qualidade, será estudado e desenvolvido.

1.4 Contribuições desta Tese

Esta tese apresenta duas contribuições principais. A primeira é o desenvolvimento de uma nova classe de métodos simultâneos. A característica principal dos métodos simultâneos existentes, que ocasiona a melhora na qualidade dos resultados, é estudada e aprimorada, visando a, principalmente, um desempenho mais robusto aos problemas de movimento. Ao mesmo tempo, os fatores que elevam o custo computacional, são sensíveis aos grandes erros e são redundantes para a estimativa foram removidos, formando uma nova classe de métodos que proporcionam a mesma qualidade superior dos métodos simultâneos existentes, mas com menor custo computacional e com maior robustez à erros de movimento. Essas contribuições foram publicadas em [26], [31]¹ e [32], e submetida para publicação em [33].

A segunda contribuição é o desenvolvimento de um método de determinação dos coeficientes utilizados nos métodos simultâneos desenvolvidos nesta tese. A determinação dos coeficientes adequados é importante para garantir o melhor desempenho no método simultâneo proposto. Os métodos de determinação dos coeficientes utilizados em outros algoritmos de super-resolução não são diretamente aplicáveis a esta nova classe [34, 35, 36, 37, 38]. Além disso, a utilização de métodos matemáticos clássicos [39, 40], que podem ser adaptados para o problema, eleva o custo computacional a níveis proibitivos. Logo, o desenvolvimento de um novo método para a determinação dos coeficientes, que seja confiável, com baixo custo computacional e que garanta uma solução adequada, foi uma necessidade. Contudo, o novo método desenvolvido não só atingiu um bom desempenho no método simultâneo proposto, mas também alcançou um bom desempenho quando adaptado a outros métodos de super-resolução, como os tradicionais. A segunda contribuição ainda não foi publicada, mas foi submetida em [41].

1.5 Organização da Tese

No Capítulo 2 serão apresentadas revisões dos modelos utilizados em super-resolução, das técnicas de super-resolução mais relevantes na área e da fundamentação das técnicas de super-resolução regularizadas, que são a base dos algoritmos simultâneos estudados. No Capítulo 3, serão enumerados os problemas encontrados nos algoritmos de super-resolução simultâneos existentes e será apresentada a primeira contribuição obtida nesta pesquisa: a nova classe de métodos simultâneos desenvolvida. Ao final do capítulo, o método simultâneo proposto é comparado aos métodos existentes na literatura. No Capítulo 4, serão mostrados os experimentos realizados contrastando os algoritmos desenvolvidos com

¹Este trabalho foi premiado com o IBM BEST STUDENT PAPER AWARD, na IEEE International Conference on Image Processing, realizada em Atlanta (Georgia, EUA), de 8 a 11 de outubro de 2006.

os algoritmos de super-resolução existentes. No Capítulo 5, será apresentado o problema da determinação dos coeficientes de regularização e a segunda contribuição da tese, que é o método de determinação dos coeficientes desenvolvido durante esta pesquisa. No Capítulo 6, os experimentos, comparando-se o método de determinação dos coeficientes proposto com outros métodos existentes, são apresentados. E, finalmente, no Capítulo 7, são oferecidas as conclusões e pontos para futuras pesquisas.

Capítulo 2

Revisão Bibliográfica e Fundamentação Teórica

Neste capítulo, apresenta-se a revisão bibliográfica dos algoritmos de super-resolução e sua fundamentação teórica. Inicialmente, são mostrados os modelos utilizados nos algoritmos de super-resolução; em seguida, alguns dos métodos de super-resolução mais relevantes na literatura da área e, ao final, os fundamentos dos métodos regularizados.

Na seção 2.1, são apresentados os modelos utilizados nos algoritmos de super-resolução. Esses modelos são a base para os métodos de super-resolução revisados nesta tese e serão frequentemente referenciados.

Na seção 2.2, são revisados os métodos de super-resolução mais adequados às aplicações enfocadas nesta tese. São eles: Interpolação Não-Uniforme e Restauração, *Iterative Back Projection*, métodos baseados em algoritmos adaptativos, Projeção em Conjuntos Convexos (POCS - *Projection Onto Convex Sets*) e os Métodos Regularizados.

Na seção 2.3, os fundamentos dos métodos regularizados são apresentados. Os métodos regularizados foram escolhidos para desenvolvimento dentro dos métodos de super-resolução, devido à grande variedade de modelos e ferramentas disponíveis. Duas linhas de algoritmos regularizados são apresentadas: a linha estatística, dos algoritmos Bayesianos de *maximum a posteriori* (MAP), e a linha determinística, dos algoritmos baseados em regularização generalizada de Tikhonov. Será demonstrado que as duas linhas levam a algoritmos matematicamente idênticos, porém com pontos de vista diferentes sobre o problema. Essas duas interpretações serão utilizadas para entender melhor o problema e para formular as soluções.

2.1 Modelos Usados em Super-Resolução

2.1.1 Aquisição de Imagem

O modelo de aquisição descreve as transformações que ocorrem com as imagens originais durante o processo de aquisição. Neste modelo, os dados observados são as imagens de baixa resolução. O modelo é descrito pela seguinte equação:

$$\mathbf{g}_k = \mathbf{D}_k \mathbf{f}_k + \boldsymbol{\eta}_k \quad (2.1)$$

Em (2.1), \mathbf{g}_k é um vetor, de tamanho $N \times 1$, que representa a imagem LR no instante temporal k . Os elementos de \mathbf{g}_k correspondem aos pixels da imagem LR, $g_k[n_1, n_2]$, ordenados lexicograficamente. $N = N_1 N_2$ é o número de pixels na imagem LR, sendo $N_1 \times N_2$ o tamanho da imagem LR nas direções horizontal e vertical, respectivamente. O vetor \mathbf{f}_k , de tamanho $M \times 1$, representa a imagem HR, $f_k[m_1, m_2]$, de tamanho $M_1 \times M_2$, com $M = M_1 M_2$ pixels ordenados lexicograficamente. Define-se a redução da resolução da imagem na direção horizontal como $R_1 = M_1/N_1$ e a redução na direção vertical como $R_2 = M_2/N_2$. Quando a redução é a mesma nas duas dimensões, então $R = R_1 = R_2$, no qual $R > 1$. O vetor $\boldsymbol{\eta}_k$ representa o ruído de aquisição, assumido como estatisticamente independente de $\boldsymbol{\eta}_j$, para $j \neq k$.

A matriz \mathbf{D}_k , de tamanho $N \times M$, representa o sistema de aquisição para a imagem k . A coluna m , em \mathbf{D}_k , representa a *Função de Espalhamento* do pixel m da imagem HR durante a aquisição. A linha n , em \mathbf{D}_k , representa a *Função de Aquisição* do pixel n da imagem LR; a qual contém a ponderação de cada pixel da imagem HR na formação do pixel n . A matriz \mathbf{D}_k é uma versão discretizada do sistema de aquisição contínuo [8]. No Apêndice A.1 é apresentado o processo de discretização de \mathbf{D}_k .

Assume-se que o processo de aquisição degrada a imagem. Em outras palavras, \mathbf{D}_k representa as degradações sofridas na aquisição. São consideradas degradações como a distorção ótica, causada nas lentes e fotossensores, e a subamostragem, causada no sensor. A subamostragem implica a redução do número de amostras ou pixels.

Na Figura 2.1, é mostrado um exemplo com uma imagem de alta resolução e sua versão degradada pela aquisição. Tal modelo é, geralmente, chamado de *temporalmente coincidente*, porque a imagem HR e a imagem LR capturada correspondem ao mesmo instante temporal.

Em alguns casos específicos, quando R é um número inteiro e as distorções causadas nas lentes e no sensor podem ser modeladas como operadores invariantes no espaço, a matriz \mathbf{D}_k pode ser separada,

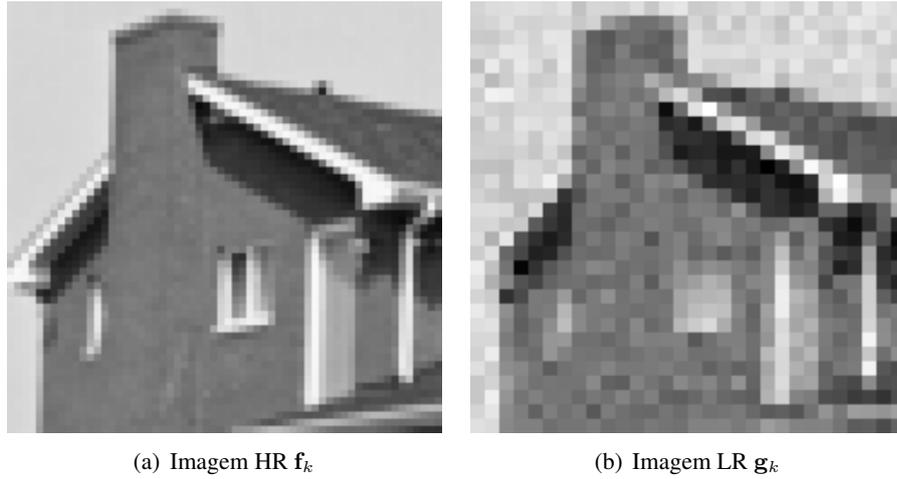


Figura 2.1: Ilustração da degradação sofrida na aquisição.

segundo:

$$\begin{aligned} \mathbf{g}_k &= \mathbf{D}_k \mathbf{f}_k + \boldsymbol{\eta}_k \\ \mathbf{g}_k &= \mathbf{S}_k \mathbf{B}_k \mathbf{f}_k + \boldsymbol{\eta}_k \end{aligned} \quad (2.2)$$

sendo que \mathbf{B}_k é uma matriz $M \times M$, geralmente Bloco-Toeplitz¹, que modela a convolução bidimensional da distorção óptica com a imagem. A matriz \mathbf{S}_k é a matriz de subamostragem, de tamanho $N \times M$, representa a redução da resolução pela dizimação das amostras da imagem.

2.1.2 Movimento entre Imagens

O modelo de movimento descreve a transformação causada pelo movimento entre as imagens da seqüência. Assume-se que a imagem no instante temporal j pode ser composta pela imagem do instante temporal k , com movimento compensado, mais uma nova informação, que não pode ser obtida da imagem no instante k . A equação que descreve o modelo é:

$$\mathbf{f}_j = \mathbf{M}_{j,k} \mathbf{f}_k + \mathbf{e}_{j,k} \quad (2.3)$$

Em (2.3), \mathbf{f}_j e \mathbf{f}_k são vetores que representam as imagens $f_j[m_1, m_2]$ e $f_k[m_1, m_2]$ nos instantes temporais j e k , respectivamente. A matriz $\mathbf{M}_{j,k}$, de tamanho $M \times M$, representa a transformação de movimento. Essa matriz é, normalmente, bastante esparsa [43]. A linha m é quase toda composta por zeros, exceto por:

- um elemento com valor unitário, na coluna m , que representa o deslocamento do pixel n em \mathbf{f}_k

¹Em alguns casos a matriz é aproximada por uma matriz Bloco-Circulante [8, 4, 42].

para a posição m em \mathbf{f}_j ;

- poucos elementos não-nulos, que correspondem à combinação de determinados pixels de \mathbf{f}_k para formar o pixel m em \mathbf{f}_j . Isso ocorre em movimentos cujo deslocamento do pixel é fracionário na grade de alta resolução.

A matriz $\mathbf{M}_{j,k}$ é uma versão discretizada do operador contínuo de transformação de movimento. No Apêndice A.2, encontra-se mais informação sobre a transformação de movimento contínua e sua representação discreta. O vetor $\mathbf{e}_{j,k}$ é chamado de erro de movimento e representa a nova informação que não pode ser obtida da imagem no instante k . Esta nova informação pode ser dividida em duas partes:

1. pequenas diferenças na imagem com movimento compensado, as quais podem ser originadas por pequenos erros na estimação dos parâmetros de movimento, limitações do operador discretizado em representar o movimento contínuo, ou outras pequenas variações temporais não relacionadas ao movimento;
2. erros de grande magnitude, (*outliers*²) [45]. No contexto do movimento, um *outlier* é, geralmente, uma região ou objeto que recebeu oclusão pelo movimento de outros objetos; objetos que aparecem repentinamente na cena ou regiões que sofrem um movimento não esperado. O movimento translacional também gera *outliers* na borda das imagens.

Os *outliers* provocam distorções nas imagens estimadas. Ou algoritmos robustos [44], ou técnicas de pré-deteção e eliminação, como em [45], devem ser utilizados para evitar as distorções. A equação seguinte apresenta o modelo de movimento no qual os *outliers* são removidos:

$$\begin{aligned}\mathbf{I}_{j,k}\mathbf{f}_j &= \mathbf{I}_{j,k}(\mathbf{M}_{j,k}\mathbf{f}_k + \mathbf{e}_{j,k}) \\ \mathbf{I}_{j,k}\mathbf{f}_j &= \mathbf{I}_{j,k}\mathbf{M}_{j,k}\mathbf{f}_k + \mathbf{I}_{j,k}\mathbf{e}_{j,k} \\ \bar{\mathbf{f}}_{j,k} &= \bar{\mathbf{M}}_{j,k}\mathbf{f}_k + \bar{\mathbf{e}}_{j,k}\end{aligned}\tag{2.4}$$

sendo que o operador $\mathbf{I}_{j,k}$ é uma matriz diagonal que preserva os *inliers*, i.e., pixels observados em ambas as imagens, e elimina os *outliers*. $\bar{\mathbf{f}}_{j,k}$, $\bar{\mathbf{M}}_{j,k}$ e $\bar{\mathbf{e}}_{j,k}$ são os *inliers* de \mathbf{f}_j , $\mathbf{M}_{j,k}$ e $\mathbf{e}_{j,k}$.

Na Figura 2.2, são demonstradas duas imagens, \mathbf{f}_j e \mathbf{f}_k , o valor absoluto da nova informação, $|\mathbf{e}_{j,k}| = |\mathbf{f}_j - \mathbf{M}_{j,k}\mathbf{f}_k|$, e a representação, na forma de imagem, dos elementos da matriz diagonal $\mathbf{I}_{j,k}$. Pode-se identificar facilmente os grandes erros, ou *outliers*, nos cantos da imagem $|\mathbf{e}_{j,k}|$ e nas regiões em que

²O termo *outlier* foi, inicialmente, utilizado em estatística. Na literatura estatística um *outlier* é um elemento ou uma medida ruim, geralmente um erro de grande magnitude, que não segue o modelo assumido [44]. Esta definição será considerada quando a estatística de $\mathbf{e}_{j,k}$ for modelada.

aparece o carro, e os pequenos erros, ou *inliers*, na região central da imagem $|e_{j,k}|$. Na imagem formada pelos elementos da diagonal de $\mathbf{I}_{j,k}$, são observadas a região de *inliers* em branco, que são preservados, e a região de *outliers* em preto, que são eliminados.

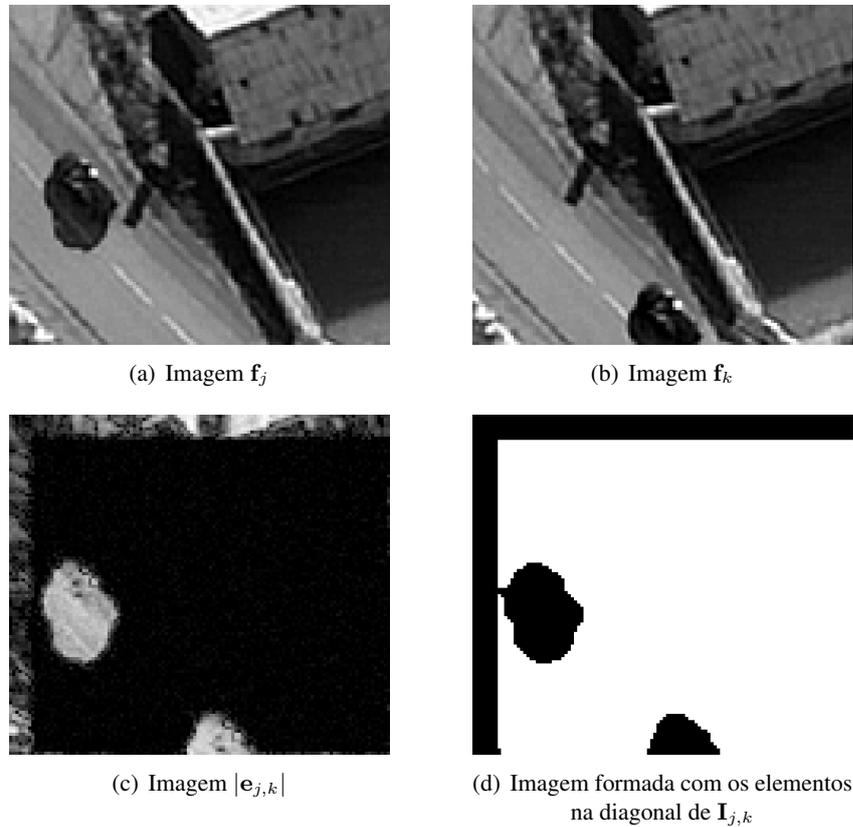


Figura 2.2: Ilustração da nova informação temporal $e_{j,k}$.

Em algumas aplicações, nas quais o movimento é translacional global e pequeno, é possível desconsiderar a presença dos *outliers* e assumir condições especiais para as bordas da imagem; como, por exemplo, a periodicidade da imagem [42], que é implicitamente assumida ao se aplicar Transformada Discreta de Fourier. Somente nesses casos se pode utilizar modelo da equação (2.3) sem se preocupar com a presença de *outliers* em $e_{j,k}$. Nas demais aplicações, os grandes erros causados pelos *outliers* precisam ser considerados.

2.1.3 Modelo de Aquisição e de Movimento Combinados

Os algoritmos de super-resolução que seguem a forma tradicional [2], alguns dos algoritmos seqüenciais [25] e o algoritmo simultâneo proposto em [1] utilizam o modelo de aquisição combinado com o modelo de movimento. O modelo combinado é utilizado para relacionar uma imagem LR a uma outra imagem HR, em um instante temporal diferente. Este modelo é descrito pela equação:

$$\begin{aligned}
\mathbf{g}_j &= \mathbf{D}_j(\mathbf{M}_{j,k}\mathbf{f}_k + \mathbf{e}_{j,k}) + \boldsymbol{\eta}_j \\
\mathbf{g}_j &= \mathbf{D}_j\mathbf{M}_{j,k}\mathbf{f}_k + \mathbf{D}_j\mathbf{e}_{j,k} + \boldsymbol{\eta}_j \\
\mathbf{g}_j &= \mathbf{D}_j\mathbf{M}_{j,k}\mathbf{f}_k + \boldsymbol{\epsilon}_{j,k} + \boldsymbol{\eta}_j \\
\mathbf{g}_j &= \mathbf{D}_j\mathbf{M}_{j,k}\mathbf{f}_k + \boldsymbol{\epsilon}_{j,k} \\
\mathbf{g}_j &= \mathbf{C}_{j,k}\mathbf{f}_k + \boldsymbol{\epsilon}_{j,k}
\end{aligned} \tag{2.5}$$

Em (2.5), $\mathbf{C}_{j,k} = \mathbf{D}_j\mathbf{M}_{j,k}$ é a matriz de *transformação conjunta* de movimento e aquisição. O erro $\boldsymbol{\epsilon}_{j,k}$ é uma soma de dois erros de naturezas diferentes: o ruído de aquisição $\boldsymbol{\eta}_j$ e o erro de movimento projetado nos dados $\boldsymbol{\epsilon}_{j,k}$, o qual pode conter *outliers*. Esse modelo é mais complexo do que o modelo em (2.1) e apresenta dificuldades na modelagem estatística de $\boldsymbol{\epsilon}_{j,k}$ em razão da presença de *outliers*. No Capítulo 3, esse problema será melhor detalhado.

Por causa da existência dos *outliers*, os dados requerem um tratamento especial. Os *outliers* precisam ser detectados e removidos, como em (2.4). Nesse caso, a seguinte equação deve ser adotada para o modelo combinado:

$$\bar{\mathbf{g}}_{j,k} = \bar{\mathbf{C}}_{j,k}\mathbf{f}_k + \bar{\boldsymbol{\epsilon}}_{j,k} \tag{2.6}$$

sendo que $\bar{\mathbf{g}}_{j,k}$, $\bar{\mathbf{C}}_{j,k}$ e $\bar{\boldsymbol{\epsilon}}_{j,k}$ são versões de \mathbf{g}_j , $\mathbf{C}_{j,k}$ e $\boldsymbol{\epsilon}_{j,k}$ nas quais o problema dos *outliers* foi corrigido. No modelo combinado em (2.6) a detecção dos *outliers* é feita em baixa-resolução, de acordo com [46, 1], no modelo de imagem (2.4) a detecção é feita em alta-resolução, de acordo com [1].

O modelo combinado é, geralmente, chamado de *temporalmente não-coincidente*, porque a imagem LR é associada ao instante temporal j enquanto que a imagem HR é associada ao instante temporal k .

2.1.4 Modelos de Imagem

Em geral, super-resolução é um problema mal posto³, seja porque tem infinitas soluções possíveis ou porque a solução possui grande sensibilidade ao ruído [2, 10, 40]. Para resolver tal problema e obter uma solução única e estável, os algoritmos de super-resolução usam informações adicionais sobre a imagem. Nos algoritmos estatísticos Bayesianos, a informação adicional é definida na forma de uma distribuição *a priori* [47, 48], enquanto nos algoritmos determinísticos a informação adicional é tratada como uma restrição [49, 50] ou como uma penalidade de regularização [8, 42, 40].

A informação adicional mais comumente assumida é de que as imagens são suaves; em outras palavras, as variações na intensidade de \mathbf{f}_k são relativamente pequenas [51, 42]. Essa informação adicional

³Um problema mal posto (*ill-posed*) é um problema matemático que tem, pelo menos, uma das seguintes características: não tem solução; tem infinitas soluções; ou a solução é instável em relação à entrada [40, 39]

é, geralmente, expressa na forma:

$$\|\mathbf{R}_k \mathbf{f}_k\| = \delta_{\mathbf{f}_k} \quad (2.7)$$

na qual $\|\mathbf{x}\|$ é uma norma ou outra medida de distância escolhida e \mathbf{R}_k é uma matriz de tamanho $P \times M$ que representa um operador “passa-alta”, como os operadores de diferenças finitas ou Laplaciano discreto e $\delta_{\mathbf{f}_k}$ é uma medida conhecida de $\|\mathbf{R}_k \mathbf{f}_k\|$. O vetor resultante de $\mathbf{R}_k \mathbf{f}_k$ representa as variações na intensidade de \mathbf{f}_k . Em muitos casos, o mesmo operador $\mathbf{R}_k = \mathbf{R}$ é considerado para todas imagens da seqüência. As formas construtivas mais comuns da matriz \mathbf{R} são discutidas no Apêndice A.3.

Entretanto, a definição de um modelo mais completo para a imagem \mathbf{f}_k pode ser feita através de um campo aleatório. Assim sendo, um Campo de Markov (MRF - *Markov Random Field*) [47] é especificado através de uma distribuição de Gibbs [42, 52] com a seguinte densidade de probabilidade:

$$\rho(\mathbf{f}_k) = \frac{1}{q} e^{-\frac{Q(\mathbf{f}_k)}{\beta}} \quad (2.8)$$

na qual q é a constante para normalização da distribuição; β , o parâmetro de controle relacionado ao desvio-padrão da distribuição, também chamado de hiperparâmetro, e $Q(\mathbf{f}_k)$ pode ser definido como:

$$Q(\mathbf{f}_k) = \sum_i V([\mathbf{R}_k \mathbf{f}_k]_i) \quad (2.9)$$

sendo que $[\mathbf{R}_k \mathbf{f}_k]_i$ é um elemento do resultado de $\mathbf{R}_k \mathbf{f}_k$, $V([\mathbf{R}_k \mathbf{f}_k]_i)$ é uma função que aplica um potencial a esse elemento. As funções de potencial mais estudadas são: a quadrática, valor absoluto, Huber [53], valor absoluto elevado à potência $1 \leq p \leq 2$ [54], entre outros [8, 42, 55].

O modelo de imagem, com a especificação de \mathbf{R}_k , do conjunto de β e do potencial $V(x)$, descreve a informação adicional necessária para ser utilizada na busca da imagem de alta resolução.

2.2 Métodos de Super-Resolução

Nesta seção é apresentada uma breve revisão dos principais algoritmos de super-resolução para seqüência de imagens. São eles: Interpolação Não-Uniforme e Restauração, *Iterative Back Projection*, os Métodos Baseados em Algoritmos Adaptativos, Projeção em Conjuntos Convexos (POCS - *Projection Onto Convex Sets*) e os Métodos Regularizados.

2.2.1 Interpolação Não-Uniforme e Restauração

Os métodos de Interpolação Não-Uniforme e Restauração estão entre os primeiros aplicados em super-resolução [56, 57, 58, 59]. Nestes métodos, o problema de super-resolução é separado em dois: um problema de interpolação e um de restauração.

Na etapa de interpolação, uma imagem \mathbf{h}_k com as dimensões da imagem HR é criada a partir da seqüência de baixa resolução. Entretanto, a imagem \mathbf{h}_k ainda possui distorções ópticas que precisam ser corrigidas. O passo seguinte é a restauração para correção das distorções ópticas.

Este método é baseado no modelo combinado e restrito a aplicações nas quais o sistema de aquisição é invariante, ou seja, é utilizada a equação (2.2) para representar a aquisição no modelo combinado, conforme:

$$\begin{aligned}\mathbf{g}_j &= \mathbf{C}_{j,k}\mathbf{f}_k + \boldsymbol{\varepsilon}_{j,k} \\ \mathbf{g}_j &= \mathbf{D}_j\mathbf{M}_{j,k}\mathbf{f}_k + \boldsymbol{\varepsilon}_{j,k} \\ \mathbf{g}_j &= \mathbf{S}_j\mathbf{B}_j\mathbf{M}_{j,k}\mathbf{f}_k + \boldsymbol{\varepsilon}_{j,k}\end{aligned}\tag{2.10}$$

Assume-se que o movimento $\mathbf{M}_{j,k}$ e a degradação óptica \mathbf{B}_j são intercambiáveis, i.e., $\mathbf{B}_j\mathbf{M}_{j,k} = \mathbf{M}_{j,k}\mathbf{B}_k$. Entretanto, essa consideração só é válida em casos restritos, como nas aplicações em que as matrizes são bloco-circulantes e $\mathbf{B}_j = \mathbf{B}_k$. Em [60] é afirmado que, caso o movimento seja estimado a partir das imagens LR, é preferível a utilização de $\mathbf{M}_{j,k}\mathbf{B}_k$. Portanto, em vez de:

$$\mathbf{g}_j = \mathbf{S}_j\mathbf{B}_j\mathbf{M}_{j,k}\mathbf{f}_k + \boldsymbol{\varepsilon}_{j,k},\tag{2.11}$$

assume-se:

$$\mathbf{g}_j = \mathbf{S}_j\mathbf{M}_{j,k}\mathbf{B}_k\mathbf{f}_k + \boldsymbol{\varepsilon}_{j,k}\tag{2.12}$$

como modelo.

O problema é, então, separado em interpolação não-uniforme e restauração. A etapa de interpolação utiliza a seqüência de imagens de baixa resolução, através do seguinte modelo:

$$\mathbf{g}_j = \mathbf{S}_j\mathbf{M}_{j,k}\mathbf{h}_k + \boldsymbol{\varepsilon}_{j,k}\tag{2.13}$$

na qual a interpolação não-uniforme pode ser executada através da utilização do operador linear:

$$\hat{\mathbf{h}}_k = \mathbf{E}_k\mathbf{g}\tag{2.14}$$

sendo que $\mathbf{g} = [\mathbf{g}_1^T \cdots \mathbf{g}_L^T]^T$ é a seqüência de baixa resolução e \mathbf{E}_k é a matriz de interpolação. Métodos utilizados para encontrar a matriz \mathbf{E}_k são vistos em [61, 62, 63].

A imagem interpolada \mathbf{h}_k pode, assim, ser restaurada, segundo [4, 42]:

$$\hat{\mathbf{f}}_k = \tilde{\mathbf{B}}_k^{-1} \mathbf{h}_k \quad (2.15)$$

sendo que $\tilde{\mathbf{B}}_k^{-1}$ é a inversa aproximada, ou regularizada, de \mathbf{B}_k [42, 40].

O único trabalho pertencente a essa classe que aplica super-resolução simultaneamente é o [59]. Todavia, a formulação do problema de super-resolução não foi bem sucedida, resultando em um algoritmo computacionalmente custoso.

O método de Interpolação Não-Uniforme e Restauração é vantajoso nas situações em que é possível usar métodos rápidos para interpolação e para a restauração. Os métodos propostos em [57, 58, 64] proporcionam uma interpolação não-uniforme rápida. A inversa aproximada utilizada na restauração pode ser calculada através da Transformada Rápida de Fourier, dado que a matriz \mathbf{B}_k é bloco-circulante e as matrizes bloco-circulantes são diagonalizáveis através da Transformada de Fourier [4, 42].

As desvantagens do referido procedimento são as limitações impostas pela utilização de matrizes bloco-circulantes e a utilização da mesma distorção óptica para todas as imagens. Além disso, a validade do intercâmbio entre as matrizes é questionável nos casos em que o movimento é complexo ou quando existem *outliers*. Outra desvantagem é que a solução do problema de super-resolução de forma separada não necessariamente leva a uma solução adequada para o problema conjunto. Portanto, tais técnicas não foram consideradas para desenvolvimento, em virtude dessas limitações.

2.2.2 IBP-Iterative Back Projection

A IBP é uma técnica iterativa criada a partir de modificações das técnicas de *Back Projection* utilizadas em tomografia computadorizada [8, 65]. Estas técnicas também estão entre as mais antigas aplicadas na área de super-resolução [66, 67, 68, 69]. O IBP utiliza um procedimento iterativo de minimização do erro entre dados e a saída do modelo teórico, segundo:

$$\mathbf{f}_k^{n+1} = \mathbf{f}_k^n + \sum_{j=1}^L \mathbf{H}_{k,j}^{BP} (\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k^n) \quad (2.16)$$

sendo que n é a iteração corrente e $\mathbf{H}_{k,j}^{BP}$ é operador de *Back Projection*.

Os métodos IBP são bastante similares aos métodos iterativos para solução de mínimos quadrados,

como *steepest descent*, Jacobi e Gauss-Seidel [8, 50, 70, 43]. Uma das grandes vantagens desses métodos é a velocidade de convergência. Entretanto, para que se obtenha uma grande velocidade de convergência, o operador de *back projection* deve ser escolhido para tal. Até o presente momento, nenhum trabalho de super-resolução simultânea utilizando o IBP foi encontrado na literatura.

Devido ao problema de super-resolução ser mal posto, o operador de *back projection* deve ser muito bem especificado; caso contrário, a solução pode divergir ou será dependente da estimativa inicial. Isso ocorre porque esses métodos nem sempre apresentam uma formulação adequada para a função custo. Sem isso não se pode garantir unicidade e estabilidade da solução. Outro problema é que se está restrito a uma mesma ferramenta iterativa para encontrar a solução e não há condições de inserir facilmente informações *a priori* sobre a solução. Devido a essas limitações, esses métodos foram desconsiderados para desenvolvimento neste trabalho.

2.2.3 Métodos Baseados em Algoritmos Adaptativos

Os métodos baseados em algoritmos adaptativos seguem, em geral, a estrutura sequencial. Em [24] um método de reconstrução baseada no algoritmo LMS é utilizado em SR. O objetivo principal destes métodos é reduzir o custo computacional da estimativa, chegando próximo de uma estimativa em tempo real. Contudo, a qualidade da estimativa e a robustez aos *outliers* desses métodos são pequenas. O algoritmo LMS aplicado em super-resolução é apresentado na Tabela 2.1

Em [71], um modelo determinístico que expressa o comportamento destes algoritmos é apresentado. Tal modelo permite melhorar o projeto desses métodos visando à melhora da robustez aos erros no movimento, por exemplo.

Ainda não existem métodos simultâneos baseados em algoritmos adaptativos. Além disso, devido ao problema de super-resolução ser mal posto, é necessário que o algoritmo seja muito bem projetado; caso contrário, a solução pode divergir ou ser dependente da estimativa inicial. A superfície de desempenho também precisa ser analisada para garantir unicidade e estabilidade da solução. Porque esses métodos são direcionados principalmente para o baixo custo computacional, e não para alta qualidade das estimativas, eles foram desconsiderados para desenvolvimento nesta tese.

Tabela 2.1: Algoritmo LMS aplicado em super-resolução

$k := 1$; índice da imagem inicial
 $n := 1$; iteração inicial da imagem k
 $\mathbf{f}_1^1 :=$ imagem inicial é uma interpolação de \mathbf{g}_1
 Algoritmo LMS:
 Laço em $k = 1 \dots L$;
 Laço em $n = 1 \dots N$;
 $\mathbf{f}_k^{n+1} := \mathbf{f}_k^n + \mu \mathbf{D}_k^T (\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k^n)$
 Termina Laço em n ;
 $\mathbf{f}_{k+1}^1 := \mathbf{M}_{k+1,k} \mathbf{f}_k^N$
 Termina Laço em k

2.2.4 POCS - *Projection Onto Convex Sets*

O método POCS é um dos mais utilizados em super-resolução [72, 73, 74, 76] e foi um dos primeiros a apresentar modelos de aquisição espaço-variantes e informações *a priori* na super-resolução. A ideia dos algoritmos POCS é utilizar todos os modelos e informações *a priori* disponíveis para compor uma série de conjuntos convexos [8, 50, 73]. Assim, em um procedimento iterativo, o resultado da iteração corrente é projetado em cada um dos conjuntos, segundo:

$$\mathbf{f}_k^{n+1} = \mathcal{P}_q \dots \mathcal{P}_0 \mathbf{f}_k^n \quad (2.17)$$

sendo que \mathcal{P}_m é o operador de projeção para o m -ésimo conjunto convexo, assumindo que são usados $q + 1$ conjuntos convexos.

A solução desse problema está na intersecção dos conjuntos. Se os conjuntos forem adequadamente especificados haverá, pelo menos, uma solução para o problema.

Entretanto, o método POCS tem algumas desvantagens importantes. Uma das dificuldades é a determinação dos operadores de projeção, os quais podem requerer o cálculo de matrizes inversas ou, em alguns casos, a execução de procedimentos iterativos [8, 50]. Veja um exemplo da projeção no conjunto formado pelos dados capturados. O conjunto é expresso como:

$$\{\mathbf{f}_k : \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2^2 \leq \alpha^2\} \quad (2.18)$$

sendo que o respectivo operador de projeção é:

$$\mathcal{P}\mathbf{f}_k^n = \lambda(\mathbf{D}_k^+ \mathbf{D}_k \mathbf{f}_k^n - \mathbf{D}_k^+ \mathbf{g}_k) + (\mathbf{I} - \mathbf{D}_k^+ \mathbf{D}_k) \mathbf{f}_k^n \quad (2.19)$$

na qual \mathbf{D}_k^+ é a pseudo-inversa de \mathbf{D}_k [8], \mathbf{I} é a matriz identidade e $\lambda \leq \alpha / \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k^n\|_2$.

Em [73, 74, 76] o custo computacional foi reduzido. Em vez de assumir um conjunto convexo para cada imagem LR, assumiu-se um conjunto convexo para cada pixel nas imagens LR, eliminando a necessidade do cálculo da pseudo-inversa e tornando o método POCS mais interessante para super-resolução.

Contudo, no método POCS pode haver problemas de não-unicidade de solução ou inexistência de solução, dependendo da intersecção dos conjuntos. Devido a esse problema, o desenvolvimento através do método POCS foi desconsiderado, mas em razão de sua grande utilização e aceitação não são descartados futuros trabalhos de super-resolução simultânea, utilizando o método POCS. Até o presente momento, nenhum trabalho de super-resolução simultânea baseado em POCS foi encontrado na literatura.

2.2.5 Métodos Regularizados

Em virtude do problema de super-resolução ser mal posto (*ill-posed*), os métodos regularizados buscam uma solução única e estável através da minimização do erro entre os dados e a imagem, acrescentando uma penalidade de regularização [2, 40, 39]. O erro entre dados e imagem é proveniente das equações do modelo de aquisição ou do modelo combinado, ao passo que a penalidade de regularização é proveniente da informação adicional sobre a imagem. A formulação típica dos métodos regularizados [2] é:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \alpha_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_2^2 \quad (2.20)$$

na qual $\|\mathbf{x}\|_2$ representa a norma ℓ_2 , $\alpha_{j,k}$ é uma ponderação aplicada em razão do aumento do erro de movimento com o aumento da distância temporal entre j e k e λ_k é o coeficiente de regularização.

A equação (2.20) é minimizada utilizando métodos de otimização. O método mais comumente encontrado na literatura é o *steepest descent* [40, 49].

Em alguns trabalhos, [21, 46, 75], a norma ℓ_2 da penalidade de regularização é substituída por uma medida de distância que não penaliza, exageradamente, grandes variações de amplitude. São utilizadas métricas como a norma ℓ_1 [75], ou a norma Huber [21, 46], que têm a propriedade de preservar descontinuidades nas variações de amplitude. Em outras palavras, a nitidez das bordas dos objetos existentes nas

imagens estimadas com penalidades desse tipo é maior do que nas imagens estimadas com a norma ℓ_2 .

Em [75] a norma ℓ_1 é utilizada na minimização do erro no modelo combinado, o qual pode conter *outliers*. A propriedade de preservar descontinuidades dessa norma é utilizada para aumentar a robustez do algoritmo aos *outliers*. A robustez tem grande importância prática, já que os procedimentos de pré-deteção e eliminação de *outliers* podem falhar.

Os métodos regularizados têm grande versatilidade na escolha dos modelos e na inclusão de informação adicional. Além disso, há uma grande variedade de ferramentas e métodos de otimização que podem ser considerados. Devido a essas vantagens, optou-se por tais métodos para o desenvolvimento do algoritmo de super-resolução proposto neste trabalho.

Método Regularizado de Super-Resolução Simultânea

Em [1], Borman aprimorou os algoritmos regularizados propondo a utilização de uma penalidade espaço-temporal para regularização. O aprimoramento levou a um aumento significativo da qualidade na estimativa, destacando os métodos simultâneos. Esse tipo de penalidade não era novo; penalidades espaço-temporais já eram empregadas na área de restauração de vídeo [42]. Porém, Borman foi o primeiro a sugerir esse tipo de penalidade em super-resolução. Em [1], também foi observado que, para que bons resultados fossem obtidos, a otimização precisava ser feita de forma simultânea para todas as imagens da seqüência.

A proposta original de [1] é:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \sum_{j=1}^L \alpha_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda \left[\sum_{k=1}^L \mathcal{V}(\mathbf{R}_k \mathbf{f}_k) + \sum_{k=2}^{L-1} \mathcal{V}(-\mathbf{M}_{k,k-1} \mathbf{f}_{k-1} + 2 \mathbf{f}_k - \mathbf{M}_{k,k+1} \mathbf{f}_{k+1}) \right] \quad (2.21)$$

Sendo que a medida de distância utilizada $\mathcal{V}(\mathbf{x})$ é a norma Huber [53]. O método de otimização proposto foi o *Steepest Descent*.

Após a proposta [1], nenhum outro esforço no sentido de melhorar os algoritmos simultâneos foi encontrado na literatura.

Determinação do Coeficiente de Regularização

Um problema existente nos métodos regularizados, tanto nos tradicionais quanto nos simultâneos é a determinação do coeficiente de regularização λ e das ponderações $\alpha_{j,k}$. Na literatura de problemas

inversos, dos quais os métodos regularizados surgiram, diversos métodos têm sido desenvolvidos para encontrar o coeficiente de regularização, como o Generalized Cross-Validation (GCV) [40] e os métodos baseados na Curva-L [39, 77, 78]. Na literatura estatística busca-se a estimação dos parâmetros das distribuições estatísticas assumidas no problema, os quais formam o coeficiente de regularização. Os métodos desenvolvidos para este fim são: Método Evidence [79] e JMAP [80].

Algumas outras alternativas, desenvolvidas de forma intuitiva, para determinar o parâmetro de regularização e as ponderações aparecem em [37, 38]. Contudo, estes métodos foram desenvolvidos apenas para os algoritmos SR tradicionais; ainda estão indisponíveis para os algoritmos SR simultâneos. A busca por um método de determinação dos coeficientes de regularização é a segunda parte desta tese. No Capítulo 5, é apresentada uma revisão mais aprofundada e específica sobre os métodos de determinação dos coeficientes de regularização.

2.3 Fundamentos Teóricos dos Métodos Regularizados

Nesta seção são apresentadas duas linhas de pesquisa dentro dos métodos regularizados de super-resolução. A primeira linha de pesquisa é a determinística regularizada [8, 40, 39], cujas principais vantagens são a simplicidade e a interpretação geométrica do problema. A segunda é a linha estatística Bayesiana [8, 40, 48], cuja principal vantagem é a facilidade no tratamento das imprecisões e incertezas nos dados medidos e nos modelos assumidos. Será demonstrado que as duas linhas levam a soluções idênticas em muitos aspectos, porém mantêm diferentes pontos de vista sobre o problema.

2.3.1 Métodos Determinísticos Regularizados

Através dos métodos determinísticos, tem-se uma forma simples e direta de especificar uma solução para o problema de super-resolução. É preciso, simplesmente, definir uma medida de discrepância entre os dados e a saída do modelo teórico e minimizá-la. Como, por exemplo, a norma ℓ_2 , segundo:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2^2 = \arg \min_{\mathbf{f}_k} \sum_{i=1}^N |[\mathbf{g}_k]_i - [\mathbf{D}_k \mathbf{f}_k]_i|^2 \quad (2.22)$$

sendo que $[\mathbf{x}]_i$ é o elemento i do vetor \mathbf{x} . Tem-se, assim, a estimativa tradicional de mínimos quadrados.

Em uma forma mais generalizada, pode-se definir:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} J(\mathbf{g}_k, \mathbf{D}_k \mathbf{f}_k) \quad (2.23)$$

sendo que $J(\mathbf{x}, \mathbf{y})$ é uma medida de distância ou discrepância entre o vetor \mathbf{x} e o vetor \mathbf{y} . As medidas de discrepância mais utilizadas são: Mínimos Quadrados, Mínimos Quadrados Ponderados ou Generalizados, Norma ℓ_p , Norma Huber e a Função de Kullback-Leibler. Essas funções de discrepância estão relacionadas no Apêndice A.4.

Fazendo uma simples observação na matriz \mathbf{D}_k , de tamanho $N \times M$, sendo $N < M$, sabe-se que o sistema é subdeterminado, ou seja, ele tem mais incógnitas do que equações. Portanto, (2.23) é um problema mal posto, pois possui infinitas soluções.

Para restringir o número de soluções possíveis e tentar obter uma solução única, os algoritmos de super-resolução utilizam o modelo combinado, apresentado em (2.5). A equação de aquisição, em (2.1), juntamente com as equações combinadas de aquisição e movimento, são utilizadas para aumentar o número de equações do sistema e restringir o número de soluções possíveis para \mathbf{f}_k .

Assumindo que o conjunto de imagens capturadas é formado por L imagens, tem-se:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L J(\mathbf{g}_j, \mathbf{C}_{j,k} \mathbf{f}_k) \quad (2.24)$$

Tal equacionamento pode ser representado através da seguinte concatenação:

$$\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_L \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{1,k} \\ \vdots \\ \mathbf{C}_{L,k} \end{bmatrix} \mathbf{f}_k + \begin{bmatrix} \boldsymbol{\varepsilon}_{1,k} \\ \vdots \\ \boldsymbol{\varepsilon}_{L,k} \end{bmatrix} \quad (2.25)$$

sendo que, de acordo com (2.5), $\mathbf{C}_{j,k} = \mathbf{D}_j \mathbf{M}_{j,k}$. Observe que, quando os subíndices são os mesmos, a matriz de transformação de movimento equivale à matriz Identidade, ou seja, $\mathbf{C}_{k,k} = \mathbf{D}_k \mathbf{M}_{k,k} = \mathbf{D}_k$ e o erro $\boldsymbol{\varepsilon}_{k,k} = \boldsymbol{\eta}_k$, reduzindo-se ao modelo da aquisição.

A equação (2.25) pode ser representada de forma compacta por:

$$\mathbf{g} = \mathbf{C}_k \mathbf{f}_k + \boldsymbol{\varepsilon}_k \quad (2.26)$$

sendo $\mathbf{g} = [\mathbf{g}_1^T \dots \mathbf{g}_L^T]^T$ e $\boldsymbol{\varepsilon}_k = [\boldsymbol{\varepsilon}_{1,k}^T \dots \boldsymbol{\varepsilon}_{L,k}^T]^T$. A matriz \mathbf{C}_k , definida pelo conjunto de matrizes em (2.25), tem tamanho $LN \times M$.

A minimização em (2.24) é equivalente a:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k) \quad (2.27)$$

No entanto, as equações (2.24) e (2.27) podem trazer resultados insatisfatórios, porque, mesmo quando $LN > M$, não se pode garantir a unicidade da solução ou, ainda, a solução pode ser extremamente sensível aos erros nos dados. A unicidade da solução não pode ser garantida porque não se pode afirmar como certo que o posto da matriz \mathbf{C}_k é igual a M . Essa condição é necessária para a unicidade [40]. A estabilidade da solução, com respeito aos erros de medição, também não pode ser assegurada porque depende do condicionamento do sistema, que, em geral, é mal condicionado.

Para garantir unicidade e estabilizar a solução é possível introduzir uma penalidade de regularização, segundo:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k) + \lambda Q(\mathbf{f}_k) \quad (2.28)$$

sendo que $Q(\mathbf{x})$ é a penalidade de regularização usada para estabilizar a solução do problema, o λ é o coeficiente de regularização, que controla a influência da penalidade de regularização. Tal forma de regularização é conhecida como Regularização Generalizada de Tikhonov [8, 40, 39, 48]. Mais à frente será demonstrado que $Q(\mathbf{f}_k)$ pode ser a mesma utilizada no modelo de imagem, apresentado na seção 2.1.4.

Contudo, não é fácil definir qual é a melhor escolha para a medida de discrepância $J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k)$ nem para a penalidade $Q(\mathbf{f}_k)$. Pode-se escolher a medida de discrepância baseando-se no conhecimento de características comumente obtidas por essas medidas [48]. Por exemplo: mínimos quadrados levam a soluções lineares e computacionalmente mais simples; norma ℓ_p , para $1 < p < 2$, e norma Huber são mais robustas a *outliers* [44]; ou, ainda, a função de Kullback-Leibler garante que a solução seja positiva, considerando que a intensidade da imagem sempre é positiva [48].

Pode-se escolher a penalidade de regularização da mesma forma. Porém, isso não proporciona o tratamento mais adequado para os erros nos dados, nem para as incertezas nos modelos assumidos. Por exemplo, não se está explorando o fato de que determinado erro possa seguir uma distribuição conhecida. Essas noções de imprecisão e incerteza são muito mais fáceis de tratar através dos métodos estatísticos. Conforme [48], os métodos estatísticos foram inventados para tratar exatamente dessas noções.

2.3.2 Métodos Estatísticos Bayesianos

A principal característica dos métodos estatísticos é considerar os dados \mathbf{g} como a observação de um vetor aleatório. Em particular, os métodos de inferência Bayesianos consideram que \mathbf{f}_k também é um vetor aleatório e que todas as estimativas de \mathbf{f}_k são baseadas na densidade de probabilidade *a posteriori* $\rho(\mathbf{f}_k | \mathbf{g})$.

A densidade de probabilidade *a posteriori* é, geralmente, desconhecida e difícil de determinar, entretanto ela pode ser mais facilmente escrita, através da Regra de Bayes, como:

$$\rho(\mathbf{f}_k|\mathbf{g}) = \frac{\rho(\mathbf{g}|\mathbf{f}_k)\rho(\mathbf{f}_k)}{\rho(\mathbf{g})} \quad (2.29)$$

sendo que $\rho(\mathbf{g}|\mathbf{f}_k)$ é a densidade de probabilidade dos dados condicionados a \mathbf{f}_k , $\rho(\mathbf{f}_k)$ é a densidade de probabilidade *a priori*, e $\rho(\mathbf{g})$ é a densidade marginal de \mathbf{g} .

As densidades $\rho(\mathbf{g}|\mathbf{f}_k)$ e $\rho(\mathbf{f}_k)$ são, em geral, mais fáceis de encontrar ou definir. Todavia, determinar a densidade de probabilidade *a posteriori* $\rho(\mathbf{f}_k|\mathbf{g})$ não é uma tarefa fácil [8, 47, 48]. Além disso, está-se buscando uma boa solução para \mathbf{f}_k e não, necessariamente, sua probabilidade de ocorrência, que é especificada por $\rho(\mathbf{f}_k|\mathbf{g})$.

Tendo esse objetivo, a solução para \mathbf{f}_k pode ser definida como, por exemplo, o vetor \mathbf{f}_k que possui a máxima probabilidade na densidade de probabilidade *a posteriori*. Esta estimativa é chamada de estimativa MAP (*maximum a posteriori*):

$$\hat{\mathbf{f}}_k = \arg \max_{\mathbf{f}_k} \rho(\mathbf{f}_k|\mathbf{g}) \quad (2.30)$$

As semelhanças entre os métodos determinísticos e os estatísticos MAP têm sido bastante discutidas [8, 48, 55], e podem ser demonstradas através das seguintes manipulações:

$$\begin{aligned} \hat{\mathbf{f}}_k &= \arg \max_{\mathbf{f}_k} \rho(\mathbf{f}_k|\mathbf{g}) \\ \hat{\mathbf{f}}_k &= \arg \max_{\mathbf{f}_k} \frac{\rho(\mathbf{g}|\mathbf{f}_k)\rho(\mathbf{f}_k)}{\rho(\mathbf{g})} \end{aligned} \quad (2.31)$$

nas quais $\rho(\mathbf{g})$ não influencia na maximização e pode ser eliminado. Aplicando o logaritmo, que não muda o ponto de máximo, tem-se:

$$\begin{aligned} \hat{\mathbf{f}}_k &= \arg \max_{\mathbf{f}_k} \rho(\mathbf{g}|\mathbf{f}_k)\rho(\mathbf{f}_k) \\ \hat{\mathbf{f}}_k &= \arg \max_{\mathbf{f}_k} \ln \rho(\mathbf{g}|\mathbf{f}_k) + \ln \rho(\mathbf{f}_k) \end{aligned} \quad (2.32)$$

e transformando a maximização em minimização, tem-se:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} -\ln \rho(\mathbf{g}|\mathbf{f}_k) - \ln \rho(\mathbf{f}_k) \quad (2.33)$$

As semelhanças entre (2.28) e (2.33) já são evidentes. Observe que a solução dos métodos determinísticos regularizados é equivalente à solução de máxima probabilidade *a posteriori* para distribuições do tipo:

$$\rho(\mathbf{g}|\mathbf{f}_k) = \frac{1}{j} e^{-\frac{J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k)}{\theta}} \quad (2.34)$$

no qual θ é o parâmetro de controle, ou hiperparâmetro, da distribuição $\rho(\mathbf{g}|\mathbf{f}_k)$, e

$$\rho(\mathbf{f}_k) = \frac{1}{q} e^{-\frac{Q(\mathbf{f}_k)}{\beta}} \quad (2.35)$$

que é o modelo de imagem, apresentado na seção 2.1.4.

Observe que, de acordo com a equivalência entre (2.24) e (2.27), se está assumindo independência na densidade dos dados condicionados à \mathbf{f}_k , ou seja:

$$\rho(\mathbf{g}|\mathbf{f}_k) = \prod_{j=1}^K \rho(\mathbf{g}_j|\mathbf{f}_k) \quad (2.36)$$

sendo

$$\rho(\mathbf{g}_j|\mathbf{f}_k) = \frac{1}{j_j} e^{-\frac{J(\mathbf{g}_j, \mathbf{C}_{j,k} \mathbf{f}_k)}{\theta}} \quad (2.37)$$

Substituindo as densidades (2.34) e (2.35) na equação (2.33), obtém-se:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \frac{J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k)}{\theta} + \ln j + \frac{Q(\mathbf{f}_k)}{\beta} + \ln q \quad (2.38)$$

Observe que j e q não influenciam na minimização e podem ser desconsiderados; logo:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k) + \frac{\theta}{\beta} Q(\mathbf{f}_k) \quad (2.39)$$

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k) + \lambda Q(\mathbf{f}_k) \quad (2.40)$$

no qual λ é o coeficiente de regularização, definido como $\lambda = \theta/\beta$.

Assim, o ponto de vista estatístico pode ser utilizado para determinar a medida de discrepância mais adequada ao problema, para determinar a escolha da penalidade de regularização, e, inclusive, para indicar o coeficiente de regularização mais apropriado. Geralmente, $J(\mathbf{g}, \mathbf{C}_k \mathbf{f}_k)$ é chamado de *termo de*

$dados$ e $Q(\mathbf{f}_k)$ é chamado de *termo de informação a priori*.

Conclusão do Capítulo

Neste capítulo, os modelos utilizados em super-resolução, as principais técnicas de super-resolução e a teoria dos algoritmos regularizados foram revisados. Os modelos utilizados em super-resolução são necessários para o resto do trabalho e serão frequentemente referenciados. Na revisão das principais técnicas de super-resolução ficou estabelecido que o trabalho proposto será desenvolvido a partir dos métodos regularizados, devido às suas vantagens. Tanto o ponto de vista determinístico quanto o ponto de vista estatístico serão utilizados para justificar as escolhas feitas no decorrer da tese.

Capítulo 3

Super-Resolução Simultânea

Neste capítulo, a primeira contribuição desta tese é apresentada: o aprimoramento dos métodos regularizados de super-resolução simultânea. Na seção 3.1, são mostrados os problemas encontrados no algoritmo simultâneo existente, proposto em [1]. São eles: o alto custo computacional para estimação da seqüência de imagens e a baixa robustez na presença de *outliers*.

Na seção 3.2, o método simultâneo proposto é introduzido, o qual reduz significativamente as dificuldades encontradas nos métodos simultâneos existentes, mantendo os benefícios trazidos pelos mesmos. Inicialmente, é apresentada a idéia principal, que se resume na utilização do movimento exclusivamente como informação *a priori*. A partir dessa idéia, é definida uma nova classe de algoritmos regularizados de super-resolução simultânea.

Finalmente, na seção 3.3, são apresentadas as comparações teóricas, contrastando o algoritmo proposto com os métodos tradicionais e simultâneos existentes na literatura. Nesta comparação, os métodos citados são avaliados quanto à qualidade obtida nos resultados, quanto ao custo computacional e quanto à robustez aos *outliers*.

3.1 Problemas Identificados no Método Simultâneo Existente

Nesta seção, são discutidos dois problemas identificados nos algoritmos existentes de super-resolução simultânea. São eles: o alto custo computacional para estimação da seqüência de imagens e a baixa robustez da estimativa na presença de *outliers*.

3.1.1 Alto Custo Computacional

Os algoritmos regularizados de super-resolução são, em geral, de alta complexidade computacional, mesmo quando a aplicação em questão requer a estimação de uma única imagem da seqüência. Quando se deseja a obtenção de toda a seqüência de imagens, o custo computacional é ainda maior. Considere o caso dos algoritmos de super-resolução regularizados tradicionais. Se o objetivo for encontrar uma única imagem, utiliza-se a seguinte equação:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \alpha_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_k \mathcal{V}(\mathbf{R}_k \mathbf{f}_k) \quad (3.1)$$

O custo computacional deste método, para solução da imagem \mathbf{f}_k , é proporcional à L , que é o número de imagens LR utilizadas, como será demonstrado na seção 3.3.3.

Porém, se o objetivo da aplicação for encontrar toda a seqüência de imagens, tem-se que utilizar a seguinte equação:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \left[\sum_{j=1}^L \alpha_{k,j} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_k \mathcal{V}(\mathbf{R}_k \mathbf{f}_k) \right] \quad (3.2)$$

Neste caso, o custo computacional passa a ser proporcional à L^2 ; em outras palavras, os métodos tradicionais não aproveitam os resultados prévios para reduzir o custo computacional.

Essa mesma estrutura é utilizada pelo algoritmo simultâneo proposto em [1], sendo que a equação utilizada é:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \sum_{j=1}^L \alpha_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda \left[\sum_{k=1}^L \mathcal{V}(\mathbf{R}_k \mathbf{f}_k) + \sum_{k=2}^{L-1} \mathcal{V}(-\mathbf{M}_{k,k-1} \mathbf{f}_{k-1} + 2 \mathbf{f}_k - \mathbf{M}_{k,k+1} \mathbf{f}_{k+1}) \right] \quad (3.3)$$

Todos os algoritmos regularizados encontrados na literatura, quando aplicados para a solução de seqüência de imagens, utilizam o mesmo conjunto de equações no termo de dados. O uso destas L^2 equações tem grau de redundância muito grande, que pode ser reduzido. A redução da redundância das equações pode ser convertida em redução de custo computacional.

3.1.2 Baixa Robustez do Método na Presença de *Outliers*

Um problema típico dos métodos de super-resolução que utilizam o movimento é a presença de *outliers*. Os *outliers* são observados com grandes erros de movimento, geralmente em regiões espaciais bem localizadas. Eles podem causar fortes distorções nas imagens estimadas.

Os algoritmos de SR que exploram o movimento forçam com que ocorra uma similaridade entre as imagens através do movimento. Isso faz com que a informação contida em uma determinada imagem seja transferida para uma outra imagem, em um instante temporal diferente, com o objetivo de recuperar detalhes perdidos no processo de aquisição. Contudo, ao forçar esta similaridade em uma região de *outliers*, onde a similaridade não existe, o efeito é contrário, ou seja, aparecem distorções nas imagens estimadas em vez de melhoria de resolução espacial.

A capacidade de um algoritmo de super-resolução de aproveitar a similaridade no movimento onde ela existe e descartá-la onde ela não existe, evitando as distorções, é conhecida, neste contexto, como *robustez aos outliers*. Quase todos os algoritmos SR regularizados, desenvolvidos até o momento, são baseados na utilização da norma ℓ_2 ao quadrado nas expressões do modelo combinado. Algoritmos desse tipo têm baixa robustez porque os erros são elevados ao quadrado e, portanto, os grandes erros, ou *outliers*, têm grande influência nos resultados, causando fortes desvios.

Uma forma de aumentar a robustez dos algoritmos baseados na norma ℓ_2 ao quadrado é utilizar uma pré-deteção e remoção dos *outliers*. Entretanto, na prática este procedimento não é muito seguro, porque o processo pode apresentar falhas na detecções, ou seja, alguns *inliers* podem ser erradamente eliminados enquanto alguns *outliers* remanescentes podem continuar causando distorções.

Uma solução recente para aumentar a robustez dos algoritmos tradicionais foi apresentada em [81, 75]. Nestes métodos, a norma ℓ_1 foi utilizada em substituição à norma ℓ_2 ao quadrado. A função custo proposta em [75] é:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_1 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_1 \quad (3.4)$$

De acordo com a teoria de estimação robusta, em [44], a norma ℓ_1 considera linearmente os erros. Assim, os *outliers*, ou grandes erros, não têm uma influência tão grande no resultado quanto no caso da norma ℓ_2 ao quadrado; aumentando significativamente a robustez a *outliers* do algoritmo de super-resolução.

Contudo, os métodos simultâneos existentes ainda não exploram esta característica e estão suscetíveis aos problemas dos *outliers*. Alternativas baseadas na teoria de estimação robusta de [44] podem ser

incorporadas nos métodos simultâneos, com a finalidade de aumentar a robustez destes métodos.

3.2 Solução Proposta de Super-Resolução Simultânea

Nesta seção, a primeira contribuição da tese é apresentada: o método simultâneo proposto. Inicialmente, a idéia principal do método é introduzida: a utilização do movimento exclusivamente como informação *a priori*. Em seguida, a nova classe de algoritmos de super-resolução simultâneos, baseada na utilização da idéia principal, é definida. São apresentados o modelo geral proposto, as opções de implementação da nova classe de algoritmos e, ao final, uma comparação teórica entre o método proposto e os métodos existentes.

3.2.1 Idéia Principal

A idéia principal do trabalho pode ser resumida como: o movimento é utilizado exclusivamente como informação *a priori*. Esta seção mostra como a idéia principal busca reduzir os dois problemas apresentados na seção 3.1.

A princípio, tem-se que a principal vantagem dos métodos simultâneos é o uso do movimento no termo de regularização, no qual estão incluídas expressões do tipo:

$$\|\mathbf{f}_k - \mathbf{M}_{k,j}\mathbf{f}_j\| = \|\mathbf{e}_{k,j}\| \quad (3.5)$$

Estas equações proporcionam a produção de uma seqüência de imagens com maior similaridade na trajetória de movimento, conforme será explicado na seção 3.3.1. Ao combinar a minimização de expressões desse tipo com a minimização das equações do modelo de aquisição, dado por:

$$\|\mathbf{g}_k - \mathbf{D}_k\mathbf{f}_k\| = \|\boldsymbol{\eta}_k\| \quad (3.6)$$

é possível obter uma solução que satisfaça às seguintes expressões:

$$\hat{\mathbf{f}}_k = \mathbf{M}_{k,j}\hat{\mathbf{f}}_j + \mathbf{e}_{k,j} \quad (3.7)$$

e

$$\mathbf{g}_k = \mathbf{D}_k\hat{\mathbf{f}}_k + \boldsymbol{\eta}_k \quad (3.8)$$

além de, indiretamente, satisfazer o modelo combinado, sem considerá-lo diretamente no problema.

Sendo:

$$\begin{aligned}\mathbf{g}_k &= \mathbf{D}_k(\mathbf{M}_{k,j}\hat{\mathbf{f}}_j + \mathbf{e}_{k,j}) + \boldsymbol{\eta}_k \\ \mathbf{g}_k &= \mathbf{D}_k\mathbf{M}_{k,j}\hat{\mathbf{f}}_j + \mathbf{D}_k\mathbf{e}_{k,j} + \boldsymbol{\eta}_k \\ \mathbf{g}_k &= \mathbf{C}_{k,j}\hat{\mathbf{f}}_j + \boldsymbol{\varepsilon}_{k,j}\end{aligned}\quad (3.9)$$

o resultado proporcionado pelo modelo combinado.

Observe a função custo do método proposto, na forma mais simples, é:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \|\mathbf{g}_k - \mathbf{D}_k\mathbf{f}_k\|_2^2 + \lambda_R \sum_{k=1}^L \mathcal{V}(\mathbf{R}_k\mathbf{f}_k) + \lambda_M \sum_{k=1}^{L-1} \mathcal{V}(\mathbf{f}_k - \mathbf{M}_{k,k+1}\mathbf{f}_{k+1}) \quad (3.10)$$

Note que a equação (3.10) envolve equações do tipo (3.5) e (3.6), sem envolver as equações do modelo combinado.

A redução do número de termos nas equações em (3.10), em comparação com (3.3), normalmente não ocasiona perdas de qualidade em razão da recursividade do modelo auto-regressivo, assumido para o algoritmo. A minimização da diferença entre quadros adjacentes, no termo de informação *a priori*, proporciona a propagação da informação necessária para super-resolução entre os quadros da seqüência. Por exemplo, o processo de otimização visa a minimização da função de custo, forçando com que \mathbf{f}_1 se aproxime de $\mathbf{M}_{1,2}\mathbf{f}_2$ e \mathbf{f}_2 se aproxime de $\mathbf{M}_{2,3}\mathbf{f}_3$; logo, também proporciona uma propagação entre quadros tal que \mathbf{f}_1 se aproxime de $\mathbf{M}_{1,2}\mathbf{M}_{2,3}\mathbf{f}_3 = \mathbf{M}_{1,3}\mathbf{f}_3$. Dessa forma, cada imagem da seqüência estará contribuindo para a reconstrução dos detalhes em todas as outras. Essa é uma característica exclusiva dos métodos simultâneos, tanto do método existente em [1] quanto do método proposto.

Utilizando-se desse princípio, o método proposto deve atingir uma solução que satisfaz as mesmas condições que os demais métodos, mas com um número reduzido de equações. Isso irá reduzir significativamente o custo computacional do método simultâneo proposto em relação ao método simultâneo existente, em [1]. Além disso, as equações de movimento, isoladas no termo *a priori*, permitem que os *outliers* tenham um tratamento diferente dos demais erros, ou seja, a medida de discrepância utilizada nesse termo pode ser escolhida de forma a se obter uma robustez maior aos *outliers*, sem que a medida de discrepância utilizada no termo de dados precise ser modificada. Isso permite, por exemplo, definir a medida de discrepância do termo de dados a partir da estatística do ruído de aquisição.

Dessa forma, a idéia principal permite uma redução de custo computacional a partir da eliminação das expressões contendo o modelo combinado e, ao mesmo tempo, permite um controle maior sobre o erro de movimento, podendo proporcionar maior robustez aos *outliers*.

3.2.2 Modelo Geral Proposto

Esta nova classe de algoritmos simultâneos é composta por algoritmos de super-resolução baseados na utilização da idéia principal, ou seja, em comum, todos os algoritmos utilizam o movimento somente no termo de informação *a priori*. Isso significa que não será mais utilizado o modelo combinado no termo de dados. Em contrapartida, o modelo de movimento no termo *a priori* será generalizado para permitir maior controle das relações temporais. A não-utilização do modelo combinado e a generalização do termo de movimento na regularização são, até agora, as maiores diferenças entre o método proposto e os demais algoritmos regularizados existentes na literatura.

O desenvolvimento desses algoritmos é apresentado através dos métodos estatísticos Bayesianos. De acordo com a teoria dos métodos estatísticos MAP [8, 82], tem-se:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \max_{\mathbf{f}_1, \dots, \mathbf{f}_L} \rho(\mathbf{f}_1, \dots, \mathbf{f}_L | \mathbf{g}_1, \dots, \mathbf{g}_L) \quad (3.11)$$

Considerando a seqüência de baixa resolução como $\mathbf{g} = [\mathbf{g}_1^T \dots \mathbf{g}_L^T]^T$ e a seqüência de alta resolução como $\mathbf{f} = [\mathbf{f}_1^T \dots \mathbf{f}_L^T]^T$, tem-se:

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \rho(\mathbf{f} | \mathbf{g}) \quad (3.12)$$

Aplicando a Regra de Bayes, pode-se representar (3.12) na forma:

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \frac{\rho(\mathbf{g} | \mathbf{f}) \rho(\mathbf{f})}{\rho(\mathbf{g})} \quad (3.13)$$

Como $\rho(\mathbf{g})$ não influencia na maximização, pode ser eliminada.

Em seguida, aplica-se o logaritmo:

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \ln \rho(\mathbf{g} | \mathbf{f}) + \ln \rho(\mathbf{f}) \quad (3.14)$$

e transforma-se o problema de maximização em minimização:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} - \ln \rho(\mathbf{g} | \mathbf{f}) - \ln \rho(\mathbf{f}) \quad (3.15)$$

Para continuar o desenvolvimento, tem-se que definir as densidades para o termo de dados e para o

termo de informação *a priori*. Entretanto, o objetivo é que a equação resultante tenha a seguinte forma:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} J(\mathbf{g}, \mathbf{D}\mathbf{f}) + \lambda Q(\mathbf{f}), \quad (3.16)$$

Termo de Dados

Para o termo de dados, será considerado somente o conjunto de equações da aquisição, dado pela equação (2.1). Para tornar as equações mais compactas será utilizada a seguinte concatenação:

$$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_L \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D}_L \end{bmatrix} \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_L \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_L \end{bmatrix}, \text{ ou } \mathbf{g} = \mathbf{D}\mathbf{f} + \boldsymbol{\eta} \quad (3.17)$$

A medida de discrepância escolhida para o termo de dados está relacionada com a distribuição estatística assumida para o ruído de aquisição. A lista de distribuições que podem ser consideradas neste método e suas medidas de discrepância equivalentes estão no Apêndice B.1. Como exemplo, será considerado o caso em que o ruído de aquisição é um ruído gaussiano branco. Tem-se, então, que a densidade $\rho(\mathbf{g}|\mathbf{f})$ será:

$$\rho(\mathbf{g}|\mathbf{f}) = \frac{1}{(2\pi\sigma_\eta^2)^{\frac{LN}{2}}} e^{-\frac{1}{2\sigma_\eta^2} \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2^2} \quad (3.18)$$

sendo que σ_η é o desvio-padrão do ruído de aquisição.

Comparando (3.18) à seguinte densidade de probabilidade para os dados:

$$\rho(\mathbf{g}|\mathbf{f}) = \frac{1}{j} e^{-\frac{1}{\theta_\eta} J(\mathbf{g}, \mathbf{D}\mathbf{f})} \quad (3.19)$$

têm-se as seguintes equivalências,

$$J(\mathbf{g}, \mathbf{D}\mathbf{f}) = \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2^2 = \sum_{k=1}^L \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2^2 \quad (3.20)$$

e $\theta_\eta = 2\sigma_\eta^2$

Termo de Informação *A Priori*

O termo de informação *a priori* descreve o conhecimento prévio sobre a seqüência de imagens na ausência de dados. De acordo com a idéia principal, o termo de informação *a priori* deve ser composto por equações do tipo apresentado em (3.5). Além disso, esse termo deve conter as matrizes \mathbf{R}_k para proporcionar suavidade espacial.

Partindo do modelo de campos aleatórios para imagens, apresentado em (2.8), chega-se ao modelo generalizado para seqüência de imagens:

$$\rho(\mathbf{f}) = \frac{1}{q} e^{-\frac{Q(\mathbf{f})}{\beta}} \quad (3.21)$$

Este modelo descreve duas características essenciais para a seqüência de imagens:

1. a seqüência deve possuir variações espaciais de pequena amplitude, ou seja, as imagens são espacialmente suaves;
2. as variações de amplitude ao longo da trajetória de movimento são mínimas, ou seja, a diferença de amplitude entre imagens, com o respectivo movimento compensado, é bem pequena.

Para tornar essas duas características distintas, a penalidade será separada em duas: uma para representar a suavidade espacial e outra para representar a suavidade ao longo da trajetória de movimento.

$$\rho(\mathbf{f}) = \frac{1}{q} e^{-\left[\sum_i \frac{V_R([\mathbf{R}\mathbf{f}]_i)}{\beta_{Ri}} + \sum_j \frac{V_M([\mathbf{M}\mathbf{f}]_j)}{\beta_{Mj}} \right]} \quad (3.22)$$

Tem-se que o termo $\sum_i \frac{V_R([\mathbf{R}\mathbf{f}]_i)}{\beta_{Ri}}$, em (3.22), é a penalidade espacial, derivada diretamente do modelo de imagem em (2.7). O potencial $V_R(x)$ é escolhido de acordo com a estatística das variações espaciais de \mathbf{f} .

Pode-se representar \mathbf{R} como:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_L \end{bmatrix} \quad (3.23)$$

O caso mais comum, e também adotado nos experimentos, é o seguinte: assume-se operadores do mesmo tipo para todas as imagens da seqüência, $\mathbf{R}_k = \mathbf{R}$, e assume-se que as medidas de variação espacial são as mesmas para todos elementos $\beta_{Ri} = \beta_R$.

Já o termo $\sum_j \frac{V_M(\mathbf{M}\mathbf{f}_j)}{\beta_{Mj}}$, em (3.22), é a penalidade de movimento. Este termo pode ser construído de várias formas utilizando os termos do modelo de movimento, apresentado em (2.3). No Apêndice B.2, são apresentados diversos exemplos construtivos do operador \mathbf{M} .

Aqui, os dois casos mais comuns são apresentados. São eles:

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} & -\mathbf{M}_{1,2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & -\mathbf{M}_{2,3} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & -\mathbf{M}_{3,4} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I} & -\mathbf{M}_{L-1,L} \end{bmatrix} \quad (3.24)$$

que corresponde à diferença finita de primeira ordem na trajetória do movimento, chamado simplesmente de modelo de primeira ordem. Em (3.24), $\mathbf{M}_{j,k}$ é a transformação de movimento apresentada na seção 2.1.2, e \mathbf{I} é a matriz Identidade. Esta matriz aplicada no termo *a priori* em (3.27) equivale ao termo de movimento utilizado em (3.10). O modelo de segunda ordem é:

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} & -\mathbf{M}_{1,2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{M}_{2,1} & 2\mathbf{I} & -\mathbf{M}_{2,3} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{M}_{3,2} & 2\mathbf{I} & -\mathbf{M}_{3,4} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{M}_{K-1,K-2} & 2\mathbf{I} & -\mathbf{M}_{K-1,K} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{M}_{K,K-1} & \mathbf{I} \end{bmatrix} \quad (3.25)$$

que corresponde à diferença finita de segunda ordem na trajetória do movimento, equivalente ao utilizado no método simultâneo de [1], apresentado na equação (3.3). Além dos modelos de primeira e segunda ordem, modelos complexos de mais alta ordem podem ser considerados.

Assim como na penalidade espacial, na penalidade de movimento pode-se assumir que as medidas do erro na trajetória de movimento são as mesmas para toda a seqüência, ou seja, $\beta_{Mi} = \beta_M$. Também, será considerado o mesmo potencial para as duas penalidades.

Através dessas considerações, tem-se o seguinte modelo para a seqüência:

$$\rho(\mathbf{f}) = \frac{1}{q} e^{-\left[\frac{1}{\beta_R} \sum_i V([\mathbf{R}\mathbf{f}]_i) + \frac{1}{\beta_M} \sum_j V([\mathbf{M}\mathbf{f}]_j) \right]} \quad (3.26)$$

sendo que o somatório é elemento a elemento.

A penalidade equivalente de (3.26) é

$$\frac{Q(\mathbf{f})}{\beta} = \frac{\mathcal{V}(\mathbf{R}\mathbf{f})}{\beta_R} + \frac{\mathcal{V}(\mathbf{M}\mathbf{f})}{\beta_M} \quad (3.27)$$

sendo $\mathcal{V}(\mathbf{x}) = \sum_i V([\mathbf{x}]_i)$.

Pode-se expressar $\mathcal{V}(\mathbf{R}\mathbf{f}) = \sum_k \mathcal{V}(\mathbf{R}_k \mathbf{f}_k)$, porém, para expressar o termo $\mathcal{V}(\mathbf{M}\mathbf{f})$ em função das imagens da seqüência é necessário definir a forma da matriz \mathbf{M} , como, por exemplo, no caso de primeira ordem, sendo $\mathcal{V}(\mathbf{M}\mathbf{f}) = \sum_k \mathcal{V}(\mathbf{f}_k - \mathbf{M}_{k,k-1} \mathbf{f}_{k-1})$.

3.2.3 Implementação Proposta

Embora diferentes penalidades e medidas de discrepância possam ser utilizadas, foram consideradas apenas penalidades e medidas de discrepância convexas. A convexidade é uma condição suficiente para que a função custo tenha um único mínimo global [40, 49]. Garantindo a convexidade da função custo, podem-se utilizar diversos métodos de otimização, principalmente os métodos baseados no gradiente [42, 49].

Os métodos de otimização mais adequados para utilização nos algoritmos regularizados de super-resolução são os métodos iterativos baseados no gradiente, como *Steepest Descent* [42, 40], Gradiente Conjugado [40, 70, 83] e alguns métodos do tipo Quasi-Newton [40]. Em virtude da grande dimensão do problema de otimização, os métodos baseados no uso de Hessiana, como Método de Newton [40], não são adequados devido ao alto custo computacional da inversão da Hessiana.

Encontrar o método de otimização mais apropriado para o problema, principalmente quando a discrepância e a penalidade não são quadráticas, ainda é um problema em aberto. Nesta tese, a implementação proposta foi feita utilizando o método do Gradiente Conjugado Linear e Não-Linear [40, 83], que estão descritos no Apêndice B.3.

O Gradiente Conjugado Linear é utilizado quando todas as penalidades são baseadas na norma ℓ_2 ao quadrado. Quando qualquer uma das normas for diferente da ℓ_2 ao quadrado, então o Gradiente

Conjugado Não-Linear deve ser considerado. É importante lembrar que o uso do Gradiente Conjugado Linear já foi proposto em [35] para super-resolução, contudo o uso do GC Não-Linear em super-resolução é uma inovação na área.

3.3 Comparações Teóricas

Nesta seção, as equações do método proposto são comparadas com as equações dos métodos existentes e, além disso, o desempenho teórico dos métodos é contrastado, quanto à qualidade da estimativa, quanto ao custo computacional e quanto à robustez aos *outliers*.

3.3.1 Comparações entre as Equações

Se o algoritmo proposto for comparado com os demais algoritmos apresentados, utilizando as mesmas medidas de discrepância e o mesmo potencial na penalidade de regularização, podem-se notar claramente as diferenças entre eles. Abaixo são apresentadas as equações:

Primeiramente, o algoritmo de super-resolução de única imagem:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \left[\frac{J(\mathbf{g}_k, \mathbf{D}_k \mathbf{f}_k)}{\theta_{\eta_k}} + \frac{\mathcal{V}(\mathbf{R}_k \mathbf{f}_k)}{\beta_R} \right] \quad (3.28)$$

O algoritmo de super-resolução tradicional:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \left[\sum_{j=1}^L \frac{J(\mathbf{g}_j, \mathbf{C}_{j,k} \mathbf{f}_k)}{\theta_{\epsilon_{j,k}}} + \frac{\mathcal{V}(\mathbf{R}_k \mathbf{f}_k)}{\beta_R} \right] \quad (3.29)$$

A versão generalizada do algoritmo simultâneo existente, usando o modelo de segunda ordem:

$$\begin{aligned} \hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} & \sum_{k=1}^L \sum_{j=1}^L \frac{J(\mathbf{g}_j, \mathbf{C}_{j,k} \mathbf{f}_k)}{\theta_{\epsilon_{j,k}}} + \sum_{k=1}^L \frac{\mathcal{V}(\mathbf{R}_k \mathbf{f}_k)}{\beta_R} \\ & + \sum_k \frac{\mathcal{V}(-\mathbf{M}_{k,k-1} \mathbf{f}_{k-1} + 2 \mathbf{f}_k - \mathbf{M}_{k,k+1} \mathbf{f}_{k+1})}{\beta_M} \end{aligned} \quad (3.30)$$

Na proposta original, em [1], o autor utiliza diferença de segunda ordem tanto para a suavização espacial quanto para a suavização ao longo da trajetória de movimento e $\beta_R = \beta_M$. Além disso, é considerado apenas norma ℓ_2 ao quadrado no termo de dados

O algoritmo proposto neste trabalho, usando o modelo de segunda ordem:

$$\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_L = \arg \min_{\mathbf{f}_1, \dots, \mathbf{f}_L} \sum_{k=1}^L \frac{J(\mathbf{g}_k, \mathbf{D}_k \mathbf{f}_k)}{\theta_{\eta_k}} + \sum_{k=1}^L \frac{\mathcal{V}(\mathbf{R}_k \mathbf{f}_k)}{\beta_R} + \sum_k \frac{\mathcal{V}(-\mathbf{M}_{k,k-1} \mathbf{f}_{k-1} + 2 \mathbf{f}_k - \mathbf{M}_{k,k+1} \mathbf{f}_{k+1})}{\beta_M} \quad (3.31)$$

Observe que, no termo de dados do método proposto, apenas as matrizes \mathbf{D}_k , do sistema de aquisição, são encontradas, assim como no método de única imagem. Os demais métodos utilizam o modelo combinado no termo de dados. Nos dois métodos simultâneos, no proposto e no existente [1], a matriz de movimento é encontrada no termo de regularização, que, nesse exemplo, é o modelo de segunda ordem apresentado em (3.25). A Tabela 3.1 mostra quais modelos são considerados em cada um dos métodos contrastados.

Tabela 3.1: Modelos utilizados nos algoritmos de super-resolução

Termo	Termo de Dados		Termo <i>A Priori</i>	
	Aquisição	Combinado	Imagem	Movimento
Única imagem	Sim	Não	Sim	Não
Tradicional	Sim	Sim	Sim	Não
Simultâneo em [1]	Sim	Sim	Sim	Sim
Simultâneo proposto	Sim	Não	Sim	Sim

Pode-se observar que, em uma versão generalizada de [1], o método simultâneo dado pela equação (3.30) é mais completo que o proposto em (3.31). Contudo, este trabalho é baseado na hipótese de que o ganho trazido pelo uso das equações do modelo combinado nos algoritmos simultâneos é muito pequeno em relação ao aumento de custo computacional resultante.

3.3.2 Qualidade Obtida pelos Métodos Simultâneos

O método simultâneo proposto nesta tese, assim como o método simultâneo existente, proposto em [1], utiliza o modelo de movimento no termo de informação *a priori*. Essa penalidade permite estimações de alta qualidade, quanto comparada com os demais métodos. A diferença entre as imagens com movimento é minimizada nos quadros de alta resolução, proporcionando uma solução mais consistente com o modelo de movimento apresentado na equação (2.3). As equações minimizadas são do tipo:

$$\|\mathbf{f}_j - \mathbf{M}_{j,k} \mathbf{f}_k\| = \|\mathbf{e}_{j,k}\| \quad (3.32)$$

sendo que o erro de movimento, $\mathbf{e}_{j,k}$, é minimizado em alta resolução.

Entretanto, nos métodos tradicionais, que utilizam o movimento apenas no modelo de aquisição e movimento combinados, minimizam o erro de movimento indiretamente e em baixa resolução, juntamente com o ruído de aquisição, segundo:

$$\begin{aligned}
 \|\mathbf{g}_k - \mathbf{C}_{k,j}\mathbf{f}_j\| &= \|\boldsymbol{\epsilon}_{k,j}\| \\
 \|\mathbf{g}_k - \mathbf{D}_k\mathbf{M}_{k,j}\mathbf{f}_j\| &= \|\boldsymbol{\epsilon}_{k,j}\| \\
 \|\mathbf{D}_k\mathbf{f}_k + \boldsymbol{\eta}_k - \mathbf{D}_k\mathbf{M}_{k,j}\mathbf{f}_j\| &= \|\boldsymbol{\epsilon}_{k,j} + \boldsymbol{\eta}_k\| \\
 \|\mathbf{D}_k(\mathbf{f}_k - \mathbf{M}_{k,j}\mathbf{f}_j) + \boldsymbol{\eta}_k\| &= \|\mathbf{D}_k\mathbf{e}_{k,j} + \boldsymbol{\eta}_k\|
 \end{aligned} \tag{3.33}$$

O operador \mathbf{D}_k , que representa um mapeamento de HR para LR, e inclui distorção óptica e subamostragem. Este operador gera perdas na diferença na trajetória do movimento que está sendo minimizada. Assim, diferente dos métodos simultâneos, os métodos tradicionais nem sempre oferecem uma solução completamente consistente com o modelo de movimento apresentado em (2.3).

Outra vantagem é o comportamento recursivo do método simultâneo. Este comportamento permite que os detalhes recuperados em um quadro em particular se propaguem para os demais quadros, contribuindo para aumentar a similaridade na trajetória de movimento e obter um melhoramento de qualidade uniforme ao longo da seqüência.

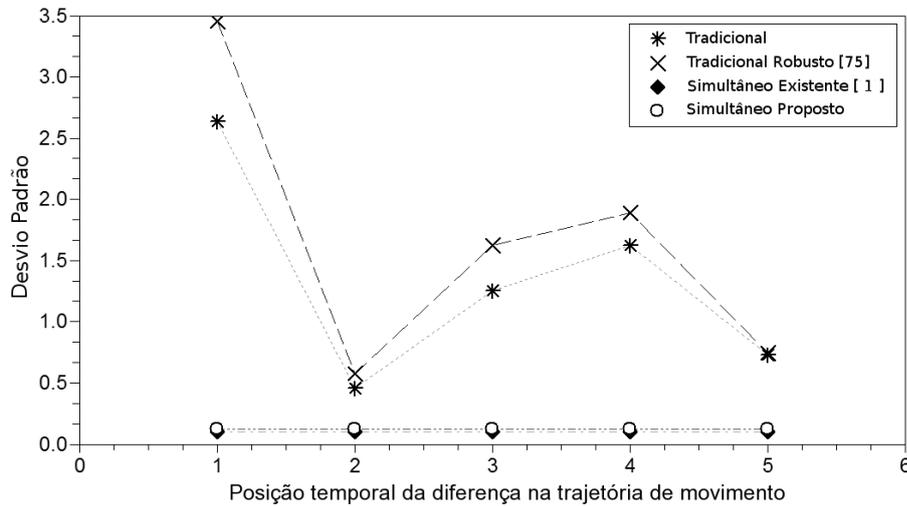


Figura 3.1: Desvio-padrão da diferença entre quadros adjacentes ao longo da seqüência.

A Figura 3.1 mostra o desvio-padrão da diferença na trajetória de movimento entre dois quadros adjacentes da seqüência recuperada com determinado método SR. Este exemplo mostra os resultados para a seqüência *Árvore*, que é faz parte dos experimentos do Capítulo 4. O desvio-padrão pequeno nas

seqüências recuperadas pelos métodos simultâneos indica que as imagens resultantes são mais similares na trajetória de movimento do que nas reconstruídas pelos métodos tradicionais.

3.3.3 Redução de Custo Computacional

Uma das principais contribuições deste trabalho é a redução computacional obtida em relação ao método simultâneo proposto em [1]. A redução de custo computacional é alcançada através da redução do número de termos na função de custo e pela otimização eficiente baseada nos métodos de Gradiente Conjugado.

A redução de complexidade computacional do método simultâneo proposto em relação ao método simultâneo existente é proveniente da diminuição do custo computacional dos produtos matriz-vetor, necessários no método do Gradiente Conjugado; que é bem menor no método proposto do que no método simultâneo existente. Em alguns casos, o custo por iteração pode chegar a ser menor do que nos métodos tradicionais.

A Tabela 3.2 apresenta o custo computacional aproximado por iteração requerido pelo GC linear, para os métodos apresentados. Nesta tabela é apresentado o número de operações, sendo que o custo de uma multiplicação é assumido igual ao de uma adição. A Tabela 3.2 considera a implementação do produto matriz-vetor, no Gradiente Conjugado, na forma expandida¹, ou seja, $\mathbf{A}\mathbf{f} = \mathbf{D}^T\mathbf{D}\mathbf{f} + \lambda_R\mathbf{R}^T\mathbf{R}\mathbf{f} + \lambda_M\mathbf{M}^T\mathbf{M}\mathbf{f}$. Observe que, nos métodos simultâneos uma iteração atualiza as L imagens da seqüência, enquanto que no método de única imagem e no método tradicional cada iteração atualiza uma imagem da seqüência, requerendo L iterações para atualizar as L imagens da seqüência.

Tabela 3.2: Custo por iteração do GC linear em operações (adição ou multiplicação)

Algoritmo SR	Custo por iteração para atualizar L imagens
Única Imagem	$L(10M+4(DN+BP))$
Tradicional	$L(10M+4(DLN+BP))$
Simult. de [1]	$10LM+4(DL^2N+BLP+QLM)$
Simult. Proposta	$10LM+4(DLN+BLP+QLM)$

Sendo L o número de imagens na seqüência, M o número de pixels na imagem HR, N o número de pixels na imagem LR, P a dimensão do vetor resultante de $\mathbf{R}_k\mathbf{f}_k$, B o número de elementos não-nulos por linha em \mathbf{R}_k , D o número de elementos não-nulos por linha em \mathbf{D}_k ou $\mathbf{C}_{k,j}$, e finalmente, Q o número de elementos não-nulos por linha em \mathbf{M} .

Na Tabela 3.3 é apresentado um exemplo com valores tipicamente encontrados nos problemas de

¹Aplicando uma série de produtos matriz-vetor, seguido da soma dos vetores resultantes, sem realizar produtos matriz-matriz

super-resolução, os quais correspondem aos valores encontrados nos experimentos realizados no Capítulo 4. Observe que, o custo computacional por iteração é proporcional a L^2 no método tradicional e no método simultâneo proposto em [1]. Por outro lado, no método de única imagem e no método simultâneo proposto o custo por iteração é proporcional a L .

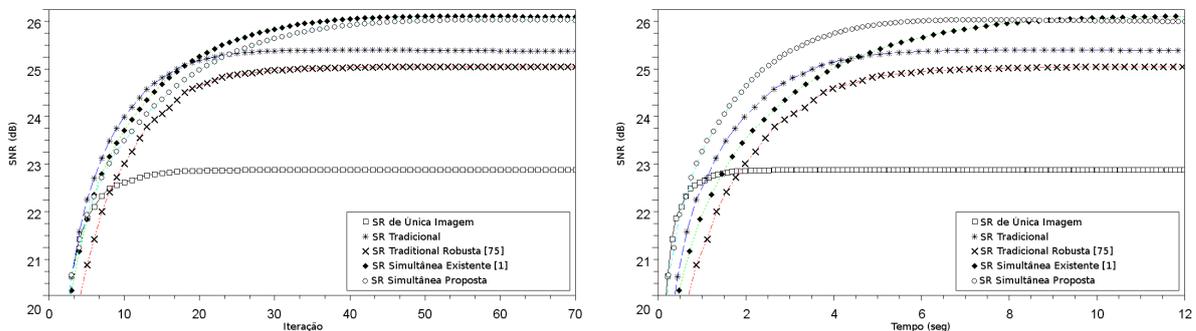
Tabela 3.3: Exemplo de custo por iteração do GC com valores típicos encontrados em SR

Algoritmo SR	Custo por iteração neste exemplo
Única Imagem	$(40 + 64 + 128)LN$
Tradicional	$(40 + 64L + 128)LN$
Simultânea de [1]	$(40 + 64L + 128 + 32)LN$
Simultânea Proposta	$(40 + 64 + 128 + 32)LN$

Sendo que os valores típicos de SR: $D = 16, B = 2, P = 4M, M = 4N, Q = 2$.

O número de iterações necessárias para a convergência é ligeiramente diferente entre os métodos contrastados. Em geral, o método de única imagem converge em menos iterações; os métodos tradicionais requerem mais iterações do que os de única imagem; enquanto os métodos simultâneos requerem ainda mais iterações do que os tradicionais.

Nos algoritmos com penalidade Huber o processo de otimização é mais complexo. Contudo, a utilização do método de Gradiente Conjugado Não-Linear proporciona um desempenho superior ao método do Gradiente Descendente, proposto em [1]. O Gradiente Conjugado Não-Linear possui uma velocidade de convergência maior do que a do método de Gradiente Descendente [83]. Além disso, a redução do custo por iteração do método proposto em relação ao método simultâneo existente também é grande no GC Não-Linear, dando mais uma vantagem ao método simultâneo proposto.



(a) Evolução do ganho em qualidade, através da SNR (dB), ao longo das iterações (b) Evolução do ganho em qualidade, através da SNR (dB), ao longo do tempo, em segundos

Figura 3.2: Observação da redução de custo computacional do método proposto.

A Figura 3.2 ilustra o desempenho obtido com a redução de custo computacional por iteração do método proposto em relação aos demais métodos. Esta figura ilustra a evolução do ganho em qualidade,

através da evolução da relação sinal ruído (SNR - *signal-to-noise-ratio*) ao longo das iterações e ao longo do tempo, utilizado o Gradiente Conjugado Não-Linear. Neste exemplo, pode-se notar que o método proposto requer algumas iterações a mais do que o método simultâneo existente (aproximadamente 10% a mais, neste exemplo). Contudo, o custo por iteração é bem menor que o método simultâneo existente (aproximadamente 50% menor, neste exemplo) e, como resultado, o método proposto converge mais rapidamente, em termos de tempo total, do que o método existente.

Além da redução de custo computacional do algoritmo em si, o método proposto requer uma quantidade menor de movimentos a serem estimados, o que reduz o custo computacional global do método proposto quando o movimento não é conhecido. Esta redução ocorre porque os modelos de movimento adotados são, geralmente, de baixa ordem, envolvendo apenas o movimento entre quadros próximos. Por exemplo, no caso dos modelos de movimento de primeira ordem, apenas $L - 1$ conjuntos de vetores de movimentos precisam ser estimados para uso no método proposto; já no método simultâneo existente, $L^2 - L$ conjuntos de vetores precisam ser estimados, um conjunto para cada matriz $\mathbf{C}_{j,k}$ em que $j \neq k$, aumentando o custo computacional global do método.

3.3.4 Aumento da Robustez aos *Outliers*

A existência de *outliers* no movimento entre as imagens é bastante comum em situações práticas. Os *outliers* causam distorções que reduzem a qualidade da estimativa. Portanto, é necessário o uso de algoritmos robustos ou de procedimentos auxiliares de detecção e remoção dos *outliers*.

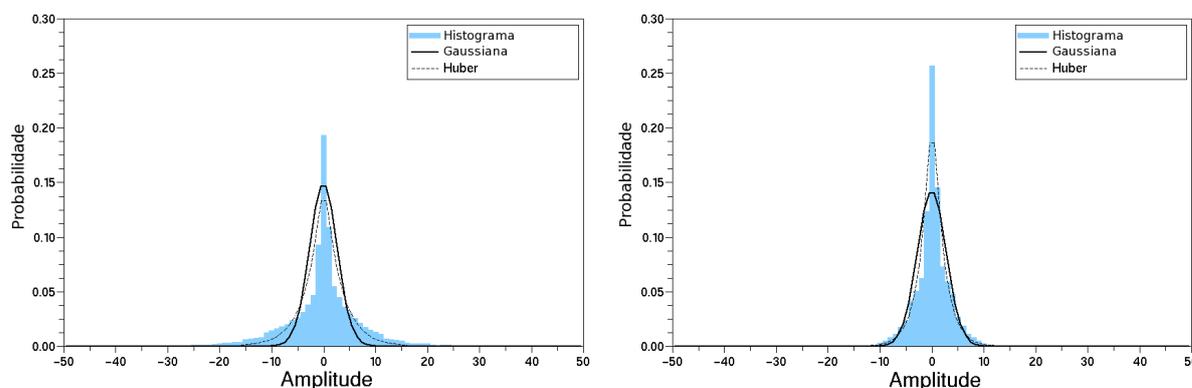
Entre os algoritmos revisados neste trabalho, o algoritmo de única imagem é o único imune aos *outliers* no erro de movimento, porque as equações de movimento não são utilizadas nesse método; todos os demais estão sujeitos aos problemas causados pelos *outliers*.

A utilização da norma ℓ_2 ao quadrado nas equações de erro de movimento, tanto no termo de dados quanto no termo *a priori*, reduz a robustez do método. Isto acontece porque os grandes erros, ou *outliers*, têm grande influência na estimativa, causando grandes desvios na solução.

De acordo com a teoria de estimação robusta em [44], a mudança da norma associada ao termo que contém *outliers* pode aumentar significativamente a robustez do algoritmo de estimação. Assim, a vantagem do método proposto está na robustez que é alcançada quando a norma Huber ou a ℓ_1 são utilizadas no termo *a priori*. Com essas normas, os erros de grande magnitude têm influência consideravelmente menor na estimação do que com a norma ℓ_2 ao quadrado.

A escolha da norma ℓ_2 ao quadrado assume, implicitamente, que o erro sendo minimizado segue uma

distribuição Gaussiana. Entretanto, é conhecido [44] que os *outliers* podem desviar o comportamento estatístico do erro do modelo Gaussiano assumido. A Figura 3.3 ilustra este problema, mostrando o histograma do erro de movimento com *outliers* e com *outliers* removidos.



(a) Histograma do erro de movimento de uma seqüência com *outliers*. (b) Histograma do mesmo erro, com *outliers* detectados e removidos.

Figura 3.3: Histograma do erro de movimento de uma seqüência com a presença de *outliers* e com os *outliers* removidos.

Pode-se observar, na Figura 3.3, que os *outliers* contribuem para estender a “cauda” do histograma. Observe que a densidade gaussiana não aproxima bem o histograma com a presença de *outliers*, mas representa consideravelmente melhor o histograma com os *outliers* removidos. Entretanto, uma densidade baseada na norma Huber proporciona uma boa aproximação do histograma tanto com a presença de *outliers* quanto com os *outliers* removidos, especialmente na região da “cauda” do histograma.

Os métodos tradicionais e o método simultâneo existente em [1] utilizam o movimento no termo de dados, associado com a norma ℓ_2 ao quadrado. Por causa disso, estes métodos ainda são suscetíveis às distorções causadas pelos *outliers*. A robustez alcançada pelo método proposto usando a norma Huber no termo *a priori* é similar à robustez alcançada pelo método tradicional robusto proposto em [75].

Conclusão do Capítulo

Neste capítulo foi apresentada a primeira contribuição desta tese: o método de super-resolução simultânea. No início do capítulo, foram apresentados os problemas encontrados nos algoritmos simultâneos existentes. Em seguida, a idéia principal e o método proposto foram apresentados. Logo após, as comparações teóricas entre o método proposto e os métodos existentes são mostradas, as quais explicam como o método proposto atua para reduzir os problemas identificados no método simultâneo existente.

Capítulo 4

Experimentos Comparativos dos Métodos Regularizados de Super-Resolução

O objetivo deste capítulo é mostrar os resultados experimentais obtidos pelo método desenvolvido. Estes experimentos ilustram a melhoria obtida em relação ao método simultâneo existente. Os resultados mais significativos são: comprovação da qualidade de imagens superior dos métodos simultâneos, redução do custo computacional e aumento da robustez aos *outliers* no método proposto. O algoritmo proposto foi comparado com os algoritmos SR de única imagem [84], com os algoritmos SR tradicionais [2], com o algoritmo SR tradicional robusto [75] e com o algoritmo simultâneo existente [1].

Na seção 4.1, os experimentos realizados são descritos. São mostradas as características das simulações e as formas de comparação consideradas. Também são apresentados os algoritmos implementados, onde foram utilizadas as normas ℓ_2 ao quadrado, Huber e, em alguns casos, ℓ_1 .

Na seção 4.2, são apresentados os resultados dos experimentos. Estes resultados comparativos buscam ilustrar três pontos: a qualidade obtida pelos métodos simultâneos, a redução de custo computacional e a maior robustez do método proposto em comparação ao método simultâneo existente.

4.1 Descrição do Experimento

4.1.1 Procedimento

Para mostrar o desempenho dos algoritmos, o seguinte experimento foi realizado:

- Considerou-se uma seqüência de imagens HR conhecida, entre as apresentadas na seção 4.1.3, cujo movimento é conhecido ou foi apropriadamente estimado, conforme descrito na seção 4.1.3.

- Simulou-se o processo de aquisição das imagens da seqüência. Considerou-se um fator de redução de resolução $R = 2$. A matriz \mathbf{D}_k utilizada possui uma estrutura invariante¹, sendo que a função de aquisição para cada pixel na imagem LR, na forma bidimensional, é a função $d_k[m_1, m_2]$, mostrada na Figura 4.1. Além disso, as matrizes \mathbf{D}_k são iguais para todas as imagens da seqüência. O ruído de aquisição $\boldsymbol{\eta}_k$ utilizado foi um ruído gaussiano branco. Considerou-se uma relação de sinal ruído (SNR - *signal-to-noise-ratio*) de 40 dB, tipicamente encontrada em sensores de imagem de boa qualidade [8, 85]. A variância do ruído de aquisição foi determinada de forma a se manter a SNR² de aquisição desejada.

$$d_k[m_1, m_2] = \frac{1}{64} \begin{bmatrix} 1 & 3 & 3 & 1 \\ 3 & 9 & 9 & 3 \\ 3 & 9 & 9 & 3 \\ 1 & 3 & 3 & 1 \end{bmatrix}$$

Figura 4.1: Função discreta de aquisição utilizada.

- Foram utilizados algoritmos descritos na seção 4.1.2 para aumentar a resolução da seqüência capturada.
- Os resultados obtidos foram comparados quanto à qualidade e ao custo computacional. Para a comparação da qualidade foi utilizada a medida de SNR dos resultados [4, 42]. A SNR, expressa em decibéis (dB), é definida como:

$$SNR = 10 \log_{10} \frac{\|\mathbf{f}\|_2^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|_2^2} \quad (4.1)$$

Para comparação do custo computacional, foi utilizado o tempo de processamento até a convergência do método. Assume-se que a convergência foi alcançada quando o melhoramento de qualidade, em SNR (dB), da iteração é menor que 10^{-2} .

4.1.2 Algoritmos Implementados

Todos algoritmos de super-resolução foram implementados através do método do Gradiente Conjugado (GC), usando GC linear para minimização quando o potencial for quadrático e GC não-linear quando o potencial for Huber ou ℓ_1 . Os métodos de GC estão descritos no Apêndice B.3.

¹Ver descrição no Apêndice A.1

²A SNR de aquisição é definida como $SNR = 10 \log_{10}(\sigma_{\mathbf{D}\mathbf{f}}^2/\sigma_{\boldsymbol{\eta}}^2)$, no qual $\sigma_{\mathbf{D}\mathbf{f}}^2$ é a variância da seqüência LR sem ruído e $\sigma_{\boldsymbol{\eta}}^2$ é a variância do ruído de aquisição.

Algoritmos de Interpolação (INT)

Para ilustrar a melhoria de qualidade obtida pelos métodos de super-resolução, serão comparados os resultados obtidos com dois métodos de interpolação [42, 18, 19]. Foram consideradas a interpolação de ordem zero (INT-0) e a interpolação linear (INT-1). O procedimento de interpolação não é considerado um procedimento de super-resolução, porque não corrige as distorções causadas no processo de aquisição. Contudo, este método é muito utilizado para aumentar o número de pixels de uma imagem.

Algoritmos SR de Única Imagem (SR-UIM)

Os algoritmos SR de única imagem [84, 86] são descritos pela equação seguinte. Para $k = 1 \dots L$:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2^2 + \lambda_R \mathcal{V}(\mathbf{R}_k \mathbf{f}_k) \quad (4.2)$$

sendo que os coeficientes λ_R são:

$$\lambda_R = \sigma_{\boldsymbol{\eta}_k}^2 / \beta_R \quad (4.3)$$

na qual $\sigma_{\boldsymbol{\eta}_k}^2$ é a variância do ruído de aquisição, que é conhecida nestes experimentos, e β_R é um hiperparâmetro do modelo da imagem, o qual é estimada utilizando a imagem original para estes experimentos. As normas $\mathcal{V}(\mathbf{x})$ utilizadas são descritas mais a frente.

A principal característica dos métodos de única imagem é que o movimento não é utilizado. O método de única imagem é capaz de reduzir grande parte das distorções da aquisição, produzindo resultados muito melhores que os métodos de interpolação.

Algoritmos SR Tradicionais (SR-TRA)

Os algoritmos tradicionais [2] implementados foram os seguintes, sendo $k = 1 \dots L$:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \alpha_{k,j} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_R \mathcal{V}(\mathbf{R}_k \mathbf{f}_k) \quad (4.4)$$

sendo que os coeficientes $\alpha_{k,j}$ e λ_R são calculados a partir de hiperparâmetros, que são conhecidos ou estimados das seqüências originais para estes experimentos, em que:

$$\alpha_{k,j} = \sigma_{\boldsymbol{\eta}_k}^2 / (\sigma_{\boldsymbol{\eta}_j}^2 + \sigma_{\boldsymbol{\epsilon}_{j,k}}^2) \quad (4.5)$$

no qual $\sigma_{\boldsymbol{\eta}_k}^2$ é a variância do ruído de aquisição e $\sigma_{\boldsymbol{\epsilon}_{j,k}}^2$ é a variância do erro combinado. Além disso:

$$\lambda_R = \sigma_{\boldsymbol{\eta}_k}^2 / \beta_R \quad (4.6)$$

sendo β_R um hiperparâmetro do modelo da imagem.

A principal característica dos métodos tradicionais é o fato do movimento ser utilizado somente no termo de dados, através do modelo combinado de aquisição e de movimento. No termo de informação *a priori* apenas a suavização espacial é utilizada.

Algoritmo SR Tradicional Robusto (SR-FAR)

O algoritmo tradicional robusto a *outliers*, apresentado em [75], implementado foi o seguinte, sendo $k = 1 \dots L$:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_1 + \lambda_R \|\mathbf{R}_k \mathbf{f}_k\|_1 \quad (4.7)$$

no qual o coeficiente λ_R é:

$$\lambda_R = \sigma_{\boldsymbol{\eta}_k}^2 / \beta_R \quad (4.8)$$

na qual $\sigma_{\boldsymbol{\eta}_k}^2$ é a variância do ruído de aquisição e β_R é um hiperparâmetro do modelo da imagem. Estes hiperparâmetros são conhecidos ou estimados das imagens originais nestes experimentos. Apenas norma ℓ_1 é utilizada nesse caso. Esse método é comparado apenas com os métodos utilizando a norma Huber, que têm a mesma finalidade.

Esse algoritmo segue o mesmo princípio dos métodos tradicionais, sendo que o movimento é utilizado somente no termo de dados, através do modelo combinado de aquisição e de movimento. No termo de informação *a priori* apenas a suavização espacial é utilizada.

Algoritmo Simultâneo Existente (SR-BOR)

Conforme a revisão da seção 2.2.5, o algoritmo proposto por Borman em [1] utiliza o movimento no termo de dados e, também, no termo de informação *a priori*. No termo de dados é utilizado o potencial quadrático, ao passo que no termo de informação *a priori* é utilizado o potencial Huber. Além disso, é considerada diferença finita de segunda ordem tanto para impor suavidade espacial quanto para impor suavidade na trajetória de movimento.

Todavia, neste trabalho foram implementadas versões generalizadas do Borman, além da versão ori-

ginal. A versão generalizada implementada é:

$$\hat{\mathbf{f}} = [\hat{\mathbf{f}}_1^T \dots \hat{\mathbf{f}}_L^T]^T = \arg \min_{\mathbf{f}} \sum_{k=1}^L \sum_{j=1}^L \alpha_{k,j} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_R \mathcal{V}(\mathbf{R}\mathbf{f}) + \lambda_M \mathcal{V}(\mathbf{M}\mathbf{f}) \quad (4.9)$$

sendo $\alpha_{k,j}$ e λ_R são os mesmos coeficientes usados nos algoritmos tradicionais e $\lambda_M = \sigma_{\boldsymbol{\eta}_k}^2 / \beta_M$. O hiperparâmetro β_M também é estimado a partir das imagens originais. No trabalho original do Borman apenas um coeficiente é utilizado para o termo de informação *a priori*. Mas, nas simulações deste capítulo, foram considerados dois coeficientes para uma comparação equilibrada.

Neste método toda a seqüência precisa ser estimada de forma simultânea. Isto ocorre porque o termo $\mathcal{V}(\mathbf{M}\mathbf{f})$, que especifica a similaridade ao longo da trajetória de movimento, não pode ser separado.

Algoritmos Simultâneos Propostos (SR-SIM)

Foi implementado o seguinte algoritmo:

$$\hat{\mathbf{f}} = [\hat{\mathbf{f}}_1^T \dots \hat{\mathbf{f}}_L^T]^T = \arg \min_{\mathbf{f}} \sum_{k=1}^L \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2^2 + \lambda_R \mathcal{V}(\mathbf{R}\mathbf{f}) + \lambda_M \mathcal{V}(\mathbf{M}\mathbf{f}) \quad (4.10)$$

Observe que apenas as equações do modelo de aquisição são utilizadas no termo de dados. O movimento é considerado apenas no termo $\mathcal{V}(\mathbf{M}\mathbf{f})$.

Particularidades das Implementações

As matrizes envolvidas foram armazenadas através de métodos de esparsidade³. A matriz \mathbf{R}_k , utilizada no termo de suavidade espacial, é a mesma para todas as imagens e a matriz \mathbf{R} é a mesma em todos algoritmos. A matriz \mathbf{M} , do termo de suavidade ao longo da trajetória de movimento, é a mesma para os algoritmos SR-BOR e SR-SIM. Considerou-se dois casos para as matrizes \mathbf{R} e \mathbf{M} :

- Modelo de primeira ordem, sendo que \mathbf{R} representa as diferenças finitas espaciais de primeira ordem, nas direções horizontal, vertical e diagonais, conforme o modelo da equação (A.21) definido no Apêndice A.3, e \mathbf{M} representa a diferença finita de primeira ordem ao longo da trajetória de movimento. Neste caso, o algoritmo é representado pela sigla SR-XXX-1;
- Modelo de segunda ordem, sendo que \mathbf{R} representa as diferenças finitas espaciais de segunda ordem, nas direções horizontal, vertical e diagonais, conforme o modelo da equação (A.22) definido

³Apenas os elementos não-nulos e suas posições na matriz são armazenados [43]

no Apêndice A.3, e \mathbf{M} representa a diferença finita de segunda ordem ao longo da trajetória de movimento. Neste caso, o algoritmo é representado pela sigla SR-XXX-2.

As matrizes \mathbf{R} de diferenças finitas espaciais de primeira e segunda ordens estão descritas no Apêndice A.3. As matrizes \mathbf{M} de diferenças finitas de primeira e segunda ordens na trajetória do movimento estão descritas no Capítulo 3.2 e no Apêndice B.2.

As normas $\mathcal{V}(\mathbf{x})$ consideradas foram: a norma ℓ_2 ao quadrado, ou $\|\mathbf{x}\|^2$ e a norma Huber [53], ou $\mathcal{H}_T(\mathbf{x})$, definida como:

$$\mathcal{H}_T(\mathbf{x}) = \sum_i H_T(x_i) \quad (4.11)$$

no qual:

$$H_T(x) = \begin{cases} |x|^2 & \text{se } |x| \leq T \\ 2T|x| - T^2 & \text{se } |x| > T \end{cases} \quad (4.12)$$

sendo que T é o parâmetro que determina o limiar entre a região de ponderação quadrática e a região de ponderação linear. Nestes experimentos foi utilizado $T = 1$, para aproveitar a região linear da penalidade, lembrando que a intensidade das imagens digitais utilizadas neste experimento são armazenadas com valores inteiros entre 0 e 255 (representação de 8 bits).

Os coeficientes de regularização utilizados são baseados da escolha da estimativa MAP. A variância do ruído é conhecida e os hiperparâmetros relacionados à derivada das imagens, β_R , e ao erro de movimento, β_M , são calculados a partir das imagens originais, para fins deste experimento. Portanto, está-se assumindo que os coeficientes são conhecidos. O cálculo dos hiperparâmetros β_R é feito utilizando-se máxima verosimilhança. O cálculo dos hiperparâmetros β_M é feito utilizando-se estimação robusta apresentada em [44]. Esta estimativa entrega hiperparâmetros mais confiáveis quando os dados contêm *outliers*, como é o caso. Isso garante que o hiperparâmetro de movimento será aproximadamente o mesmo valor se o erro de movimento for estimado a partir dos dados antes da detecção dos *outliers* ou após a detecção dos *outliers*.

A detecção de *outliers* foi feita segundo [46], utilizando os erros $\epsilon_{j,k} = \mathbf{g}_j - \mathbf{C}_{j,k}\mathbf{f}_k^*$, para os *outliers* no termo de dados, e $e_{j,k} = \mathbf{f}_j^* - \mathbf{M}_{j,k}\mathbf{f}_k^*$, para os *outliers* no termo *a priori*, sendo \mathbf{f}_k^* uma imagem produzida por interpolação linear, a partir de \mathbf{g}_k . Valores absolutos de amplitude maiores que 8 vezes o desvio-padrão estimado são considerados *outliers* e excluídos da estimativa.

4.1.3 Seqüências de Imagens Utilizadas

Seqüências com Movimento Conhecido e sem *Outliers*

Barco: Esta seqüência, especificada na Figura 4.2, possui 4 imagens de tamanho 84×72 , com movimento translacional global pequeno. Os *outliers* ocorrem apenas nas bordas e podem ser considerados desprezíveis.

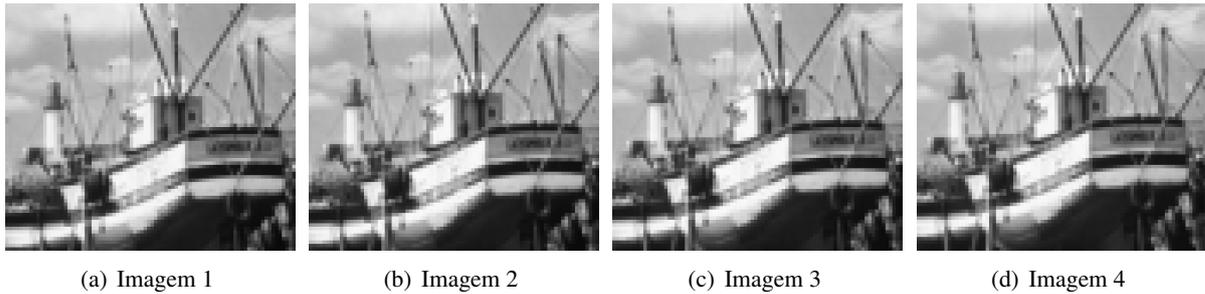


Figura 4.2: Seqüência de Imagens *Barco*.

Árvore: Esta seqüência, especificada na Figura 4.3, possui 6 imagens de tamanho 96×96 , com movimento translacional global pequeno. Os *outliers* ocorrem apenas nas bordas e podem ser considerados desprezíveis.

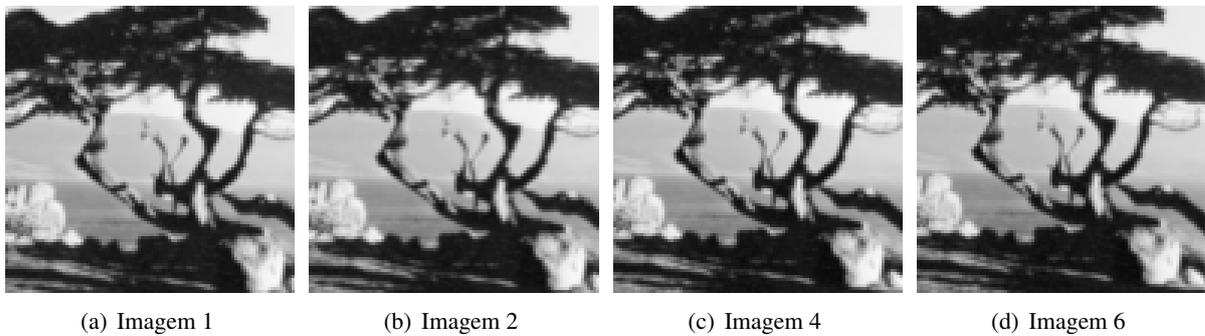


Figura 4.3: Seqüência de Imagens *Árvore*.

Seqüências com Movimento Conhecido e com *Outliers*

Sala: Esta seqüência, especificada na Figura 4.4, possui 4 imagens de tamanho 120×90 , com movimento translacional global pequeno. Um círculo preto, inserido artificialmente, que possui um movimento diferente do global, simula os *outliers* na seqüência.

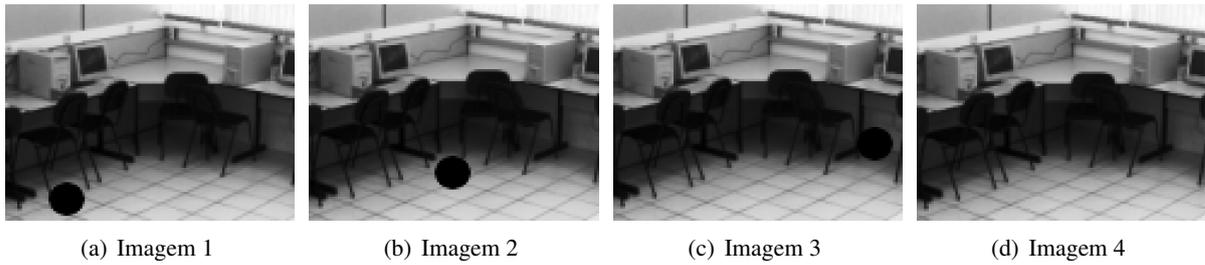


Figura 4.4: Seqüência de Imagens *Sala*.

Estrada: Esta seqüência, especificada na Figura 4.5, possui 4 imagens de tamanho 120×100 , com movimento translacional global grande. Um carro que se move com movimento diferente do movimento global da cena é considerado *outlier*.



Figura 4.5: Seqüência de Imagens *Estrada*.

Seqüências com Movimento Estimado

Jato: Esta seqüência, especificada na Figura 4.6, possui 8 imagens de tamanho 140×120 , com movimento estimado pelo algoritmo [87] de fluxo óptico. Os *outliers* ocorrem, principalmente, nas regiões cobertas e descobertas devido ao movimento da aeronave. Esta seqüência possui uma pequena mudança de intensidade ao longo do movimento em regiões consideradas *inliers*, como nas asas do jato.

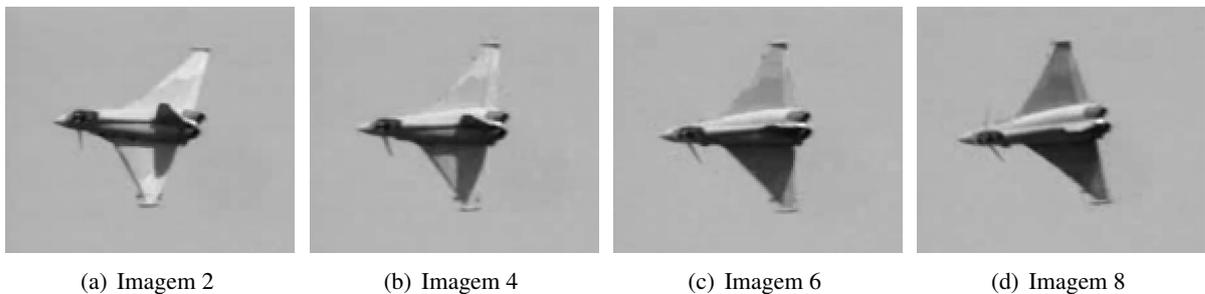


Figura 4.6: Seqüência de Imagens *Jato*.

Jardim: Esta seqüência, especificada na Figura 4.7, possui 6 imagens de tamanho 200×140 , com movimento estimado pelo algoritmo [87] de fluxo óptico. Os *outliers* ocorrem em diversos pontos da cena, mas a concentração principal está nos arredores da árvore, que encobre partes do fundo por se mover mais rapidamente que o fundo.

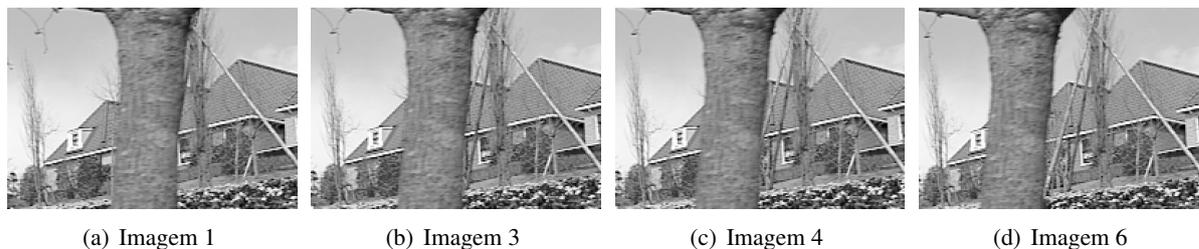


Figura 4.7: Seqüência de Imagens *Jardim*

4.2 Resultados Obtidos

4.2.1 Qualidade e Robustez das Estimativas

A qualidade da estimativa e a robustez dos métodos de super-resolução podem ser observadas nas tabelas de 4.1 a 4.4. As tabelas 4.1 e 4.2 mostram a qualidade obtida pelos métodos que usam a norma ℓ_2 ao quadrado, com a detecção de *outliers* ativada e desativada, respectivamente. As tabelas 4.3 e 4.4 mostram a qualidade obtida pelos métodos que usam a norma Huber ou ℓ_1 , também com a detecção de *outliers* ativada e desativada, respectivamente.

Os resultados numéricos da Tabela 4.1, dos métodos usando norma ℓ_2 ao quadrado e com detecção de *outliers* ativada, indicam que os métodos de super-resolução produzem imagens com mais qualidade que os métodos de interpolação. Além disso, os métodos de super-resolução tradicional, SR-TRA, entregam imagens com maior nitidez que os métodos SR de única imagem, SR-UIM. A Tabela 4.1, ainda indica que os métodos simultâneos, SR-SIM e SR-BOR, produzem imagens com praticamente a mesma qualidade, as quais são superiores às produzidas pelos demais métodos.

Comparando os resultados numéricos da Tabela 4.1 com os resultados da Tabela 4.2, é possível avaliar a pouca robustez dos métodos com norma ℓ_2 ao quadrado. Pode-se observar que todos os métodos que utilizam o movimento, como o SR-TRA, SR-BOR e SR-SIM, têm uma diminuição significativa na qualidade quando a seqüência de imagens contém *outliers*. Isso ilustra a baixa robustez destes algoritmos. A detecção é fundamental nesses casos. O método SR-UIM, que não utiliza o movimento, é o único imune aos problemas causados pelos *outliers*.

Tabela 4.1: Qualidade dos métodos SR com norma ℓ_2 ao quadrado e detecção de *outliers* **ativada**, em SNR (dB)

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média
INT-0	17.7	17.1	19.9	17.3	29.3	17.7	19.8
INT-1	17.8	17.4	20.0	17.2	29.9	17.4	19.9
SR-UIM-1	21.0	22.5	23.5	22.1	34.8	19.2	23.8
SR-UIM-2	21.1	22.6	23.6	22.2	34.8	19.2	23.9
SR-TRA-1	25.2	27.2	27.3	25.8	35.2	20.1	26.8
SR-TRA-2	25.1	27.3	27.1	25.7	35.1	20.1	26.7
SR-BOR-1	26.0	28.2	27.6	26.3	35.8	20.6	27.4
SR-BOR-2	25.9	28.1	27.4	26.3	35.5	20.6	27.3
SR-SIM-1	25.9	28.1	27.5	26.3	36.7	20.5	27.4
SR-SIM-2	25.9	28.1	27.4	26.3	36.7	20.6	27.4

 Tabela 4.2: Qualidade dos métodos SR com norma ℓ_2 ao quadrado e detecção de *outliers* **desativada**, em SNR (dB)

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média
INT-0	17.7	17.1	19.9	17.3	29.3	17.7	19.8
INT-1	17.8	17.4	20.0	17.2	29.9	17.4	19.9
SR-UIM-1	21.0	22.5	23.5	22.1	34.8	19.2	23.8
SR-UIM-2	21.1	22.6	23.6	22.2	34.8	19.2	23.9
SR-TRA-1	25.8	27.6	19.5	18.2	34.5	19.5	24.1
SR-TRA-2	26.3	27.6	20.3	19.8	34.7	19.6	24.7
SR-BOR-1	26.2	28.2	19.9	18.4	34.5	19.9	24.5
SR-BOR-2	26.6	28.2	20.4	20.4	34.6	19.9	25.0
SR-SIM-1	26.2	28.2	22.6	16.0	35.5	20.0	24.8
SR-SIM-2	26.6	28.2	22.2	21.5	35.9	19.9	25.7

As figuras 4.8 e 4.9 ilustram alguns resultados visuais dos métodos de super-resolução que usam norma ℓ_2 ao quadrado com a detecção dos *outliers* ativada. Nestas figuras pode-se observar que a nitidez das imagens produzidas pelos métodos de super-resolução é bem maior que a produzida pela interpolação. A melhora visual trazida pelos métodos SR-TRA em relação aos métodos SR-UIM é pequena, mas pode ser facilmente notada. Já a melhora visual trazida pelos métodos simultâneos, SR-BOR e SR-SIM, em relação aos métodos SR-TRA é pouco perceptível em uma imagem estática.

As figuras 4.10 e 4.11 ilustram o problema causado pelos *outliers*, mostrando alguns resultados visuais dos métodos de super-resolução que usam norma ℓ_2 ao quadrado com a detecção dos *outliers* desativada. Pode-se notar que as distorções aparecem em todos os métodos que utilizam o movimento, como o SR-TRA, SR-BOR e SR-SIM, mas não aparecem nos métodos que não utilizam o movimento,

como o SR-UIM e INT.

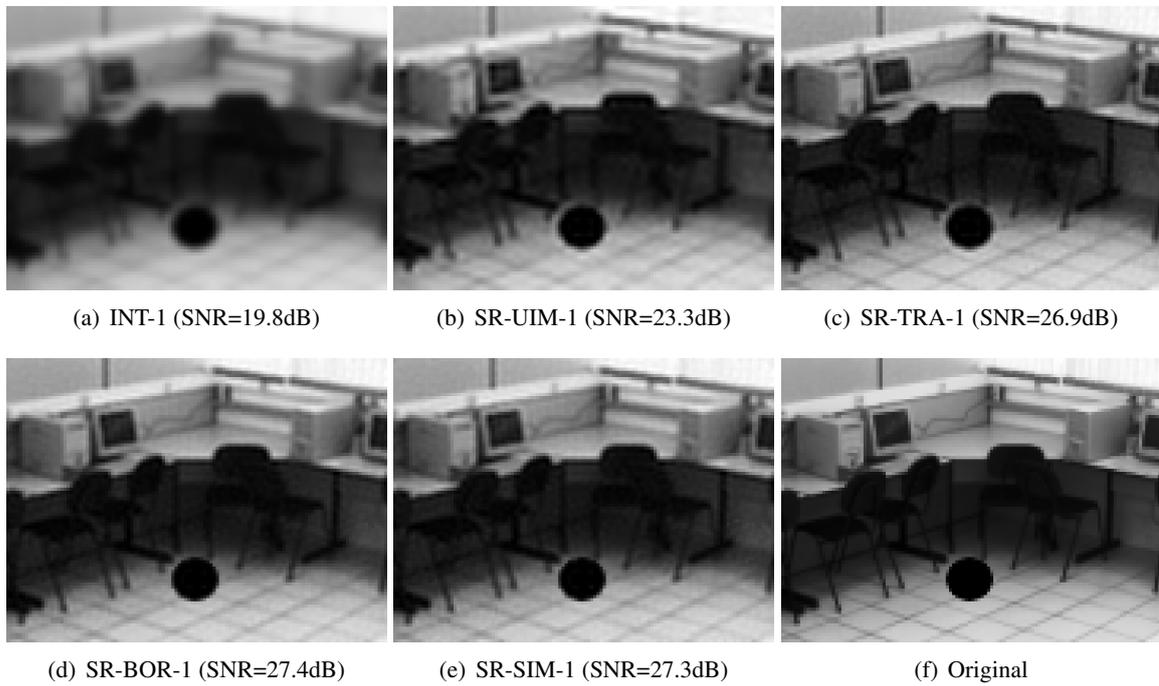


Figura 4.8: Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma ℓ_2 ao quadrado, modelos de primeira ordem e com detecção de *outliers* **ativada**.

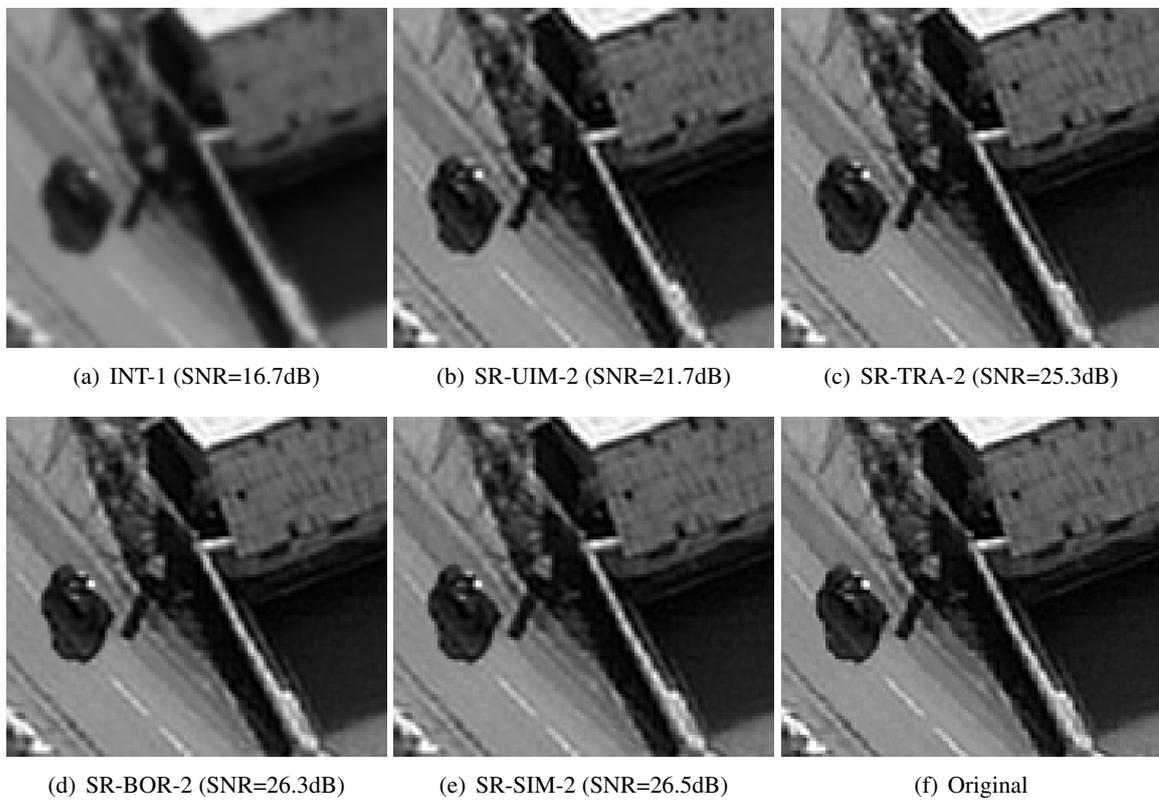


Figura 4.9: Exemplo visual mostrando a imagem 2 da seqüência Estrada, usando norma ℓ_2 ao quadrado, modelos de segunda ordem e com detecção de *outliers* **ativada**.

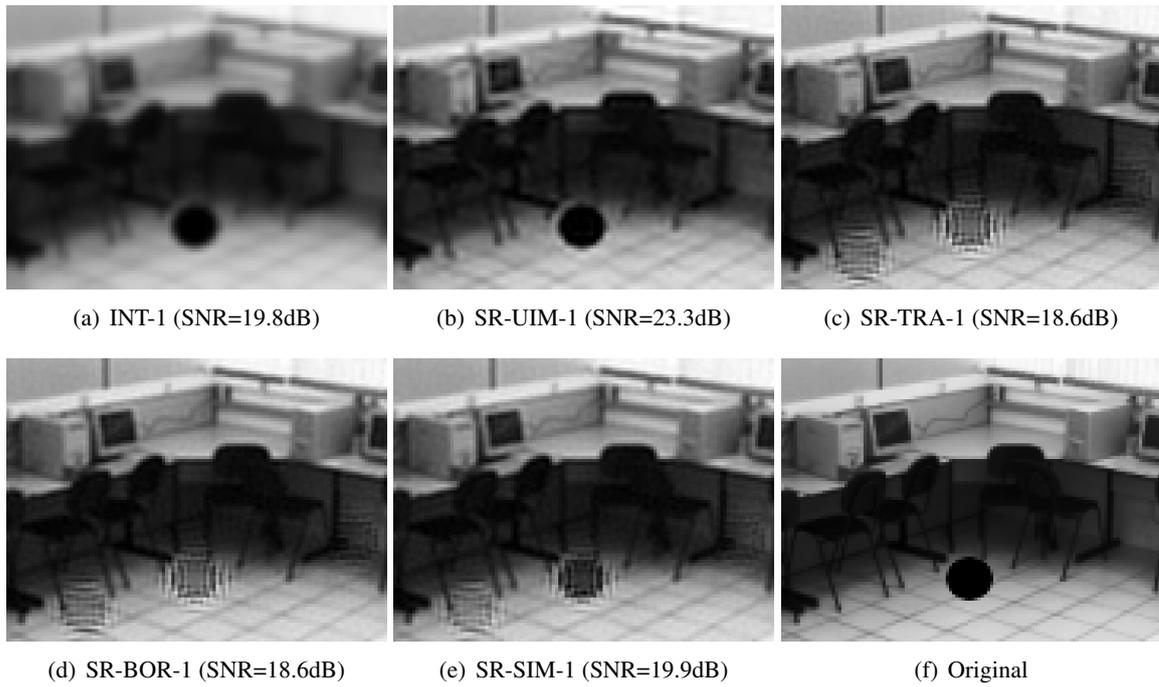


Figura 4.10: Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma ℓ_2 ao quadrado, modelos de primeira ordem e com detecção de *outliers* **desativada**.

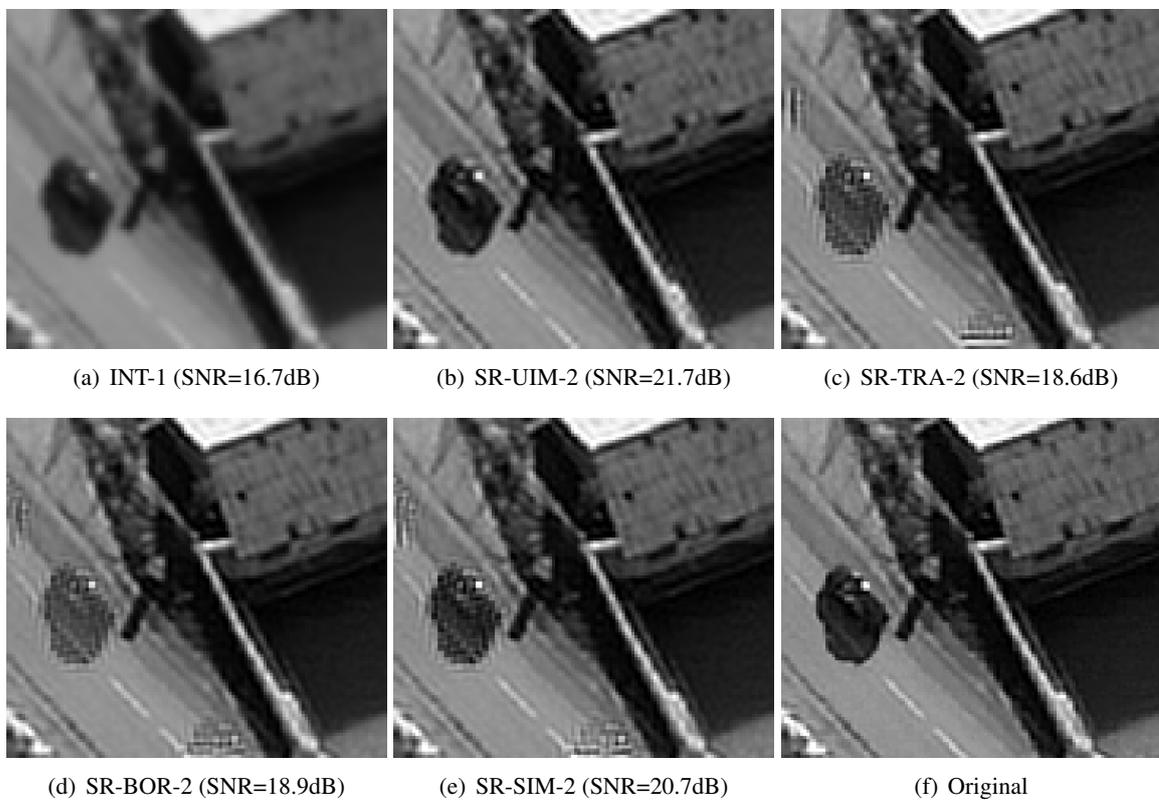


Figura 4.11: Exemplo visual mostrando a imagem 2 da seqüência Estrada, usando norma ℓ_2 ao quadrado, modelo de segunda ordem e com detecção de *outliers* **desativada**.

Os resultados numéricos da Tabela 4.3, dos métodos usando norma Huber ou ℓ_1 e com detecção de *outliers* ativada, mostram a qualidade destes métodos de super-resolução. Os métodos que usam estas normas buscam, inicialmente, uma melhora da qualidade através da produção de imagens com as bordas dos objetos bem nítidas. Os resultados da Tabela 4.3 indicam que os métodos simultâneos, SR-BOR e SR-SIM, são ligeiramente superiores aos métodos tradicionais, SR-TRA e SR-FAR; por sua vez, os métodos tradicionais são superiores ao método SR-UIM, que não explora o movimento.

Tabela 4.3: Qualidade dos métodos SR com norma Huber ou ℓ_1 e detecção de *outliers* **ativada**, em SNR (dB)

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média
SR-UIM-1	20.6	21.5	23.8	21.2	34.2	19.2	23.5
SR-UIM-2	20.8	22.8	23.9	22.7	36.1	19.4	24.1
SR-TRA-1	24.9	25.0	28.1	24.0	34.8	15.4	25.3
SR-TRA-2	25.5	27.1	28.0	25.7	37.0	16.6	26.6
SR-FAR-1	24.7	25.2	28.0	24.1	34.2	19.7	26.0
SR-FAR-2	25.2	26.5	27.4	25.7	35.8	20.0	26.7
SR-BOR-1	26.4	26.4	28.4	24.4	35.4	16.9	26.3
SR-BOR-2	26.5	28.0	28.4	26.1	37.4	18.1	27.4
SR-SIM-1	26.2	26.4	28.2	24.4	35.6	20.5	26.9
SR-SIM-2	26.5	28.0	28.3	26.2	37.7	20.5	27.8

Tabela 4.4: Qualidade dos métodos SR com norma Huber ou ℓ_1 e detecção de *outliers* **desativada**, em SNR (dB)

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média
SR-UIM-1	20.6	21.5	23.8	21.2	34.2	19.2	23.5
SR-UIM-2	20.8	22.8	23.9	22.7	36.1	19.4	24.1
SR-TRA-1	25.8	25.6	3.4	1.9	14.7	3.3	12.5
SR-TRA-2	26.3	27.4	2.0	6.0	17.0	1.9	13.4
SR-FAR-1	25.2	25.2	26.9	22.7	34.2	19.7	25.7
SR-FAR-2	25.3	26.5	26.2	24.1	35.8	19.8	26.3
SR-BOR-1	26.5	26.4	3.2	1.0	15.2	1.6	12.3
SR-BOR-2	26.7	28.0	9.0	1.9	17.6	0.2	13.9
SR-SIM-1	26.2	26.4	28.0	22.8	35.2	20.1	26.5
SR-SIM-2	26.6	28.0	28.1	25.5	37.2	19.9	27.5

A robustez dos métodos baseados na norma Huber ou ℓ_1 pode ser observada comparando-se os resultados Tabela 4.3, na qual a detecção de *outliers* foi ativada, com os resultados da Tabela 4.4, onde a detecção foi desativada. Pode-se notar que os métodos SR-FAR e o SR-SIM são mais robustos que os métodos SR-TRA e SR-BOR. Os métodos SR-TRA e SR-BOR tiveram uma grande perda de quali-

dade nas seqüências em que ocorrem *outliers* quando a detecção foi desativada. Os métodos SR-FAR e SR-SIM foram mais robustos, qualquer redução de qualidade é quase imperceptível visualmente.

A perda de qualidade devido aos *outliers* nos algoritmos não-robustos, que usam a norma Huber no operador espacial, é ainda maior. Isso acontece porque a norma ℓ_2 ao quadrado, no operador espacial, força uma suavidade maior nos resultados, evitando distorções extremamente exageradas. A norma Huber, que preserva as descontinuidades, associada ao operador de suavidade espacial, acaba permitindo distorções espaciais com uma amplitude maior, prejudicando mais os métodos não-robustos.

As figuras 4.12 e 4.13 ilustram alguns resultados visuais dos métodos de super-resolução que usam norma Huber ou ℓ_1 , com a detecção dos *outliers* ativada. Comparando a Figura 4.12 com a Figura 4.8, pode-se notar o aumento da nitidez nas bordas dos objetos com o uso da norma Huber ou ℓ_1 no operador de diferenças finitas espacial.

Os resultados das figuras 4.14 e 4.15 ilustram o problema das distorções causadas pelos *outliers* nos métodos não-robustos, como o SR-TRA e SR-BOR. A robustez dos métodos SR-FAR e SR-SIM é muito importante nestas situações, observe que esses métodos não apresentam essas distorções visuais nas figuras 4.14 e 4.15.

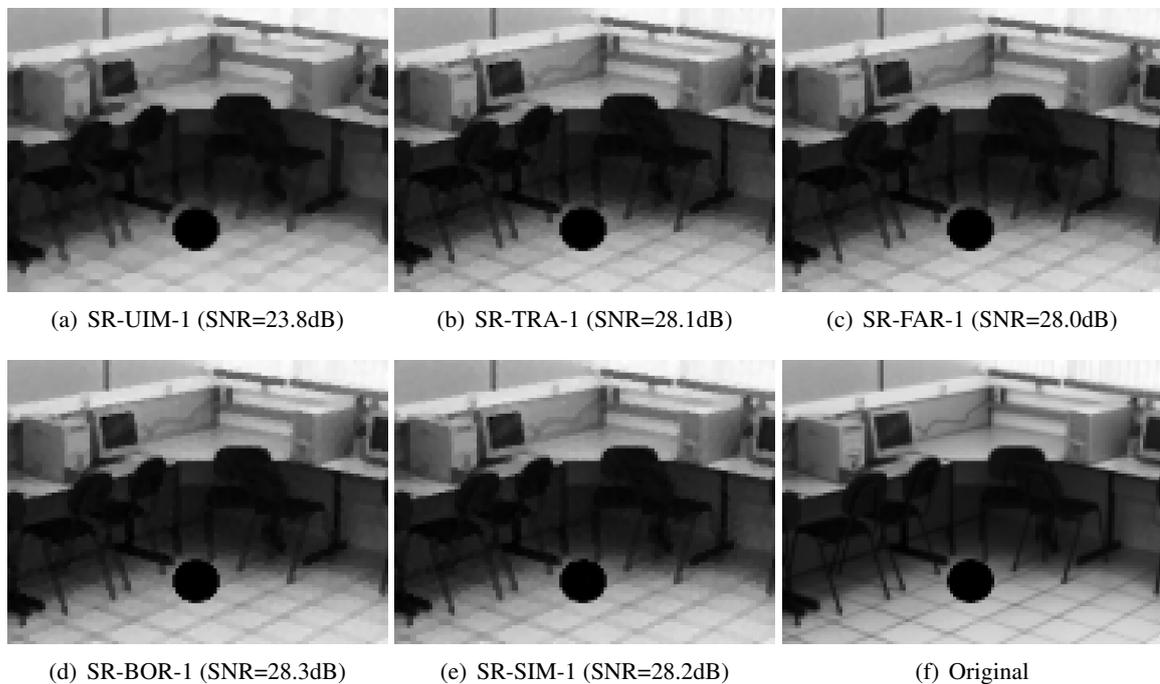


Figura 4.12: Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma Huber/ ℓ_1 , com modelos de primeira ordem e com detecção de *outliers* **ativada**.

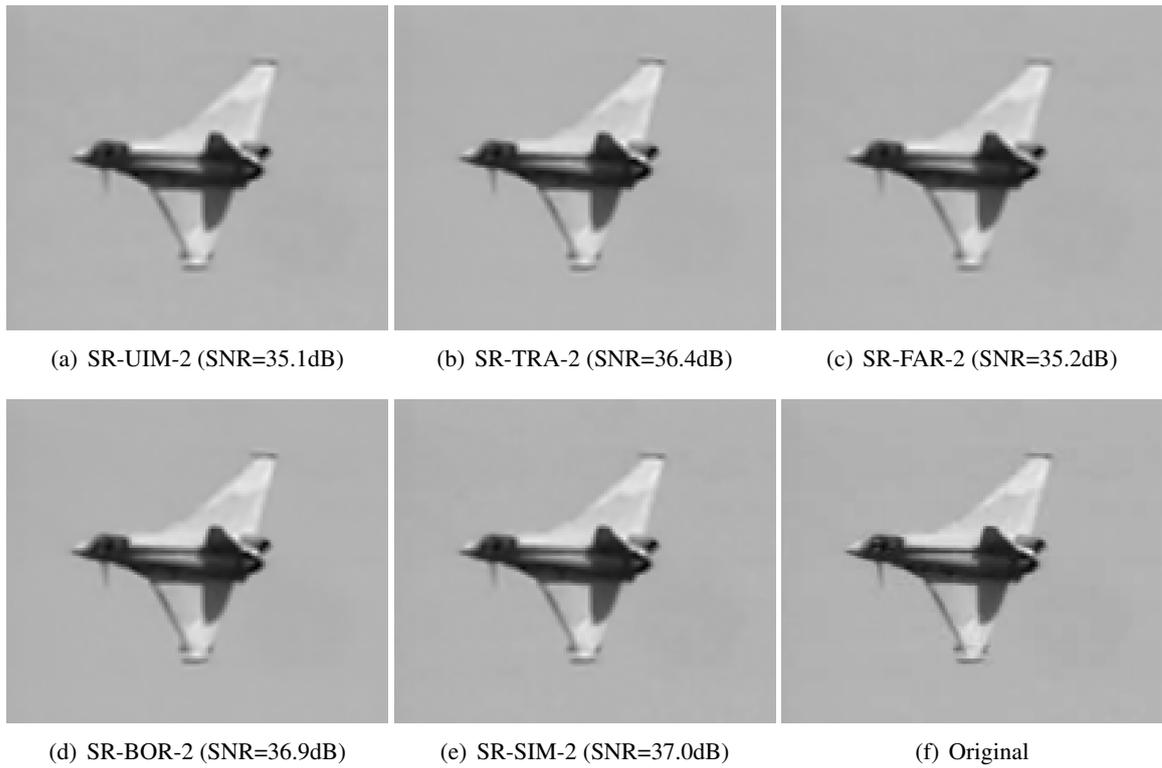


Figura 4.13: Exemplo visual mostrando a imagem 2 da seqüência Jato, usando norma Huber/ ℓ_1 , com modelos de segunda ordem e com detecção de *outliers* **ativada**.

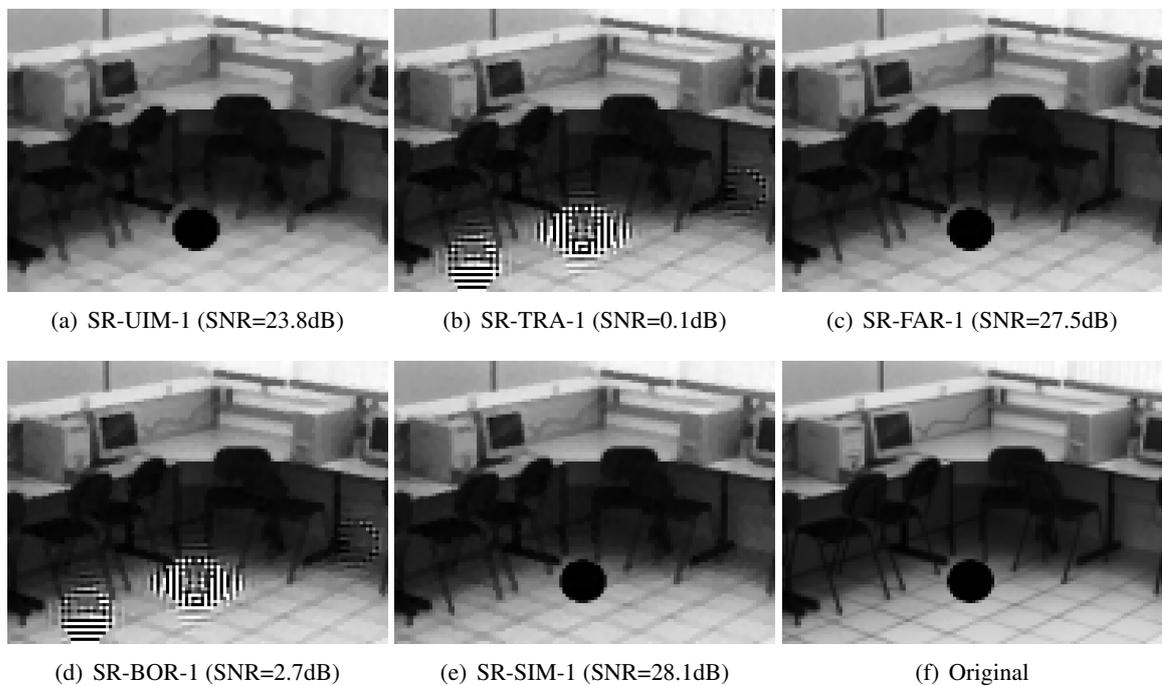


Figura 4.14: Exemplo visual mostrando a imagem 2 da seqüência Sala, usando norma Huber/ ℓ_1 , usando modelos de primeira ordem e com detecção de *outliers* **desativada**.

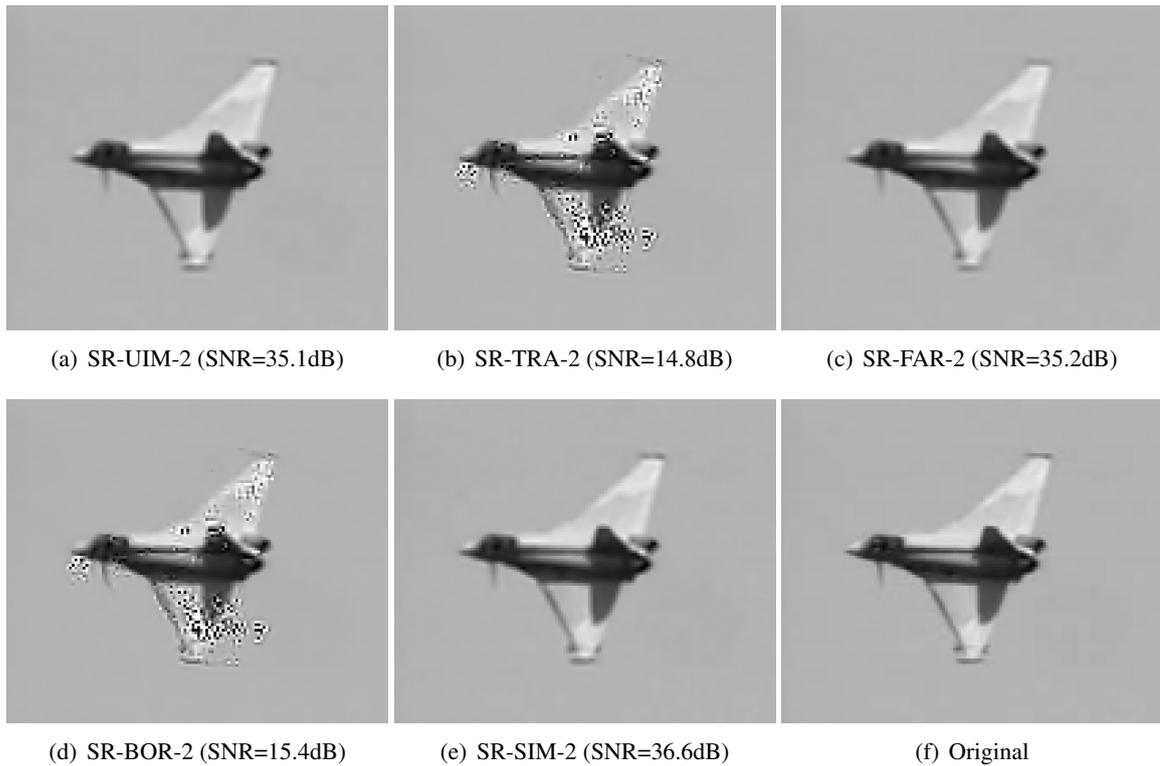


Figura 4.15: Exemplo visual mostrando a imagem 2 da seqüência Jato, usando norma Huber/ ℓ_1 , modelos de segunda ordem e com detecção de *outliers* **desativada**.

Uma observação importante deve ser feita com respeito aos resultados obtidos na seqüência “Jardim”. Observando os resultados nas tabelas 4.1 e 4.3, nota-se que houve uma perda significativa de qualidade em alguns algoritmos. Esta perda ocorreu porque a detecção dos *outliers* não foi eficiente, e alguns *outliers* não foram completamente eliminados. Contudo, a utilização da norma ℓ_2 ao quadrado, no operador espacial, reduziu as distorções através de uma maior suavização dos resultados. Já a norma Huber não suavizou devidamente as distorções causadas pelos *outliers* remanescentes, produzindo distorções nos algoritmos não-robustos, como o SR-TRA e o SR-BOR, mesmo com a detecção de *outliers* ativada. Este exemplo é mostrado nas figuras 4.16 e 4.17. Ele reforça a importância dos métodos robustos e mostra que o processo de detecção pode falhar.



Figura 4.16: Exemplo visual mostrando a imagem 6 da seqüência Jardim, usando norma ℓ_2 ao quadrado, modelos de primeira ordem e com detecção de *outliers* **ativada**.

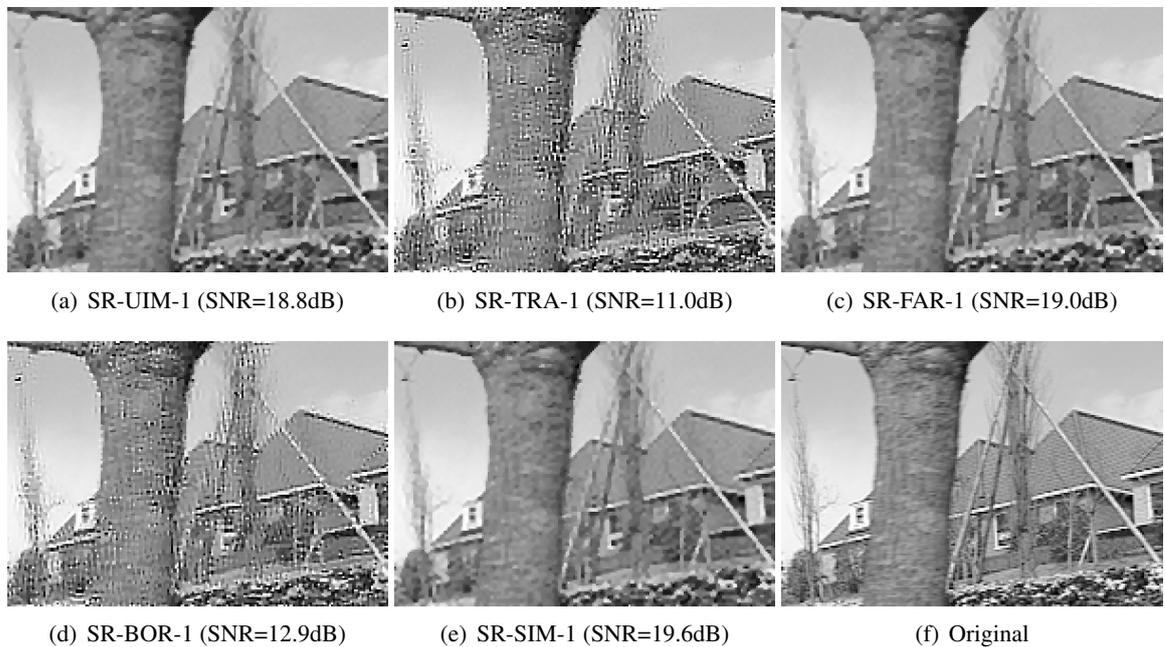


Figura 4.17: Exemplo visual mostrando a imagem 6 da seqüência Jardim, usando norma Huber/ ℓ_1 , modelos de primeira ordem e com detecção de *outliers* **ativada**.

4.2.2 Custo Computacional

O custo computacional dos métodos é ilustrado através das tabelas 4.5 e 4.6, mostrando o tempo necessário para convergência, em segundos. Nestas tabelas são mostrados o tempo necessários para

convergência de cada seqüência e a média do tempo relativo, que é o tempo necessário para convergência de um dado método em relação ao tempo que levou o método SR-UIM-1 com norma ℓ_2 ao quadrado.

Tabela 4.5: Custo Computacional dos métodos SR com norma ℓ_2 ao quadrado, em segundos

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
SR-UIM-1	2.0	4.7	2.5	4.3	7.0	20.0	1.0
SR-UIM-2	2.9	4.8	3.3	3.9	6.5	22.2	1.2
SR-TRA-1	4.7	10.1	5.4	6.7	36.5	51.5	2.5
SR-TRA-2	4.1	10.6	5.5	7.1	36.5	55.3	2.7
SR-BOR-1	9.2	19.0	8.5	12.5	62.0	79.3	4.6
SR-BOR-2	9.5	20.9	10.5	12.4	64.8	75.7	4.8
SR-SIM-1	8.0	12.1	5.0	7.6	14.1	40.1	2.4
SR-SIM-2	7.9	13.9	6.0	7.3	19.7	48.2	2.8

Tabela 4.6: Custo Computacional dos métodos SR com norma Huber ou ℓ_1 , em segundos

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
SR-UIM-1	16.0	25.0	12.0	27.9	86.0	91.5	6.9
SR-UIM-2	15.0	26.2	17.4	25.5	79.5	89.0	7.0
SR-TRA-1	21.0	40.0	32.2	50.1	170.0	265.0	13.5
SR-TRA-2	20.1	37.5	36.5	48.9	168.5	260.5	13.6
SR-FAR-1	30.0	64.0	40.9	55.3	298.0	210.0	18.2
SR-FAR-2	33.0	65.5	45.0	65.1	295.0	230.0	19.5
SR-BOR-1	34.5	68.5	47.5	70.5	310.0	325.0	21.2
SR-BOR-2	32.5	63.9	55.0	72.8	298.5	295.0	21.1
SR-SIM-1	28.5	63.1	29.7	50.4	151.0	180.0	13.6
SR-SIM-2	26.5	59.2	37.5	51.0	148.0	210.0	14.2

Na Tabela 4.5, na qual é apresentado o custo computacional dos métodos com norma ℓ_2 ao quadrado, pode-se notar a redução de custo computacional do método simultâneo proposto, o SR-SIM, em relação ao método simultâneo existente, o SR-BOR. Na maioria das seqüências, o custo computacional do método proposto, o SR-SIM, foi ligeiramente maior que o custo do método tradicional, o SR-TRA. Contudo, conforme o aumento do número de imagens na seqüência, o custo dos métodos tradicionais aumenta muito mais do que o custo do método proposto. Isso pode ser observado no tempo necessário para convergência nas seqüências “Jato” e “Jardim”.

Na Tabela 4.6, em que é apresentado o custo computacional dos métodos com norma Huber ou ℓ_1 , também se pode notar a redução de custo computacional do método simultâneo proposto, o SR-SIM, em relação ao método simultâneo existente, o SR-BOR. O custo computacional do método proposto é ligeiramente maior que o custo do método tradicional, o SR-TRA, mas menor que o custo do método

tradicional robusto, o SR-FAR. Nesta tabela, é possível observar que os métodos com norma Huber, ou ℓ_1 , têm custo computacional bem maior que os métodos baseados na norma ℓ_2 ao quadrado.

Conclusão do Capítulo

Neste capítulo foram apresentados resultados experimentais comparando o método proposto de super-resolução simultânea com os métodos regularizados existentes. Os resultados numéricos e visuais confirmaram o aumento da robustez e a redução de custo computacional do método proposto em relação ao método simultâneo existente, proposto em [1]. O método proposto obtém resultados com a alta qualidade dos métodos simultâneos existentes, mas com o baixo custo computacional dos métodos tradicionais. Além disso, o método simultâneo proposto, com norma Huber, tem uma robustez aos *outliers* maior que o método simultâneo existente. A robustez aos *outliers* do movimento no método proposto com norma Huber é semelhante à do método tradicional robusto, apresentado em [75].

Capítulo 5

Determinação dos Coeficientes de Regularização

A determinação dos coeficientes de regularização é um fator fundamental na busca por uma solução regularizada. Neste capítulo será apresentada a segunda contribuição desta tese: um novo método de determinação dos coeficientes de regularização. A técnica proposta pode ser aplicada tanto no problema de determinação de um único coeficiente de regularização, quanto no problema da determinação de vários coeficientes de regularização, encontrado no algoritmo de super-resolução simultânea proposto no Capítulo 3. Contudo, o método desenvolvido neste capítulo é aplicável apenas aos algoritmos SR baseados na norma ℓ_2 ao quadrado na função custo.

Na seção 5.1, uma revisão dos métodos existentes para determinação do coeficiente de regularização mais conhecidos é apresentada, enfatizando os métodos já testados em problemas de super-resolução. Estes métodos serão classificados quanto ao seu desempenho em problemas de super-resolução, considerando a qualidade da estimativa, o custo computacional envolvido e a estabilidade na busca da solução.

Na seção 5.2, o método proposto de determinação dos coeficientes é desenvolvido. O método é baseado na teoria estatística Bayesiana de *joint maximum a posteriori* (JMAP). No método JMAP, os coeficientes são estimados em conjunto com as imagens HR, sendo que uma nova informação *a priori* sobre os hiperparâmetros é incluída para proporcionar a estabilidade global da estimativa. O método desenvolvido apresenta, principalmente, baixo custo computacional e grande estabilidade, proporcionando uma qualidade de estimativa próxima à dos melhores métodos existentes.

5.1 Métodos Existentes de Determinação dos Coeficientes de Regularização para os Algoritmos de Super-Resolução

Esta seção apresenta os métodos existentes de determinação dos coeficientes de regularização para os algoritmos de super-resolução. Nesta seção, os objetivos dos métodos de determinação dos coeficientes são apresentados, os métodos mais utilizados são revisados e classificados, enfatizando aqueles já testados em problemas de super-resolução. Esta revisão é importante para definir um esquema apropriado para ser utilizado no algoritmo de SR simultânea proposto.

Embora o algoritmo de SR simultânea proposto necessite da determinação de múltiplos coeficientes de regularização, esta revisão irá abordar, primeiramente, os métodos desenvolvidos para um único coeficiente. O problema de super-resolução tradicional de um coeficiente é dado pela seguinte equação:

$$\begin{aligned}\hat{\mathbf{f}}_k &= \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_2^2 \\ \hat{\mathbf{f}}_k &= \arg \min_{\mathbf{f}_k} \|\mathbf{g} - \mathbf{C}_k \mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_2^2\end{aligned}\quad (5.1)$$

sendo $\mathbf{g} = [\mathbf{g}_1^T \dots \mathbf{g}_L^T]^T$ e $\mathbf{C}_k = [\mathbf{C}_{1,k} \dots \mathbf{C}_{L,k}]^T$. Para simplificar o desenvolvimento, o subíndice k será omitido das equações seguintes, ficando claro no contexto que a imagem HR estimada e seu coeficiente estão no mesmo instante de tempo.

5.1.1 Objetivo dos Métodos de Determinação dos Coeficientes de Regularização

Em geral, uma solução regularizada, $\hat{\mathbf{f}}_\lambda$, é considerada uma boa solução quando o λ correspondente produz o menor erro de estimativa $\mathbf{f} - \hat{\mathbf{f}}_\lambda$ possível, sendo \mathbf{f} a imagem HR original e $\hat{\mathbf{f}}_\lambda$ a imagem HR estimada com um determinado coeficiente, ou seja, o coeficiente λ escolhido faz com que a imagem HR estimada seja a mais parecida possível com a imagem HR original, segundo algum critério predefinido. Este coeficiente será chamado de λ_{opt} . A Figura 5.1 ilustra o erro $\|\mathbf{f} - \hat{\mathbf{f}}_\lambda\|_2$, em função do coeficiente. Na Figura 5.2 é ilustrado o aspecto visual dos resultados com diferentes coeficientes. Quando o λ é pequeno, bem menor que o λ_{opt} , a imagem estimada fica corrompida por ruído amplificado; quando o λ é grande, bem maior que o λ_{opt} , a imagem é excessivamente regularizada, geralmente ela fica exageradamente suave; quando o λ_{opt} , ou valores próximos a este, são utilizados, a imagem fica mais parecida com a imagem HR original.

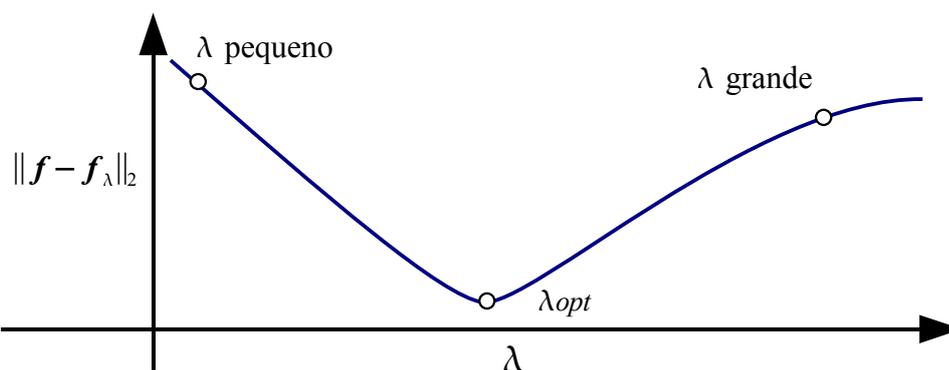
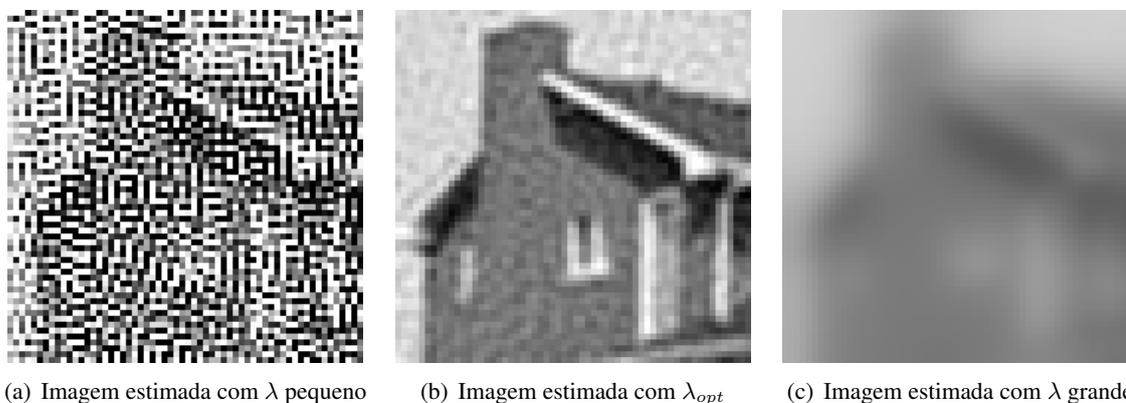


Figura 5.1: Comportamento esperado da Norma ℓ_2 do erro de estimativa.



(a) Imagem estimada com λ pequeno

(b) Imagem estimada com λ_{opt}

(c) Imagem estimada com λ grande

Figura 5.2: Resultados ilustrativos com λ muito pequeno, λ_{opt} e λ muito grande.

Contudo, o erro de estimativa só pode ser avaliado em situações experimentais. Na prática, esse erro não pode ser calculado porque a imagem HR original, f , é desconhecida. Por isso, um esquema de determinação dos coeficientes alternativo, que não requer a imagem HR original, é necessário.

Segundo a definição em [39], um esquema apropriado para escolha do coeficiente de regularização é aquele que obtém um coeficiente próximo do λ_{opt} da forma mais precisa e confiável possível, e, pode-se acrescentar ainda, com custo computacional aceitável. Quase todos os métodos revisados nesta seção foram desenvolvidos para alcançar uma precisão adequada, entretanto a confiabilidade (estabilidade¹) e o custo computacional ainda são questões graves nas aplicações de super-resolução, que precisam ser resolvidas, principalmente no caso de múltiplos coeficientes.

¹A *estabilidade*, considerada nesta tese, é definida como a capacidade do método de encontrar valores bem próximos quando o método é repetido em um determinado problema, sem divergir para zero ou para infinito, mesmo que as condições iniciais se modifiquem.

5.1.2 Revisão dos Métodos Existentes

Métodos Baseados na Curva-L

A Curva-L [39] é uma curva no plano 2D, denotada $l(\lambda) = (x(\lambda), y(\lambda))$, na qual cada ponto mostra o resíduo do termo de dados, $x(\lambda) = \Phi(\|g - Cf_\lambda\|_2^2)$ no eixo horizontal, e o resíduo do termo de regularização, $y(\lambda) = \Phi(\|Rf_\lambda\|_2^2)$ no eixo vertical. A Curva-L pode ser apresentada em diversas escalas, sendo que Φ pode ser um operador de logaritmo, raiz quadrada ou a identidade.

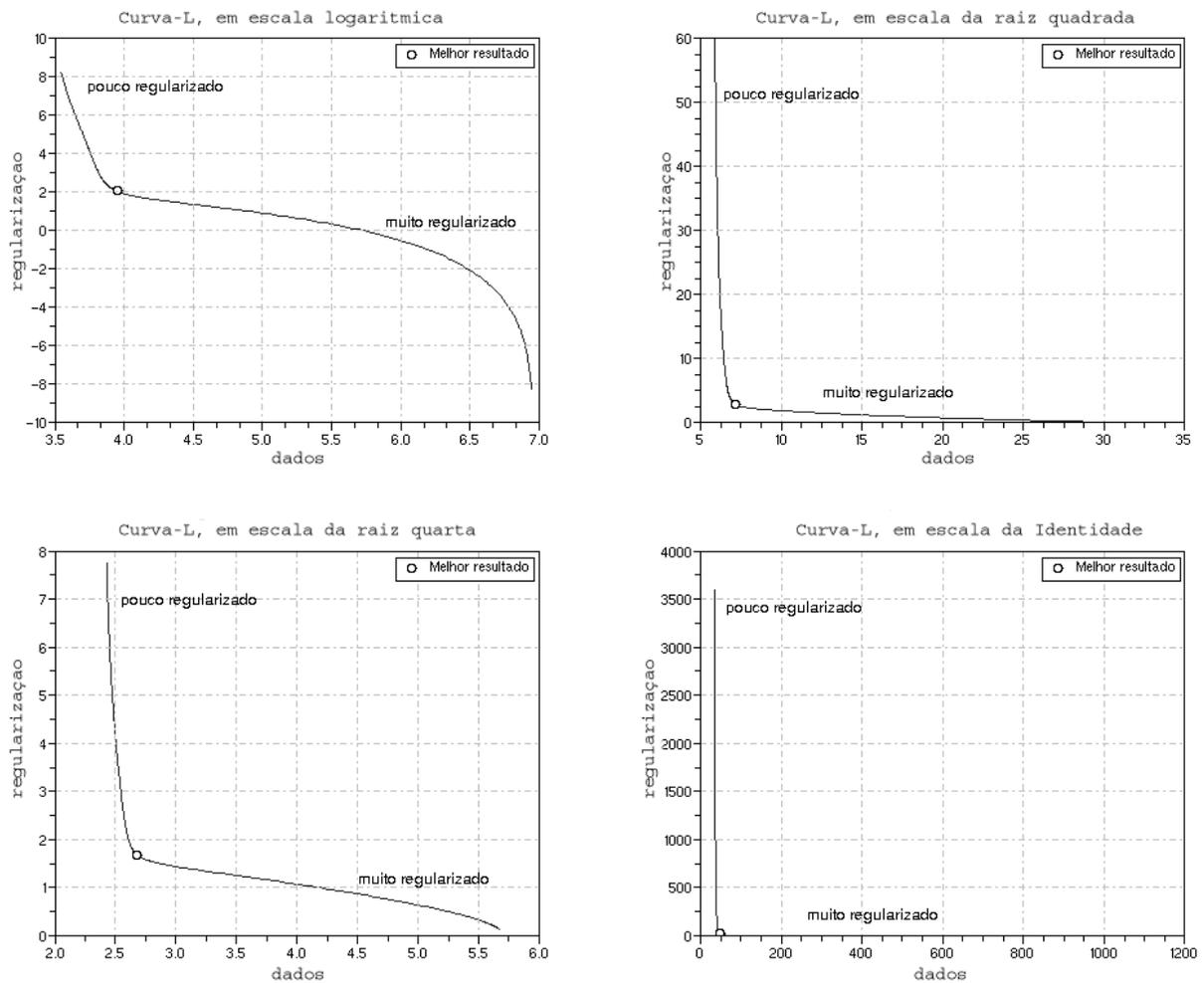


Figura 5.3: Curva-L em diversas escalas.

Exemplos típicos da Curva-L, em diversas escalas, são mostrados na Figura 5.3. Esta curva é chamada de Curva-L devido ao seu aspecto, tendo uma parte vertical, que representa as soluções ruidosas e com fortes oscilações (pouco regularizadas), que ocorrem quando λ é muito pequeno; e uma parte horizontal, que representa as soluções exageradamente suaves (muito regularizadas), que ocorrem quando λ é muito grande. Considera-se que as soluções mais adequadas estão no canto da Curva-L, na região

de transição entre a parte vertical e horizontal. Nessa região estão as soluções próximas do λ_{opt} , onde o resíduo do termo de dados e o resíduo do termo de regularização são pequenos. Pode-se observar na Figura 5.3 que o λ_{opt} está sempre próximo do “canto” da curva em todas as escalas.

Existem três critérios encontrados na literatura, baseados na Curva-L. Eles se diferenciam, basicamente, pela definição do ponto de solução, ou seja, na definição de “canto”. Todos exploram a Curva-L em escala log-log. São eles:

- **L-MC** - Ponto de máxima curvatura [39];
- **L-MD** - Ponto com menor distância Euclidiana da origem [78];
- **L-TN** - Ponto tangente a uma reta com decaimento negativo [77, 88].

Em [34] o principal critério baseado na Curva-L, o L-MC, apresentado em [39] para problemas inversos gerais, foi aplicado nos problemas de super-resolução. Este método considera que o ponto de máxima curvatura na Curva-L, em escala log-log, é uma escolha próxima do λ_{opt} . O método L-MC é, em geral, bastante preciso e estável, contudo está limitado a aplicações com um único coeficiente devido ao alto custo computacional. Em [78], em que o problema de múltiplos coeficientes é estudado, é mencionado que o custo computacional do método de máxima curvatura da Curva-L, chamada de Hipercurva-L quando múltiplos coeficientes são considerados, é geometricamente elevado com o aumento do número de coeficientes, tornando-se praticamente inviável.

Em [78] o critério L-MD, com menor custo computacional que o L-MC, foi desenvolvido para problemas inversos gerais envolvendo múltiplos coeficientes. Este método nunca foi aplicado em super-resolução, mas é um bom candidato para isso.

Um outra alternativa, de baixo custo computacional, é o critério L-TN, apresentada em [77]. Este método foi desenvolvido para problemas inversos gerais de um coeficiente apenas, mas pode ser expandido para aplicações com múltiplos coeficientes e também parece uma boa opção para o problema de super-resolução.

Apesar do baixo custo computacional, a estabilidade desses dois métodos alternativos, o L-MD e L-TN, não foi comprovada. Inclusive, em [77] é mencionado que, sem uma restrição apropriada e sem algumas condições específicas no problema inverso, o método pode divergir. Esta instabilidade ocorre devido ao método depender da convexidade da Curva-L para convergência, condição que não é satisfeita em escala log-log. Portanto, não há garantias de obtenção de uma solução confiável com os métodos L-MD e L-TN.

Generalized Cross-Validation (GCV)

O método *Generalized Cross-Validation* (GCV) é bastante utilizado em problemas inversos [40, 89]. Seu objetivo é encontrar um coeficiente de regularização que minimize o seguinte critério:

$$\hat{\lambda} = \arg \min_{\lambda} \frac{\|\mathbf{g} - \mathbf{C}\hat{\mathbf{f}}_{\lambda}\|_2^2}{\text{tr}(\mathbf{I} - \mathbf{C}\mathbf{C}_{\lambda}^{\#})^2} \quad (5.2)$$

sendo que $\text{tr}(\mathbf{A})$ é o traço da matriz \mathbf{A} , a matriz $\mathbf{C}_{\lambda}^{\#} = (\mathbf{C}^T\mathbf{C} + \lambda\mathbf{R}^T\mathbf{R})^{-1}\mathbf{C}^T$ é o sistema inverso regularizado e $\hat{\mathbf{f}}_{\lambda} = \mathbf{C}_{\lambda}^{\#}\mathbf{g}$ é a solução estimada com um dado λ .

Este método também é bastante preciso e estável. A superfície de desempenho típica é ilustrada na Figura 5.4, acompanhada da norma do erro de estimativa. Contudo, este método tem alto custo computacional devido ao cálculo do traço, o qual requer a matriz inversa do problema [40, 89].

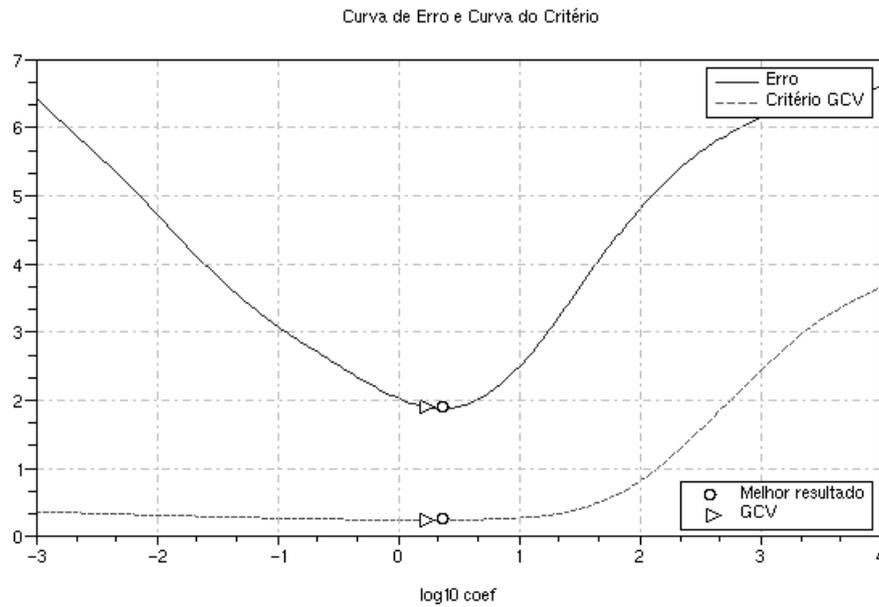


Figura 5.4: Curva da superfície de desempenho do método GCV.

Em [35] o GCV é aplicado em super-resolução. Devido às dificuldades em obter a matriz inversa, o traço não é diretamente calculado, mas estimado estatisticamente. A estimação estatística do traço permite aplicação do método em problemas com grandes dimensões [89]. Em [35], apenas problemas com um único coeficiente de regularização são tratados. Até o presente momento não foram encontrados trabalhos que apliquem o método GCV nos casos com múltiplos coeficientes em super-resolução.

Método de Kang e He (K-HE) e Outros

Outras alternativas de baixo custo computacional foram propostas em [38] e em [37]. Estes métodos fazem uma estimação conjunta dos coeficientes e das imagens HR usando minimização baseada no método do Gradiente. O método K-HE, que é uma extensão de [90], tem se mostrado estável nos experimentos relatados [38]. O coeficiente escolhido é:

$$\lambda(\mathbf{f}) = \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{\frac{1}{\gamma} - \|\mathbf{R}\mathbf{f}\|_2^2} \quad (5.3)$$

sendo que o λ é atualizado por esta equação a cada iteração do gradiente. Em [38], algumas análises que mostram a convergência do método do gradiente com a atualização do λ são apresentadas. Uma desvantagem desse método é a baixa velocidade de convergência do gradiente, uma limitação necessária para manter a estabilidade no método K-HE [38].

Em [37] uma outra aproximação é utilizada, na qual o coeficiente é atualizado por:

$$\lambda(\mathbf{f}) = \phi \left(\frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2 + \epsilon}{\|\mathbf{R}\mathbf{f}\|_2^2 + \epsilon} \right) \quad (5.4)$$

sendo que a função ϕ deve comprimir a atualização do λ , pois foi observado que se a função ϕ for a identidade, então a estimativa rapidamente diverge. Funções ϕ do tipo raiz quadrada e logarítmico são testadas com sucesso em [37]. Um valor pequeno deve ser dado ao ϵ para evitar λ extremos. Este método também tem baixa velocidade de convergência devido ao uso da minimização do Gradiente, além disso, não são apresentadas análises de estabilidade.

Métodos Estatísticos Bayesianos

Nos métodos estatísticos Bayesianos o λ recomendado não está diretamente relacionado com o λ_{opt} . Nestes métodos, o coeficiente depende da relação entre os hiperparâmetros das densidades assumidas para o termo de dados e para o termo de regularização. Entretanto, os resultados práticos observados mostram que os métodos Bayesianos obtêm praticamente a mesma qualidade dos métodos determinísticos apresentados anteriormente.

O método estatístico *Evidence* (EVID), proposto inicialmente em [79] para o problema de restauração de imagens com um único coeficiente [42], foi apresentado em [36] para o algoritmo de super-resolução tradicional com múltiplos coeficientes. Este método tem se mostrado bastante preciso e estável, entretanto tem alto custo computacional.

No *Evidence* o λ é atualizado a cada vez que uma nova imagem HR é encontrada. As equações são:

$$\hat{\theta}_{n+1} = \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}_n\|_2^2 + \text{tr}((\mathbf{C}^T\mathbf{C}/\theta_n + \mathbf{R}^T\mathbf{R}/\beta_n)^{-1}\mathbf{C}^T\mathbf{C})}{LN} \quad (5.5)$$

e

$$\hat{\beta}_{n+1} = \frac{\|\mathbf{R}\mathbf{f}_n\|_2^2 + \text{tr}((\mathbf{C}^T\mathbf{C}/\theta_n + \mathbf{R}^T\mathbf{R}/\beta_n)^{-1}\mathbf{R}^T\mathbf{R})}{M} \quad (5.6)$$

na qual $\lambda_{n+1} = \theta_{n+1}/\beta_{n+1}$, sendo θ e β os hiperparâmetros das distribuições dos dados e *a priori* das imagens, respectivamente. A imagem \mathbf{f}_n é encontrada com o λ_n da iteração anterior.

Da mesma forma que o GCV, o *Evidence* requer o cálculo de traços envolvendo a matriz inversa. Assim, em [36], o método *Evidence* é aplicado em problemas baseados em matrizes \mathbf{C} e \mathbf{R} bloco-circulantes, as quais podem ser diagonalizadas através da Transformada de Fourier. Contudo, utilizando estimação estatística do traço, como em [35], este método pode ser aplicado em matrizes que não são bloco-circulantes.

O método estatístico de *joint maximum a posteriori* (JMAP) [80, 91] é, em geral, uma alternativa de menor custo computacional que o *Evidence*. Em [91] foi proposta para estimação conjunta do movimento e das imagens HR. O trabalho [80] utiliza este método para definição dos coeficientes em reconstrução tomográfica. O grande problema deste método é a instabilidade da estimativa, conforme relatado em [79]. Na seção 5.2 a versão existente do JMAP é detalhada e o problema da instabilidade é explicado.

5.1.3 Classificação dos Métodos Revisados

Após a avaliação teórica e experimental desses métodos nos problemas de super-resolução, foi feita uma classificação quanto ao seu desempenho. Foram considerados: qualidade da estimativa, custo computacional e estabilidade na busca dos coeficientes. A Tabela 5.1 resume as características dos métodos avaliados.

Na Tabela 5.1 foram classificados como instáveis os métodos que, conforme o valor inicial do coeficiente ou nível de ruído, podem divergir para zero ou para infinito, em vez de convergir para um valor intermediário. A estabilidade é fundamental porque, caso ocorra divergência na estimativa do coeficiente, a imagem HR estimada pode ficar extremamente deteriorada; em muitos casos, fica pior que a imagem LR capturada, o que é proibitivo para um método considerado de “super-resolução”.

Tabela 5.1: Classificação dos Métodos de Determinação dos Coeficientes em Super-Resolução

Método	Qualidade	Custo Computacional	Estabilidade
GCV	Muito bom	Alto	Estável
L-MC	Muito bom	Alto	Estável
L-TN	Bom	Baixo	Instável
L-MD	Bom	Baixo	Instável
K-HE	Razoável	Baixo	Estável
EVIDENCE	Muito bom	Alto	Estável
JMAP	Bom	Baixo	Instável

Problema Identificado nos Métodos de Determinação dos Coeficientes

A classificação dos métodos encontrados na literatura mostra que, dentre os métodos estáveis, não foram encontrados na literatura métodos de baixo custo computacional que produzam imagens com qualidade satisfatória.

Esse problema motivou a busca por um método alternativo que seja, sobretudo, estável, mantendo uma boa qualidade da estimativa e um baixo custo computacional. Os estudos sobre o método JMAP demonstraram que ele pode ser estabilizado. Além disso, o desenvolvimento apresentado no Capítulo 3, utilizando a teoria Bayesiana MAP, é diretamente aproveitado. Logo, optou-se pelo desenvolvimento do método proposto a partir do JMAP.

5.2 Método Proposto de Determinação dos Coeficientes

Nesta seção, o método proposto de determinação dos coeficientes de regularização é apresentado. A proposta desenvolvida é baseada na estimação estatística JMAP. Além disso, o método desenvolvido está fortemente relacionado com o método da Curva-L proposto em [77].

O JMAP existente [80] assume uma função densidade de probabilidade *a priori* uniforme, entre zero e infinito, para os hiperparâmetros, o que acaba levando a um algoritmo instável. No JMAP proposto nesta seção, uma densidade *a priori* alternativa é assumida, que leva a um algoritmo estável e com solução global única. A seguir, será apresentada uma revisão do método JMAP existente e a apresentação do JMAP proposto.

5.2.1 Revisão do JMAP Existente

O JMAP é um método de estimação Bayesiana que visa à estimação conjunta das imagens HR e dos hiperparâmetros que formam os coeficientes [8, 80]. Para encontrar as imagens e os coeficientes,

o JMAP maximiza a densidade *a posteriori* conjunta. No problema de super-resolução tradicional com um único coeficiente, o λ é formado pela relação entre os hiperparâmetros θ e β , sendo que a estimativa global é:

$$\begin{aligned}\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} &= \arg \max_{\mathbf{f}, \theta, \beta} \rho(\mathbf{f}, \theta, \beta | \mathbf{g}) \\ \hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} &= \arg \max_{\mathbf{f}, \theta, \beta} \rho(\mathbf{g} | \mathbf{f}, \theta) \rho(\mathbf{f} | \beta) \rho(\theta) \rho(\beta)\end{aligned}\quad (5.7)$$

na qual $\rho(\mathbf{f}, \theta, \beta | \mathbf{g})$ é a densidade *a posteriori*; \mathbf{g} é o vetor representando os dados ou as imagens LR; \mathbf{f} é um vetor representando a imagem HR a ser estimada; θ é o hiperparâmetro da densidade dos dados $\rho(\mathbf{g} | \mathbf{f}, \theta)$, e β é o hiperparâmetro da densidade *a priori* da imagem HR, $\rho(\mathbf{f} | \beta)$. As densidades $\rho(\theta)$ e $\rho(\beta)$ são as densidades *a priori* dos hiperparâmetros. O hiperparâmetro dos dados, θ , é o mesmo em todas as imagens capturadas no problema de super-resolução tradicional com um único coeficiente. Neste desenvolvimento, assim como em [79, 80], é assumido que θ e β são independentes. Esta hipótese implica que o hiperparâmetro do modelo *a priori* de \mathbf{f} não está estatisticamente relacionado ao hiperparâmetro da densidade dos dados.

A densidade dos dados e a densidade *a priori* das imagens são, em geral, as mesmas utilizadas na estimativa MAP. Neste desenvolvimento foram consideradas as seguintes densidades Gaussianas:

$$\rho(\mathbf{g} | \mathbf{f}, \theta) = \frac{1}{(2\pi\theta)^{LN/2}} e^{-\frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{2\theta}} \quad (5.8)$$

para a densidade dos dados, na qual θ é o hiperparâmetro associado, que neste caso corresponde à variância do ruído, e:

$$\rho(\mathbf{f} | \beta) = \frac{1}{(2\pi\beta)^{M/2} |\mathbf{R}^T \mathbf{R}|^{-1/2}} e^{-\frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta}} \quad (5.9)$$

para a densidade *a priori* da imagem, sendo que β é o hiperparâmetro associado.

Na estimativa MAP é assumido que os hiperparâmetros são conhecidos e têm valores fixos; logo, não há necessidade de estimá-los. Neste caso, o coeficiente resultante é $\lambda = \theta/\beta$. Por outro lado, na estimativa JMAP os hiperparâmetros são variáveis aleatórias que precisam ser estimadas a partir dos dados, assim como a imagem HR. Por consequência, da mesma forma que uma densidade *a priori* precisa ser definida para as imagens, uma densidade *a priori* também precisa ser definida para os hiperparâmetros.

Em [79, 80] é mencionado que a definição de uma densidade “não-informativa” para os hiperparâmetros é adequada, dado que os hiperparâmetros não são conhecidos *a priori*. Isso é feito assumindo que os hiperparâmetros têm densidade uniforme, ou seja, $\rho(\theta) \propto cte$ e $\rho(\beta) \propto cte$, para $0 < \theta, \beta < \infty$. A

estimativa JMAP é:

$$\begin{aligned}\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} &= \arg \max_{\mathbf{f}, \theta, \beta} \rho(\mathbf{g}|\mathbf{f}, \theta) \rho(\mathbf{f}|\beta) \rho(\theta) \rho(\beta) \\ \hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} &= \arg \min_{\mathbf{f}, \theta, \beta} -\ln \rho(\mathbf{g}|\mathbf{f}, \theta) \rho(\mathbf{f}|\beta) \rho(\theta) \rho(\beta) \\ \hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} &= \arg \min_{\mathbf{f}, \theta, \beta} -\ln \rho(\mathbf{g}|\mathbf{f}, \theta) - \ln \rho(\mathbf{f}|\beta) - \ln \rho(\theta) - \ln \rho(\beta)\end{aligned}\quad (5.10)$$

Substituindo as densidades assumidas:

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg \min_{\mathbf{f}, \theta, \beta} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{2\theta} + \frac{LN}{2} \ln \theta + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} + \frac{M}{2} \ln \beta + cte + cte \quad (5.11)$$

A partir da equação (5.11) é possível encontrar os hiperparâmetros para um vetor fixo \mathbf{f} . Para isso basta diferenciar a equação (5.11), em função do respectivo hiperparâmetro, e igualar a zero, o que leva às seguintes soluções:

$$\hat{\theta} = \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{LN} \quad (5.12)$$

para o hiperparâmetro dos dados, e:

$$\hat{\beta} = \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{M} \quad (5.13)$$

para o hiperparâmetro da imagem. Pode-se observar que a estimativa JMAP dos hiperparâmetros, com densidade uniforme, é similar à estimativa de Máxima Verossimilhança (ML) dos hiperparâmetros [8].

Substituindo esses resultados novamente na equação (5.11), obtém-se a seguinte função custo global para a imagem HR:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \ln(\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2) + \frac{M}{LN} \ln(\|\mathbf{R}\mathbf{f}\|_2^2) \quad (5.14)$$

na qual a imagem HR de solução é encontrada quando:

$$\mathbf{C}^T \mathbf{C} + \frac{M}{LN} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{\|\mathbf{R}\mathbf{f}\|_2^2} \mathbf{R}^T \mathbf{R} \mathbf{f} = \mathbf{C}^T \mathbf{g} \quad (5.15)$$

sendo que o λ encontrado é:

$$\lambda = \frac{\theta}{\beta} = \frac{M}{LN} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{\|\mathbf{R}\mathbf{f}\|_2^2} \quad (5.16)$$

A dificuldade no uso de (5.14) está na não-convexidade desta equação [79]; logo, ela pode ter vários mínimos. Mesmo sendo desenvolvida do ponto de vista estatístico, a equação (5.14) tem grande

similaridade com o método determinístico da Curva-L proposto em [77]. O critério proposto em [77] é:

$$\hat{\lambda} = \arg \min_{\lambda} (\|\mathbf{g} - \mathbf{Cf}_{\lambda}\|_2) (\|\mathbf{Rf}_{\lambda}\|_2)^{\gamma} \quad (5.17)$$

o qual pode ser escrito como:

$$\hat{\lambda} = \arg \min_{\lambda} \ln(\|\mathbf{g} - \mathbf{Cf}_{\lambda}\|_2^2) + \gamma \ln(\|\mathbf{Rf}_{\lambda}\|_2^2) \quad (5.18)$$

A análise da Curva-L em [77] mostra que a curva, em escala logarítmica, não é convexa, ou seja, é necessária uma restrição na estimativa de λ para encontrar o mínimo local apropriado.

A restrição de λ , no sentido estatístico, pode ser expressa através da definição de densidades *a priori* para os hiperparâmetros. Isso significa que a definição de $\rho(\theta) = \rho(\beta) \propto cte$ não restringe adequadamente a estimativa dos hiperparâmetros, o que se traduz em um algoritmo instável. Logo, uma informação *a priori* ligeiramente mais restritiva para os hiperparâmetros pode tornar a estimativa mais estável, sendo que o objetivo final é alcançar uma função custo com um único mínimo global.

5.2.2 Método JMAP Proposto

A instabilidade da estimativa em (5.14) é mencionada em [79]. Algumas tentativas de estabilização do JMAP através do uso de informação *a priori* para os hiperparâmetros são apresentadas em [92] para problemas inversos gerais. Seguindo essa idéia, buscamos uma densidade *a priori* para os hiperparâmetros que seja capaz de estabilizar o JMAP, levando a uma solução única para a estimativa conjunta.

Na estimativa JMAP, a densidade dos dados ou a densidade *a priori* das imagens estão conectadas com as densidades dos respectivos hiperparâmetros. Por exemplo, a densidade *a priori* da imagem fornece ao algoritmo a informação de que a imagem é suave, restringindo a estimativa da imagem. O hiperparâmetro associado fornece a informação de “o quanto suave” é a imagem desejada. Todavia, se for assumida uma densidade uniforme, como $\rho(\beta) \propto cte$, então se está explicitamente assumindo que uma imagem extremamente suave, como uma imagem com intensidade constante (tipo um tom de cinza único), é tão provável de ocorrer quanto uma imagem altamente ruidosa, como a resultante da estimativa sem regularização. Em outras palavras, uma densidade uniforme não é uma densidade *a priori* adequada para os hiperparâmetros neste problema, porque não restringe a ocorrência de valores extremos nos hiperparâmetros.

Nesse sentido, uma informação *a priori* apropriada deve evitar que a estimativa dos hiperparâmetros

divirja para valores extremos. A densidade de probabilidade *a priori* desejada para os hiperparâmetros deve especificar que estes são sempre positivos e têm baixa probabilidade de ocorrência de valores extremamente altos ou extremamente baixos. Assim, diversas densidades com essas propriedades foram analisadas, entre elas: gama, gama-inversa, log-normal, Maxwell, Rayleigh e Weibull. Dentre as densidades analisadas, a gama, com parâmetros específicos que a torna similar à densidade chi-quadrada, se mostrou mais interessante para desenvolvimento, tanto do ponto de vista prático quanto do ponto de vista teórico. A Figura 5.5 ilustra a densidade gama sugerida.

A densidade gama proporciona a restrição necessária para estabilizar o problema, tornando a função custo convexa em todo o domínio e com solução única. Além disso, o processo de estimação resultante é bastante simples, o que resulta em uma estimação de baixo custo computacional e torna as análises mais simples. Conceitualmente, é mostrado em [93] que uma variável aleatória resultante da soma dos quadrados de variáveis aleatórias Gaussianas, como observado nas equações (5.12) e (5.13), tem densidade chi-quadrada (um caso particular da densidade gama [93]). Isso sugere que a densidade gama é uma densidade apropriada para o problema. Assim, nesta tese é proposto o uso da densidade gama para os hiperparâmetros, com parâmetros específicos que a torna similar à densidade chi-quadrada.

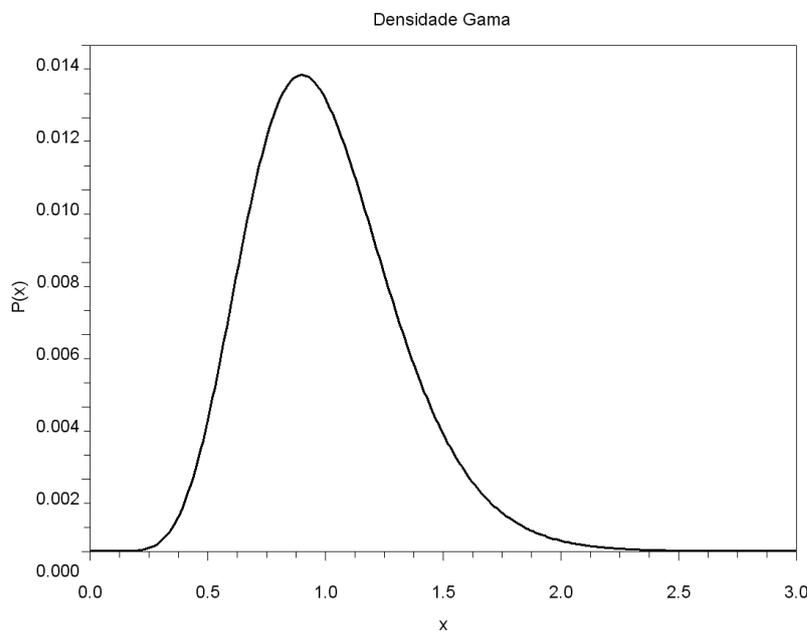


Figura 5.5: Densidade Gama, sugerida para os hiperparâmetros, com média 1 e variância 0.1.

A densidade gama para os hiperparâmetros é dada por:

$$\rho(\theta) = \frac{\theta^{a-1} b^{-a}}{\Gamma(a)} e^{-\frac{\theta}{b}} \quad (5.19)$$

sendo que a é o fator de escala e b é o fator de forma, e $E\{\theta\} = ab$ e $\text{var}\{\theta\} = ab^2$; e também:

$$\rho(\beta) = \frac{\beta^{c-1}d^{-c}}{\Gamma(c)}e^{-\frac{\beta}{d}} \quad (5.20)$$

sendo que c é o fator de escala e d é o fator de forma, na qual $E\{\beta\} = cd$ e $\text{var}\{\beta_R\} = cd^2$.

Para $a - 1 = LN/2$ e $c - 1 = M/2$, a densidade gama tem aproximadamente a mesma forma da chi-quadrada. Neste trabalho, os parâmetros a e b são fixos nesses valores, enquanto os outros dois parâmetros podem ser determinados pelo usuário, definindo valores esperados de θ , ou m_θ , e de β , ou m_β . A princípio, estes valores esperados podem ser retirados do conhecimento sobre o ruído no sensor de imagem [85] e do conhecimento sobre o comportamento médio do hiperparâmetro da imagem para uma classe de imagens de interesse [51, 42]. Contudo, mais adiante será visto que esses valores esperados não precisam ser conhecidos se algumas hipóteses sobre o problema forem satisfeitas.

Substituindo as densidades gama na equação (5.10) obtém-se:

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg \min_{\mathbf{f}, \theta, \beta} - \ln \rho(\mathbf{g}|\mathbf{f}, \theta) - \ln \rho(\mathbf{f}|\beta) - \ln \rho(\theta) - \ln \rho(\beta) \quad (5.21)$$

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg \min_{\mathbf{f}, \theta, \beta} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{2\theta} + \frac{LN}{2} \ln \theta - (a-1)\ln \theta + \frac{\theta}{b} + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} + \frac{M}{2} \ln \beta - (c-1)\ln \beta + \frac{\beta}{d} + cte \quad (5.22)$$

Utilizando os valores de a e b mencionados, ocorre o cancelamento dos termos com logaritmo:

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg \min_{\mathbf{f}, \theta, \beta} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{2\theta} + \frac{LN}{2} \ln \theta - \frac{LN}{2} \ln \theta + \frac{\theta(LN + 2)}{2m_\theta} + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} + \frac{M}{2} \ln \beta - \frac{M}{2} \ln \beta + \frac{\beta(M + 2)}{2m_\beta} + cte \quad (5.23)$$

Isso pode ser reduzido a:

$$\hat{\mathbf{f}}, \hat{\theta}, \hat{\beta} = \arg \min_{\mathbf{f}, \theta, \beta} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2}{2\theta} + \frac{\theta(LN + 2)}{2m_\theta} + \frac{\|\mathbf{R}\mathbf{f}\|_2^2}{2\beta} + \frac{\beta(M + 2)}{2m_\beta} \quad (5.24)$$

Diferenciando a equação (5.24) em função dos hiperparâmetros, para um vetor \mathbf{f} fixo, tem-se a se-

guinte estimativa de θ e β :

$$\hat{\theta} = \sqrt{m_\theta} \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2}{\sqrt{LN + 2}} \quad (5.25)$$

e

$$\hat{\beta} = \sqrt{m_\beta} \frac{\|\mathbf{R}\mathbf{f}\|_2}{\sqrt{M + 2}} \quad (5.26)$$

Substituindo as equações (5.25) e (5.26) novamente na equação (5.24), obtém-se:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \frac{\sqrt{LN + 2}}{\sqrt{m_\theta}} \|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2 + \frac{\sqrt{M + 2}}{\sqrt{m_\beta}} \|\mathbf{R}\mathbf{f}\|_2 \quad (5.27)$$

a qual pode ser reduzida a:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2 + \mu \|\mathbf{R}\mathbf{f}\|_2 \quad (5.28)$$

sendo:

$$\mu = \sqrt{\frac{m_\theta(M + 2)}{m_\beta(LN + 2)}} = \sqrt{m_\lambda} \frac{\sqrt{(M + 2)}}{\sqrt{(LN + 2)}} \quad (5.29)$$

Observe que, em (5.28), a função custo envolve a soma das normas, e não a soma dos quadrados das normas, como em (5.1), nem a soma do logaritmo das normas, como em (5.14). A convexidade do critério em (5.28) é demonstrada no Apêndice C.2.

O mínimo da equação (5.28) é encontrado quando:

$$\mathbf{C}^T \mathbf{C} + \mu \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2}{\|\mathbf{R}\mathbf{f}\|_2} \mathbf{R}^T \mathbf{R} = \mathbf{C}^T \mathbf{g} \quad (5.30)$$

sendo que o coeficiente de regularização é:

$$\lambda = \mu \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2}{\|\mathbf{R}\mathbf{f}\|_2} \quad (5.31)$$

o qual envolve apenas as normas, e não o quadrado das normas, como em (5.16).

Observe que é o valor de μ que precisa ser especificado, o qual pode ser definido a partir do conhecimento sobre o problema, através de m_θ e m_β . Contudo, soluções alternativas para a definição de μ , baseadas na análise do erro de estimativa, podem ser definidas. Nesta tese o μ sugerido é:

$$\mu = \frac{\sqrt{\text{tr}(\mathbf{C}^T \mathbf{C})/M}}{\sqrt{\text{tr}(\mathbf{R}^T \mathbf{R})/M}} \quad (5.32)$$

sendo $\text{tr}(\mathbf{A})$ é o traço da matriz \mathbf{A} . As análises e as hipóteses assumidas que levam a essa escolha

alternativa são apresentadas no Apêndice C.4. Esta definição alternativa é utilizada nos experimentos do Capítulo 6.

5.2.3 Métodos de Otimização Propostos

Neste trabalho dois métodos de otimização diferentes são propostos para solucionar este problema. O primeiro método utiliza uma forma iterativa alternada, com Gradiente Conjugado e Aproximações Sucessivas. Em uma etapa os coeficientes são fixos e a imagem HR é estimada, utilizando o método de Gradiente Conjugado (GC). Em outra etapa a imagem HR previamente estimada é utilizada para calcular o novo coeficiente, através de Aproximações Sucessivas. O segundo método utiliza a minimização direta da função custo em (5.28), usando Gradiente Conjugado Não-Linear (GC-NL), sendo que em cada iteração tanto os coeficientes quanto a imagem HR são atualizadas. As análises de convexidade e convergência dos dois métodos são apresentadas nos apêndices C.2 e C.3, respectivamente.

Método Iterativo Alternado

O primeiro método segue o seguinte procedimento:

- Dado um valor de λ_n , a imagem HR da iteração n é estimada:

$$\mathbf{f}_n = \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2 + \lambda_n \|\mathbf{R}\mathbf{f}\|_2^2 \quad (5.33)$$

Esta minimização é realizada, até a convergência, usando o método iterativo do Gradiente Conjugado Linear [40, 83], sendo que a estimativa inicial da imagem HR é a solução da iteração anterior.

- Usando a imagem HR estimada, o coeficiente é atualizado, através de:

$$\lambda_{n+1} = \mu \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}_n\|_2}{\|\mathbf{R}\mathbf{f}_n\|_2} \quad (5.34)$$

Essas duas etapas configuram a iteração n do método alternado. As iterações terminam quando um valor suficientemente pequeno para $|\lambda_{n+1} - \lambda_n|/\lambda_n$ é encontrado.

Método de Minimização Direta

O segundo método é o Gradiente Conjugado Não-Linear, usado para minimizar:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2 + \mu \|\mathbf{R}\mathbf{f}\|_2 \quad (5.35)$$

Este procedimento é descrito no Apêndice C.1. Espera-se que este método tenha um desempenho superior.

5.2.4 Extensão para Super-Resolução com Múltiplos Coeficientes

O método proposto pode ser facilmente estendido para problemas com múltiplos coeficientes. Estas extensões foram desenvolvidas mantendo as mesmas propriedades do caso unidimensional: a convexidade da função custo global e a similaridade com o critério da Curva-L em [77], em escala sqrt-sqrt.

Para o caso de super-resolução tradicional, em que o coeficiente e as ponderações precisam ser determinados, a função custo proposta passa a ser:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^k \gamma_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2 + \mu_k \|\mathbf{R}_k \mathbf{f}_k\|_2 \quad (5.36)$$

Observe que somente a norma, e não a norma ao quadrado, é considerada em (5.36).

As ponderações e o coeficiente resultantes são:

$$\alpha_{j,k} = \gamma_{j,k} \frac{\|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2}{\|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2} \quad (5.37)$$

para as ponderações, e:

$$\lambda_k = \mu_k \frac{\|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2}{\|\mathbf{R}_k \mathbf{f}_k\|_2} \quad (5.38)$$

para o coeficiente de regularização.

A definição dos $\gamma_{j,k}$ e dos μ_k para este problema ainda deve ser melhor desenvolvida. Para os experimentos do Capítulo 6 foram utilizados:

$$\gamma_{j,k} = \frac{1}{1 + |j - k|} \quad (5.39)$$

usado para reduzir a influência das imagens LR em instantes temporais mais afastados do instante da

imagem HR que será estimada, e:

$$\mu_k = \frac{\sqrt{\text{tr}(\mathbf{D}_k^T \mathbf{D}_k)/N}}{\sqrt{\text{tr}(\mathbf{R}_k^T \mathbf{R}_k)/M}} \quad (5.40)$$

Para o método de super-resolução simultânea, em que dois coeficientes são considerados, a função custo proposta é:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2 + \mu_R \|\mathbf{R}\mathbf{f}\|_2 + \mu_M \|\mathbf{M}\mathbf{f}\|_2 \quad (5.41)$$

sendo que os coeficientes de regularização são:

$$\lambda_R = \mu_R \frac{\|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2}{\|\mathbf{R}\mathbf{f}\|_2} \quad (5.42)$$

para a penalidade que impõe suavidade espacial, e:

$$\lambda_M = \mu_M \frac{\|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2}{\|\mathbf{M}\mathbf{f}\|_2} \quad (5.43)$$

para a penalidade que impõe similaridade na trajetória do movimento.

As definições de μ_R e de μ_M para o problema de super-resolução simultânea precisam ser mais desenvolvidas. Contudo, para os experimentos do Capítulo 6 foram utilizados:

$$\mu_R = \frac{\sqrt{\text{tr}(\mathbf{D}^T \mathbf{D})/N}}{\sqrt{2\text{tr}(\mathbf{R}^T \mathbf{R})/M}} \quad (5.44)$$

e

$$\mu_M = \frac{\sqrt{\text{tr}(\mathbf{D}^T \mathbf{D})/N}}{\sqrt{2\text{tr}(\mathbf{M}^T \mathbf{M})/M}} \quad (5.45)$$

Os dois métodos de otimização apresentados na 5.2.3 podem ser empregados para minimizar as funções custo nos problemas de múltiplos parâmetros.

Conclusão do Capítulo

Neste capítulo foi apresentada a segunda contribuição desta tese: um novo método de determinação dos coeficientes. Após uma revisão dos métodos mais utilizados na determinação dos coeficientes de regularização, seguido de uma classificação dos mesmos quanto ao desempenho em problemas de super-resolução, foi observado que o método JMAP tinha grande potencial de utilização devido ao baixo custo computacional. O método JMAP existente foi revisado e o problema da instabilidade foi explicado.

Partindo do JMAP existente, foi proposta uma versão estável do JMAP, utilizando densidade gama nos hiperparâmetros. Dois métodos rápidos de otimização foram apresentados para o problema e as extensões do método para os algoritmos SR de múltiplos coeficientes foi apresentada. No capítulo a seguir, serão apresentados os experimentos contrastando o método proposto com os métodos existentes. Os experimentos ilustram o baixo custo computacional e a estabilidade da estimativa, alcançando resultados com boa qualidade.

Capítulo 6

Experimentos Comparativos dos Métodos de Determinação dos Coeficientes

Neste capítulo os métodos de determinação dos coeficientes, apresentados no capítulo anterior, são comparados experimentalmente. Os métodos de determinação dos coeficientes são testados em três algoritmos de super-resolução diferentes: no algoritmo SR tradicional com um único coeficiente, no algoritmo SR tradicional com múltiplos coeficientes, no qual o coeficiente de regularização e as ponderações são determinados, e no algoritmo SR simultâneo proposto nesta tese, considerando dois coeficientes de regularização.

Em cada algoritmo de super-resolução, o desempenho dos métodos dos coeficientes é avaliado utilizando dois níveis de ruído diferentes e 6 seqüências de imagens com níveis de erro de movimento diferentes, para uma comparação de desempenho mais abrangente. Os resultados ilustram a melhoria obtida pelo método proposto em relação aos principais métodos existentes, principalmente nas questões de custo computacional e estabilidade da estimativa.

Na seção 6.1, o procedimento realizado para comparação dos métodos é descrito, detalhando-se as situações em que os métodos foram testados, as quais são comuns aos três algoritmos SR.

Nas seções seguintes são apresentadas as particularidades dos experimentos com cada algoritmo de super-resolução e seus respectivos resultados. Na seção 6.2 estão os detalhes do experimento com o algoritmo SR tradicional com um único coeficiente. Na seção 6.3, está o detalhamento do experimento com o algoritmo SR tradicional com múltiplos coeficientes, e, finalmente, na seção 6.4, a descrição do experimento com o algoritmo SR simultâneo proposto nesta tese, com dois coeficientes de regularização.

6.1 Descrição dos Experimentos

6.1.1 Procedimento

Para avaliar a capacidade dos métodos em encontrar os melhores coeficientes em cada um dos algoritmos SR mencionados, o seguinte experimento foi realizado:

- Definida uma seqüência de imagens HR conhecida, entre as 6 seqüências utilizadas no Capítulo 4, com o movimento conhecido ou previamente estimado, foi simulado o processo de aquisição. Utilizou-se uma matriz D , conhecida, cuja função de aquisição é a média de uma área retangular de 2×2 pixels, conforme a Figura 6.1.

$$d_k[m_1, m_2] = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Figura 6.1: Função discreta de aquisição utilizada.

- Adicionou-se um ruído gaussiano branco nas imagens LR, cuja variância é calculada de forma a se obter uma medida de SNR^1 de aquisição desejada. Duas situações foram consideradas:
 - alto ruído de aquisição, com SNR de aquisição de 20dB;
 - baixo ruído de aquisição, com SNR de aquisição de 40dB.

A Figura 6.2 mostra a aparência visual das imagens capturadas com esses níveis de ruído.

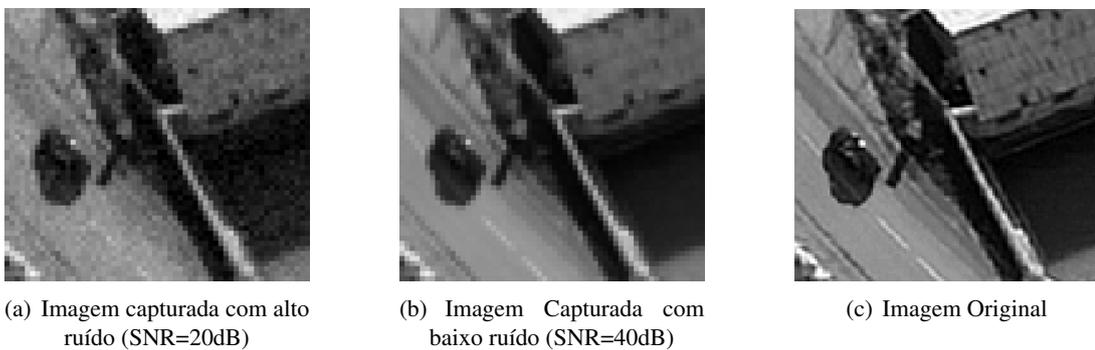


Figura 6.2: Ilustração do nível de ruído na aquisição.

- Os três algoritmos de super-resolução foram utilizados nessas seqüências de imagens LR. Em

¹A medida de SNR de aquisição é definida como $SNR = 10 \log_{10}(\sigma_{Df}^2 / \sigma_{\eta}^2)$, sendo que σ_{Df}^2 é a variância da seqüência LR sem ruído e σ_{η}^2 é a variância do ruído de aquisição.

cada algoritmo foram aplicados todos os métodos de determinação dos coeficientes viáveis a este, conforme descrito em detalhes nas seções seguintes.

- Foi calculada a qualidade das imagens HR estimadas, em termos de SNR (dB) [42], com os coeficientes encontrados e o tempo, em segundos, para cada método encontrar os respectivos coeficientes. Assume-se que a convergência foi alcançada quando o melhoramento de qualidade, em SNR (dB), da iteração é menor que 10^{-2} .
- Esse procedimento, de aquisição e estimação, foi repetido 20 vezes para a mesma seqüência, cada vez com um ruído de aquisição diferente, mas de mesma variância. Após as 20 realizações, foram calculados os resultados médios de cada método de determinação dos coeficientes, em cada um dos três algoritmos de super-resolução.
- Todo o procedimento foi repetido para outra seqüência de imagens, entre as 6 utilizadas neste experimento. As seqüências utilizadas são as mesmas utilizadas nos experimentos do Capítulo 4.

6.1.2 Métodos de Determinação dos Coeficientes Implementados

Abaixo estão listados os métodos de determinação dos coeficientes implementados nestes experimentos. Todos os métodos são aplicados no algoritmo SR tradicional de um único coeficiente; alguns deles, conforme mencionado a seguir, são aplicados nos algoritmos SR de múltiplos coeficientes.

Os métodos implementados são:

- **GCV** - O método *Generalized Cross Validation* [40, 89], bastante utilizado em diversas aplicações em problemas inversos, proposto em super-resolução em [35]. Este método é aplicado apenas no problema de um único coeficiente. Os detalhes do método estão na seção 5.1.
- **L-MC** - Método baseado na Curva-L, no qual a escolha é o ponto de máxima curvatura, conforme [39]. Este método também é amplamente utilizado em problemas inversos, proposto para super-resolução em [34]. Este método é aplicado apenas no problema de um único coeficiente. Os detalhes do método estão na seção 5.1.
- **L-TL** - Método baseado na Curva-L, no qual a escolha é o ponto tangente a uma reta com decaimento negativo, conforme [77]. Este método nunca foi utilizado em super-resolução, mas pode ser uma alternativa devido ao baixo custo computacional. Ele é aplicado nos três algoritmos de super-resolução. Os detalhes do método estão na seção 5.1.

- **L-MD** - Método baseado na Curva-L, no qual a escolha é o ponto de mínima distância da origem, conforme [78]. Ele também nunca foi utilizado em super-resolução, mas pode ser uma alternativa devido ao baixo custo computacional e possibilidade de utilização com múltiplos coeficientes. Esse método é aplicado nos três algoritmos de super-resolução. Os detalhes do método estão na seção 5.1.
- **K-HE** - Método desenvolvido inicialmente em [90] para restauração de imagens e posteriormente expandido para super-resolução em [38]. Este método é aplicado nos dois algoritmos SR tradicionais, de um coeficiente e de múltiplos coeficientes.
- **EVID** - Método estatístico Bayesiano *Evidence*, proposto em [79] para restauração de imagens (um coeficiente), e posteriormente expandido em [36] para super-resolução (múltiplos coeficientes). Este método é aplicado nos dois algoritmos SR tradicionais, de um coeficiente e de múltiplos coeficientes.
- **PROP-1** - Método proposto no Capítulo 5 com minimização alternada, usando GC para encontrar a imagem HR e Aproximações Sucessivas para atualizar o coeficiente. O método proposto é aplicado nos três algoritmos de super-resolução.
- **PROP-2** - Método proposto no Capítulo 5 com minimização direta, usando GC-NL, encontrando a imagem HR e o coeficiente conjuntamente. O método proposto é aplicado nos três algoritmos de super-resolução.

Todos esses métodos são iterativos, com exceção do K-HE, que usa minimização do Gradiente, todos os outros utilizam o Gradiente Conjugado para encontrar as imagens HR. As mesmas condições iniciais foram utilizadas em todos: a imagem HR inicial é uma imagem nula, e o coeficiente inicial foi sorteado aleatoriamente de uma densidade uniforme entre 10^{-6} e 10^6 .

Os métodos implementados foram comparados com os resultados obtidos pelos seguintes coeficientes predeterminados:

- **KNOWN** - o resultado da super-resolução usando os coeficientes recomendados pela estimativa MAP. A variância do ruído é conhecida, a variância do erro de movimento e a variância da derivada das imagens foram calculadas a partir das imagens HR originais. Este resultado é utilizado apenas como referência. Nas situações práticas ele não pode ser utilizado porque as imagens HR originais não são conhecidas.

- **INTERP** - o resultado da super-resolução em que os coeficientes seriam obtidos através da máxima verossimilhança dos hiperparâmetros. Contudo, ao invés das imagens HR originais, imagens interpoladas foram utilizadas. Esta forma assume que a imagem interpolada pode ser uma boa substituta da imagem HR original para o cálculo do coeficiente. Este resultado também é utilizado como referência, podendo ser implementado na prática, mas o resultado geralmente é ruim.

Esses resultados servem como referência. Se um determinado método obtiver uma qualidade próxima, ou superior, da qualidade do resultado KNOWN, então a qualidade obtida é considerada suficientemente boa. O resultado INTERP é, geralmente, bem inferior ao KNOWN; logo, se algum método obtiver uma qualidade inferior ao INTERP, então este método pode ser considerado inaceitável para o problema de super-resolução.

A Tabela 6.1 mostra em quais algoritmos de super-resolução os métodos de determinação dos coeficientes são aplicados.

Tabela 6.1: Listagem dos Métodos utilizados em cada algoritmo SR

Algoritmo SR Método	Tradicional de Um Coeficiente	Tradicional de Múltiplos Coeficientes	Simultâneo com 2 Coeficientes
GCV	sim	não	não
L-MC	sim	não	não
L-TN	sim	sim	sim
L-MD	sim	sim	sim
K-HE	sim	sim	não
EVID	sim	sim	não
PROP-1	sim	sim	sim
PROP-2	sim	sim	sim

6.2 Experimentos com Super-Resolução Tradicional de um Coeficiente

6.2.1 Equacionamento do Algoritmo SR e Objetivo do Coeficiente

O algoritmo de super-resolução tradicional, com um único coeficiente, é dado pela seguinte equação:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_2^2 \quad (6.1)$$

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \|\mathbf{g} - \mathbf{C}_k \mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_2^2$$

Sendo $\mathbf{g} = [\mathbf{g}_1^T \dots \mathbf{g}_L^T]^T$ e $\mathbf{C}_k = [\mathbf{C}_{1,k} \dots \mathbf{C}_{L,k}]^T$.

Neste método considera-se que todas as ponderações $\alpha_{j,k}$ são iguais e unitárias, o que deixa implícito

que não se esperam erros de movimento ou *outliers*. Considera-se, apenas, ruído de aquisição igual em todos os quadros LR. Logo, a existência de *outliers* ou erros de movimento na seqüência de imagens poderá causar distorções nos resultados. Nestas simulações, a detecção de *outliers* foi desconsiderada para observar a influência destes na determinação do coeficiente.

Neste experimento, o coeficiente de regularização, λ_k deverá proporcionar a suavização adequada, tanto para evitar a amplificação exagerada do ruído de aquisição quanto para minimizar as fortes distorções locais causadas pelos *outliers*.

6.2.2 Resultados

Qualidade e Estabilidade dos Métodos

A qualidade das imagens estimadas com o coeficiente encontrado pelos diferentes métodos é comparada nas tabelas 6.2 e 6.3. Alguns resultados visuais estão nas figuras 6.3, 6.4 e 6.5.

Nas tabelas 6.2 e 6.3 são observados os resultados médios de 20 realizações, cada realização com um ruído diferente, para cada seqüência. A média da SNR e o desvio-padrão da SNR, de todas realizações e em todas seqüências, também são mostrados nas tabelas.

Pode-se observar que a qualidade obtida pelo método proposto é similar a dos métodos clássicos, como GCV, L-MC e EVID. Nenhum método foi predominante em todos os casos, mas, na média, o método proposto ficou entre os melhores.

Tabela 6.2: Qualidade, em SNR (dB), dos Métodos para SR Tradicional de um Coeficiente com **Alto Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média	Desvio Padrão
GCV	22.2	23.9	19.7	17.6	25.0	15.6	20.8	1.8
L-MC	21.2	22.1	19.0	17.2	26.6	15.3	20.2	1.6
L-TN	5.4	22.4	11.0	7.1	5.7	13.9	10.9	10.1
L-MD	21.2	17.8	14.9	8.9	18.7	10.2	15.3	9.5
K-HE	20.3	19.3	18.2	17.0	21.8	14.4	18.5	0.8
EVID	22.1	22.2	19.8	17.7	25.1	14.4	20.2	1.1
KNOWN	19.8	18.1	17.8	15.7	25.2	15.1	18.6	0.8
INTERP	17.1	18.0	16.3	16.1	27.2	15.0	18.3	1.0
PROP-1	21.2	23.1	19.6	17.3	28.2	15.7	20.8	1.0
PROP-2	21.2	23.1	19.6	17.3	28.2	15.7	20.8	1.0

Outro ponto importante é a estabilidade, que pode ser observada através do desvio-padrão da SNR obtido pelos métodos. Quase todos os métodos apresentaram uma estabilidade adequada, com exceção

Tabela 6.3: Qualidade, em SNR (dB), dos Métodos para SR Tradicional de um Coeficiente com **Baixo Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média	Desvio Padrão
GCV	29.9	30.7	19.7	17.4	23.1	15.8	22.7	2.6
L-MC	30.0	31.1	18.6	17.2	27.7	15.4	23.3	1.4
L-TN	6.4	8.8	5.2	8.1	3.4	12.8	7.9	12.2
L-MD	25.4	30.7	16.2	5.2	20.3	10.3	18.0	10.6
K-HE	26.3	26.9	18.2	16.4	21.8	14.4	20.7	1.0
EVID	30.1	30.6	20.5	18.0	25.3	14.4	23.1	1.5
KNOWN	30.1	31.0	19.6	16.6	25.3	16.9	23.3	0.9
INTERP	22.9	24.5	18.8	15.6	27.0	15.1	20.6	1.1
PROP-1	28.7	30.3	20.3	17.6	28.4	15.7	23.5	1.3
PROP-2	28.7	30.3	20.3	17.6	28.4	15.7	23.5	1.3

do L-TN e L-MD. Eventualmente os métodos L-TN e L-MD divergiram, por isso seu desempenho foi prejudicado.

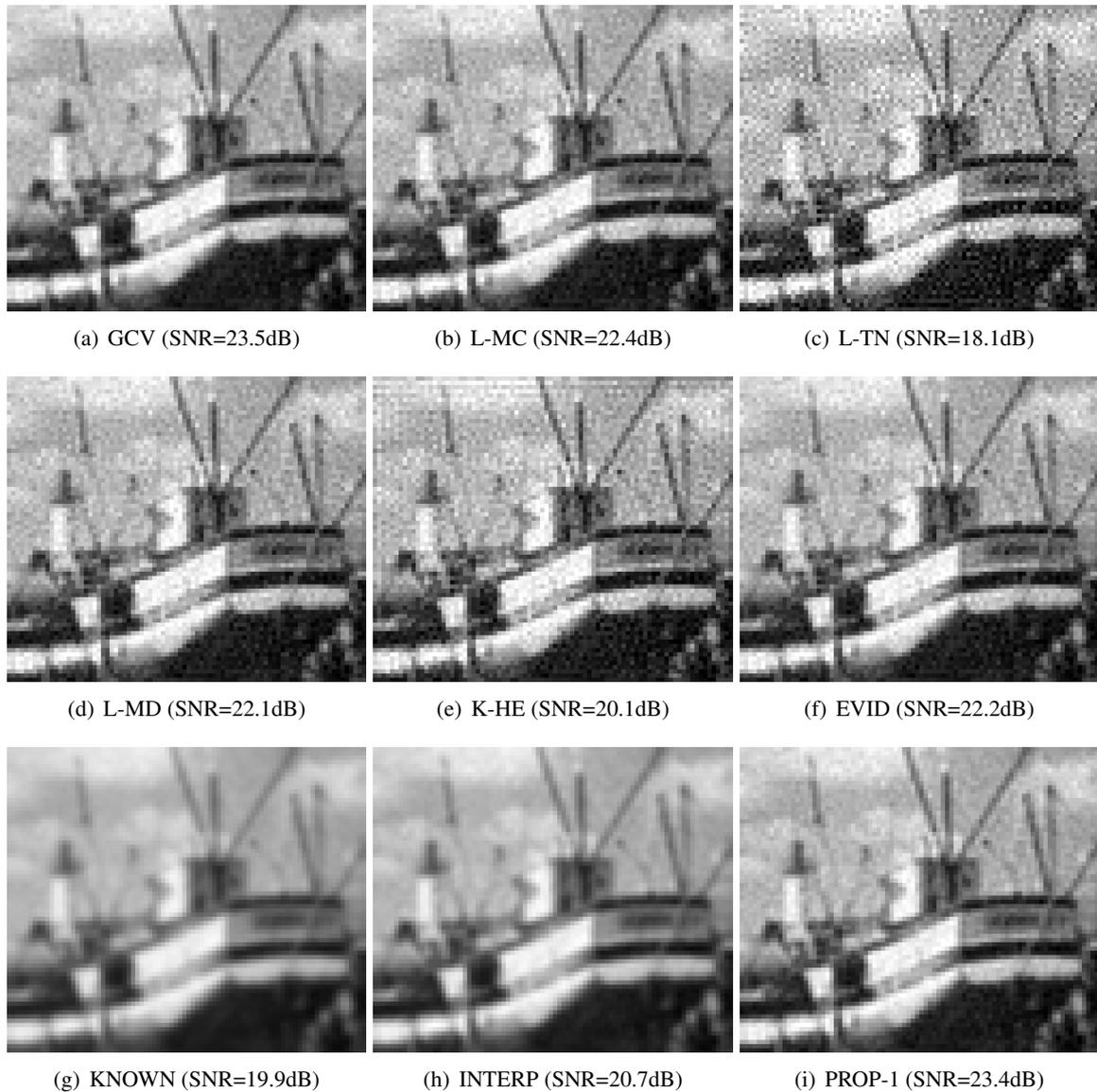


Figura 6.3: Exemplo visual mostrando a imagem 2 da seqüência Barco, com alto ruído.

Na Figura 6.3 todos os resultados visuais são aceitáveis, apesar da pequena diferença na qualidade objetiva. Neste exemplo, em particular, nenhum método divergiu. Nesta seqüência, o principal problema é a amplificação excessiva do ruído de aquisição, dado que não ocorrem *outliers* nas imagens. Observe que o desempenho do método proposto foi similar ao dos melhores métodos disponíveis.

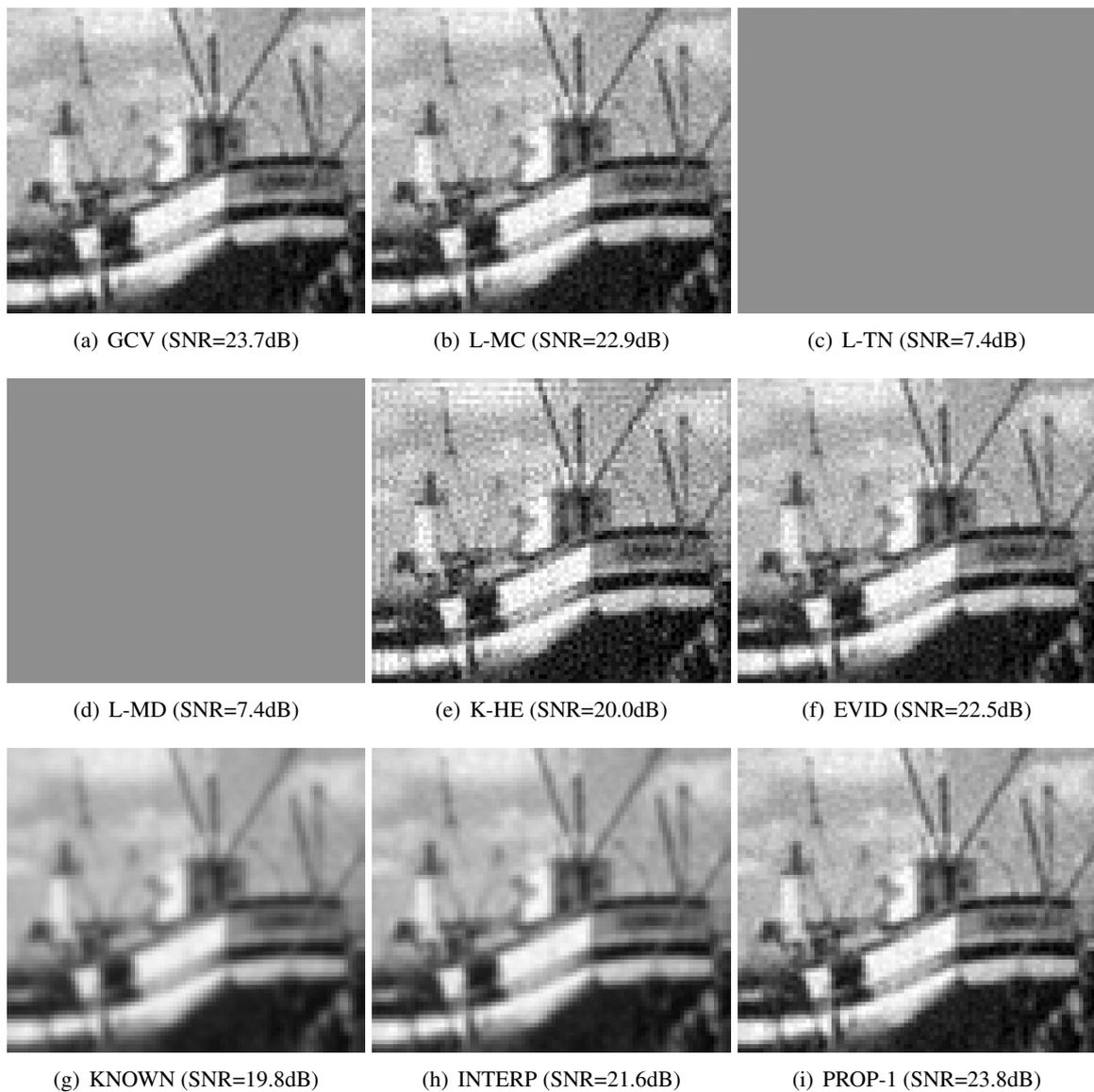


Figura 6.4: Exemplo visual mostrando a imagem 2 da seqüência Barco em outra realização, com alto ruído.

Na Figura 6.4 uma outra realização é apresentada com o mesmo nível de ruído. Nesta figura, os métodos L-TN e L-MD divergiram. Observe que as imagens resultantes têm apenas a componente constante. Este exemplo ilustra a necessidade de estabilidade na estimativa.

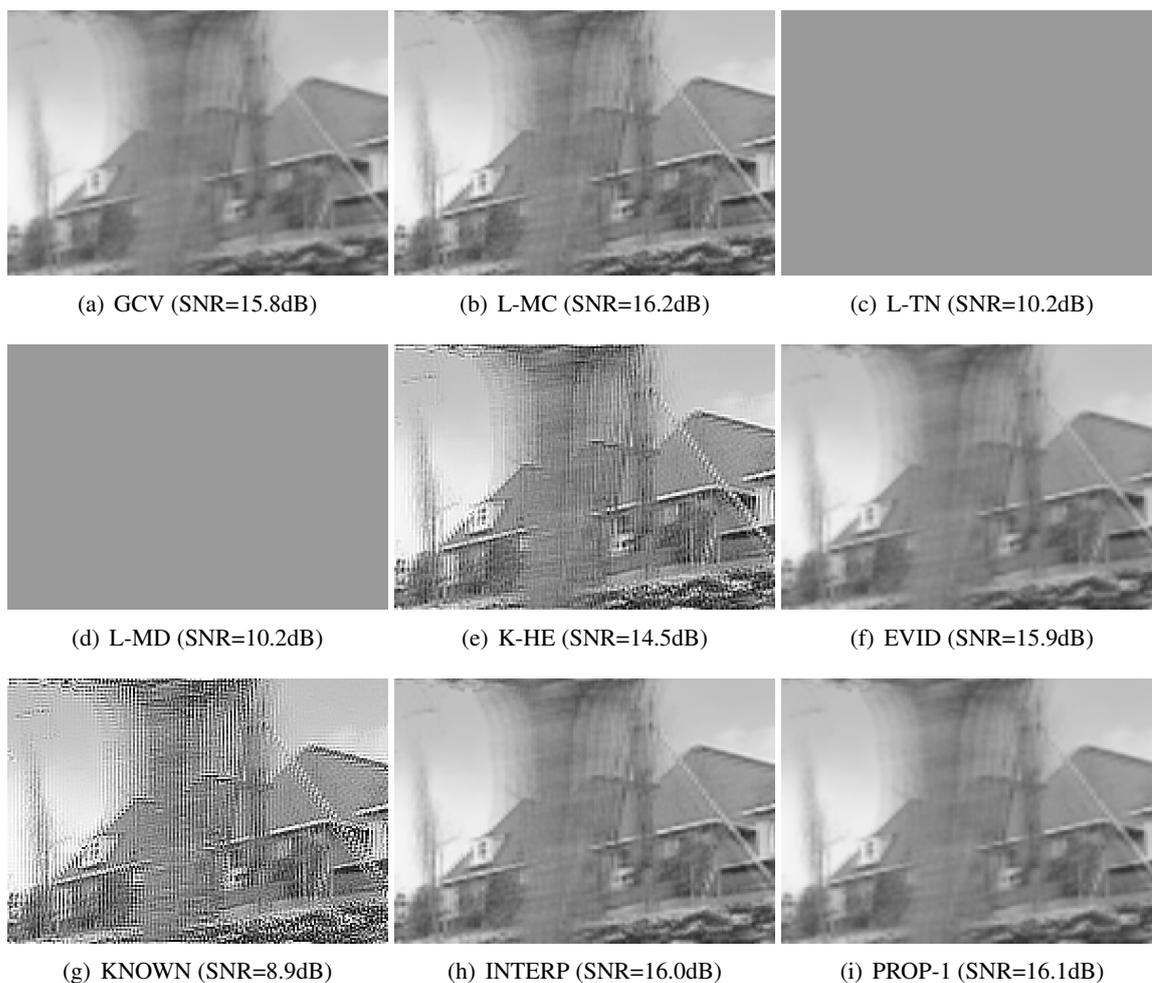


Figura 6.5: Exemplo visual mostrando a imagem 2 da seqüência Jardim, com baixo ruído.

Pode-se observar na Figura 6.5 o efeito dos *outliers* na seqüência. Neste caso, o coeficiente de regularização visa à suavização das imagens de forma a minimizar a forte amplificação das distorções locais. Observa-se que são nos contornos da árvore, a principal fonte de *outliers* dessa cena, que ocorrem as distorções mais fortes.

Na Figura 6.5, nota-se que os métodos instáveis, L-TN e L-MD, divergiram, produzindo imagens com apenas a componente constante. O resultado KNOWN teve o erro causado pelos *outliers* amplificado, o que ocorreu porque a escolha do coeficiente, neste caso, não assume a presença de *outliers*, apenas do ruído de aquisição. Contudo, quase todos os métodos conseguiram contornar o problema aumentando a suavidade, produzindo imagens com aspecto borrado nas regiões em torno da árvore, em vez de fortes oscilações.

Uma forma de reduzir este tipo de problema no algoritmo de super-resolução tradicional é através da utilização de ponderações para cada imagem LR (ou utilizando algoritmos robustos). Esta estratégia

existe no algoritmo tradicional com múltiplos coeficientes.

Custo Computacional dos Métodos

Nas tabelas 6.4 e 6.5 a complexidade computacional dos métodos é apresentada, em segundos, com o tempo necessário para a convergência dos métodos. O tempo relativo médio também é apresentado.

Tabela 6.4: Custo computacional, em segundos, dos Métodos para SR Tradicional de Um Coeficiente com **Alto Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
GCV	218.7	119.2	172.3	308.1	663.1	1820.1	63.9
L-MC	171.4	137.5	339.7	379.1	679.0	786.7	58.0
L-TN	30.1	56.2	37.1	78.1	70.2	84.1	7.2
L-MD	45.5	270.7	191.3	531.2	279.5	801.1	55.8
K-HE	47.6	86.2	66.5	98.6	225.9	366.3	21.3
EVID	145.7	79.8	248.2	255.1	605.8	1320.0	52.1
KNOWN	3.1	2.1	5.8	7.0	11.9	13.5	1.0
INTERP	4.5	3.6	6.1	8.5	10.1	13.7	1.2
PROP-1	23.7	13.6	30.1	56.1	80.2	98.6	6.9
PROP-2	3.2	4.7	5.1	5.3	9.6	9.8	1.1

Tabela 6.5: Custo computacional, em segundos, dos Métodos para SR Tradicional de Um Coeficiente com **Baixo Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
GCV	248.1	259.6	182.7	263.4	843.2	1320.1	77.6
L-MC	195.0	478.4	261.8	354.1	678.3	864.6	84.9
L-TN	32.1	59.8	27.3	65.9	84.9	120.1	11.4
L-MD	89.0	130.4	47.9	539.6	3209.0	2559.7	123.5
K-HE	94.8	180.9	186.9	106.4	222.9	381.4	35.5
EVID	93.0	99.6	276.5	294.1	625.8	919.0	51.5
KNOWN	3.3	2.3	5.4	8.8	9.1	10.9	1.0
INTERP	3.5	5.8	5.9	7.8	22.5	20.0	1.6
PROP-1	46.1	41.75	33.9	49.8	79.3	88.7	10.1
PROP-2	4.1	6.8	6.2	9.9	10.9	9.4	1.1

Os resultados das tabelas 6.4 e 6.5 ilustram o grande potencial do método proposto. O custo computacional do método proposto PROP-1, que utiliza iterações alternadas entre a Aproximação Sucessiva e o Gradiente Conjugado, é bem baixo, comparável ao método L-TN, que também tem baixo custo computacional. Além disso, o método PROP-2, que utiliza minimização direta através do gradiente conjugado não-linear, foi mais rápido ainda. O custo computacional do método PROP-2 é comparável ao custo para

se obter a estimativa usando o coeficiente fixo, como o KNOWN e o INTERP. Estes resultados ilustram o baixo custo computacional obtido pelo método proposto.

Uma observação importante, notada nos experimentos, é que o método L-MD não demonstrou ter baixo custo computacional nesse experimento. Isso ocorreu em razão das divergências e a fortes mudanças no coeficiente estimado em cada iteração, por exemplo, variando de um coeficiente muito grande para um muito pequeno e vice-versa. A velocidade de convergência do gradiente conjugado, usado por este método para encontrar a imagem HR, é prejudicada, porque a nova imagem HR é muito diferente da imagem HR encontrada na iteração anterior.

6.3 Experimentos com Super-Resolução Tradicional de Múltiplos Coeficientes

6.3.1 Equacionamento do Algoritmo SR e Objetivo dos Coeficientes

O algoritmo de super-resolução tradicional com múltiplos coeficientes é dado pela seguinte equação:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{f}_k} \sum_{j=1}^L \alpha_{j,k} \|\mathbf{g}_j - \mathbf{C}_{j,k} \mathbf{f}_k\|_2^2 + \lambda_k \|\mathbf{R}_k \mathbf{f}_k\|_2^2 \quad (6.2)$$

Neste método, tanto as ponderações $\alpha_{j,k}$ quanto o coeficiente de regularização λ_k precisam ser determinados. Neste experimento, o coeficiente de regularização deverá proporcionar a suavização adequada para evitar a amplificação exagerada dos erros, enquanto as ponderações devem reduzir a influência das imagens LR que contêm *outliers* para minimizar os erros causados por estes. Nestas simulações, a detecção de *outliers* também foi desconsiderada.

6.3.2 Resultados

Qualidade e Estabilidade dos Métodos

A qualidade, em SNR (dB), das seqüências estimadas com os coeficientes encontrados é mostrada nas tabelas 6.6 e 6.7. Nas figuras 6.6, 6.7 e 6.8 são apresentados alguns resultados visuais.

Pode-se observar nas tabelas 6.6 e 6.7 que a qualidade obtida pelo método proposto é similar à dos melhores métodos. Neste caso, o resultado KNOWN, com os coeficientes conhecidos proporcionou os melhores resultados médios. Os resultados do método proposto foram similares ao KNOWN.

Tabela 6.6: Qualidade, em SNR (dB), dos Métodos para SR Tradicional de Múltiplos Coeficientes com **Alto Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média	Desvio Padrão
L-TN	16.8	8.6	18.4	17.7	14.7	15.2	15.3	6.1
L-MD	14.6	13.0	17.7	16.0	24.7	17.2	17.2	3.8
K-HE	17.8	16.3	16.5	9.6	22.4	16.6	16.5	1.9
EVID	21.8	22.0	19.7	15.3	28.5	17.9	20.8	1.6
KNOWN	22.4	23.3	19.8	18.4	29.8	18.8	22.0	0.6
INTERP	13.5	14.3	15.8	10.4	18.1	16.0	14.7	0.6
PROP-1	21.8	22.4	21.5	19.6	32.4	18.6	22.7	0.9
PROP-2	21.8	22.4	21.5	19.6	32.4	18.6	22.7	0.9

Tabela 6.7: Qualidade, em SNR (dB), dos Métodos para SR Tradicional de Múltiplos Coeficientes com **Baixo Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média	Desvio Padrão
L-TN	25.1	22.8	16.3	18.8	10.8	16.4	18.4	6.2
L-MD	27.6	28.2	21.5	20.0	23.4	17.9	23.1	5.0
K-HE	18.6	18.9	18.7	8.6	22.6	16.4	17.3	2.2
EVID	30.5	30.7	21.2	16.1	26.5	17.0	23.7	2.5
KNOWN	30.1	31.2	23.7	22.6	33.0	19.3	26.6	0.5
INTERP	13.6	15.1	13.6	10.4	18.0	16.0	14.5	0.6
PROP-1	28.9	29.6	23.7	21.4	34.1	19.0	26.1	1.6
PROP-2	28.9	29.6	23.7	21.4	34.1	19.0	26.1	1.6

A estabilidade do método proposto foi superior à dos demais métodos, superando os métodos EVID e K-HE. Isso pode ser observado através das medida do desvio-padrão da SNR. Os métodos L-TN e L-MD obtiveram o pior desempenho em termos de estabilidade.

A Figura 6.6 ilustra o desempenho dos métodos na escolha das ponderações para evitar as distorções causadas pelos *outliers*. Neste exemplo, alguns métodos não privilegiaram a imagem LR que está no mesmo instante temporal da imagem HR, como é claramente observado nos resultados dos métodos L-TN, L-MD, K-HE e EVID. Observe que o círculo preto não é predominante nas imagens resultantes, como acontece no resultado KNOWN. Pode-se observar, também, que as ponderações escolhidas, apesar de atenuarem significativamente as distorções, não conseguiram eliminar completamente as distorções causadas pelos *outliers* em nenhum dos métodos, neste exemplo.

Nas figuras 6.7 e 6.8 são mostrados os resultados de duas imagens HR consecutivas, em uma mesma realização. Este exemplo mostra que o método proposto foi bastante estável, proporcionando praticamente o mesmo nível de ponderação nos dois quadros. Observe que a instabilidade dos métodos L-TN

e L-MD fizeram com que dois quadros consecutivos tivessem um nível de suavidade completamente diferente, prejudicando o resultado total. Apesar disso, o resultado da Figura 6.8 mostra que os métodos L-TN e L-MD obtiveram um bom desempenho neste quadro em particular.

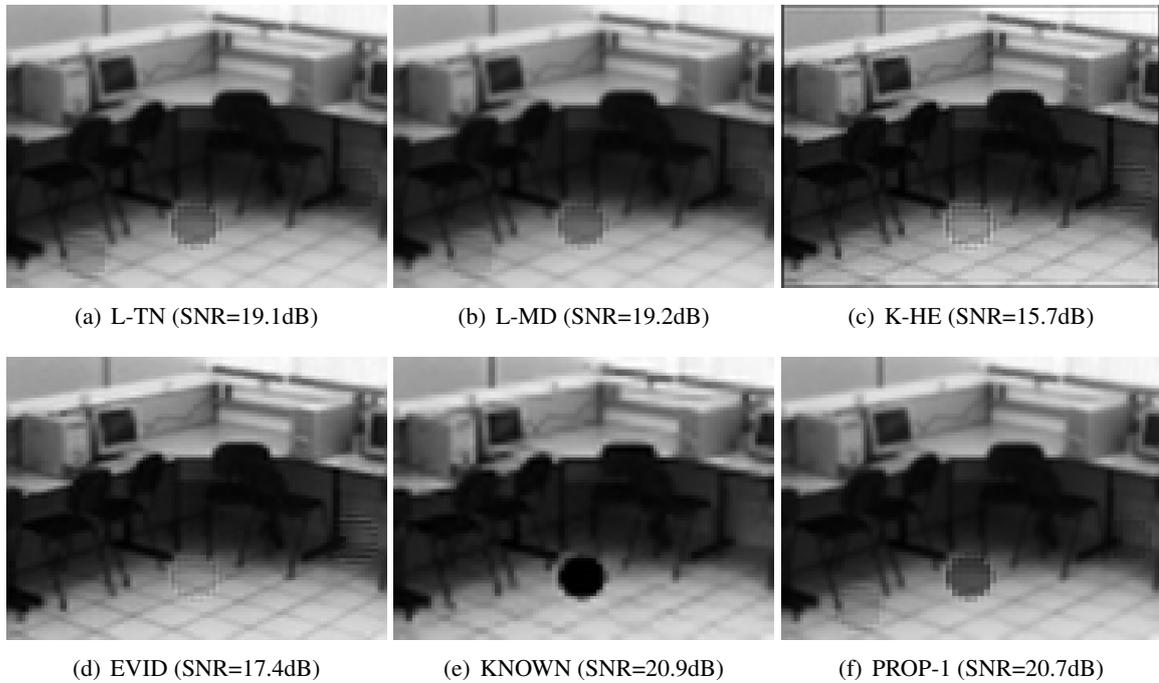


Figura 6.6: Exemplo visual mostrando a imagem 2 da seqüência Sala, com baixo ruído.

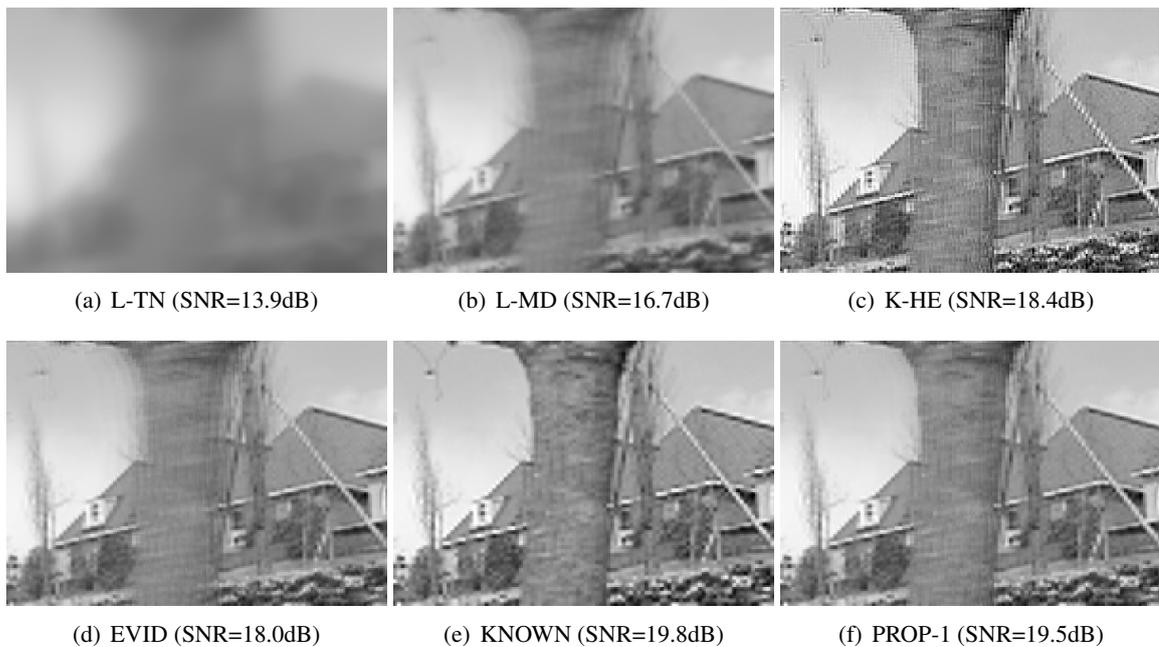


Figura 6.7: Exemplo visual mostrando a imagem 2 da seqüência Jardim, com baixo ruído.

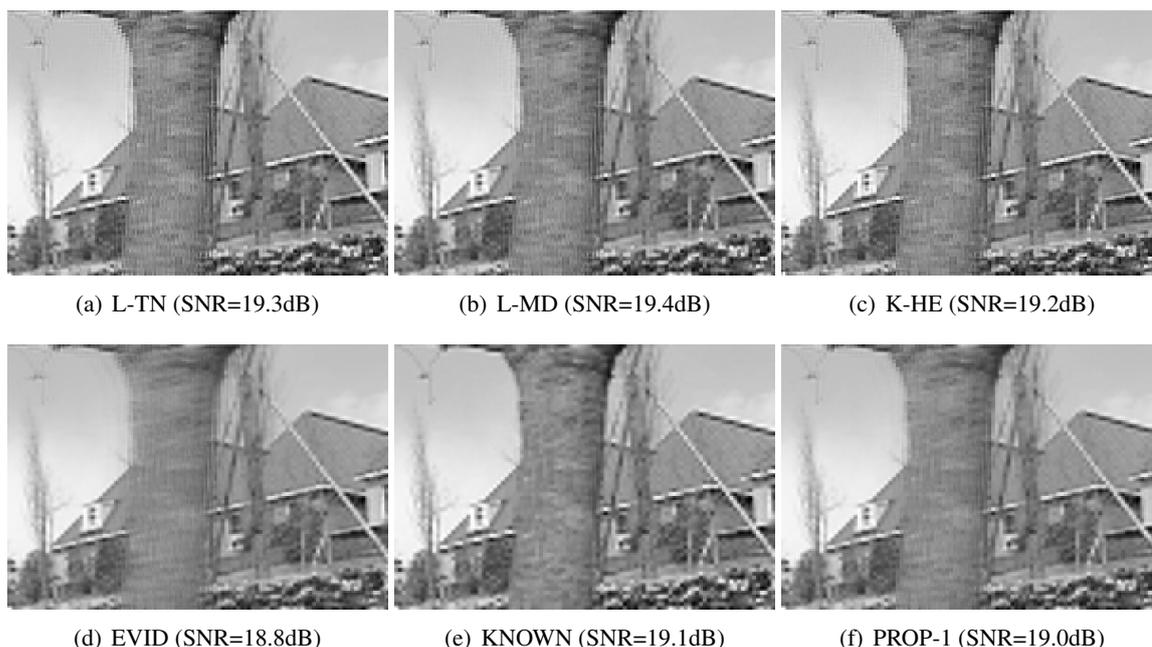


Figura 6.8: Exemplo visual mostrando a imagem 3 da seqüência Jardim, com baixo ruído.

Custo Computacional dos Métodos

Nas tabelas 6.8 e 6.9 o custo computacional dos métodos é contrastado através do tempo, em segundos, necessário para a convergência dos métodos.

Os resultados numéricos das tabelas 6.8 e 6.9 ilustram o baixo custo computacional do método proposto. O desempenho do método proposto foi superior ao do método K-HE, recentemente apresentado como um método de baixo custo para o algoritmo tradicional com múltiplos coeficientes [38].

Tabela 6.8: Custo computacional, em segundos, dos Métodos para SR Tradicional de Múltiplos Coeficientes com **Alto Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
L-TN	237.3	204.1	87.7	112.7	799.5	1504.9	66.5
L-MD	143.5	452.1	99.8	283.9	426.4	500.5	41.6
K-HE	58.6	69.4	111.9	60.6	264.7	375.2	20.6
EVID	208.9	108.8	184.7	1435.1	947.7	587.5	90.2
KNOWN	3.0	2.6	4.9	6.0	8.7	13.3	1.0
INTERP	10.1	5.0	18.8	30.2	18.9	28.5	3.1
PROP-1	53.0	65.7	51.6	69.6	124.0	154.1	15.1
PROP-2	4.1	8.4	4.9	6.5	18.1	25.1	1.8

Uma observação importante, verificada nestes experimentos, é que os métodos L-TN e L-MD obtiveram um custo computacional maior do que o esperado em algumas situações. Isso ocorreu em virtude

das dificuldades de convergência na busca da solução inerentes a esses métodos.

Tabela 6.9: Custo computacional, em segundos, dos Métodos para SR Tradicional de Múltiplos Coeficiente com **Baixo Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
L-TN	226.2	129.0	66.2	912.5	233.6	1100.2	58.6
L-MD	142.3	284.1	75.9	427.7	517.1	411.1	48.9
K-HE	41.0	120.6	55.6	59.4	250.0	502.4	21.6
EVID	874.9	281.8	220.8	3302.0	2862.9	591.7	190.0
KNOWN	3.0	2.6	7.8	7.4	12.6	14.4	1.0
INTERP	9.7	14.0	19.4	24.6	28.0	27.4	3.1
PROP-1	59.4	47.8	113.9	150.0	205.0	154.3	16.7
PROP-2	8.0	41.8	9.1	12.0	34.6	25.4	4.3

6.4 Experimentos com Super-resolução Simultânea de Dois Coeficientes

6.4.1 Equacionamento do Algoritmo SR e Objetivo dos Coeficientes

O algoritmo de super-resolução simultânea, com dois coeficientes, é dado pela seguinte equação:

$$\hat{f} = \arg \min_f \|g - Df\|_2^2 + \lambda_R \|Rf\|_2^2 + \lambda_M \|Mf\|_2^2 \quad (6.3)$$

sendo que g , f , D , R , e M foram definidos no Capítulo 3.

Neste método, existem dois coeficientes de regularização que precisam ser determinados: o coeficiente de suavidade espacial, λ_R , e o coeficiente de movimento, λ_M . O coeficiente de suavidade espacial deverá controlar o nível de suavidade das imagens HR estimadas; assim, este coeficiente deve atuar de acordo com a quantidade de ruído presente e de acordo com o grau de variação espacial da imagem HR. O coeficiente de movimento deverá controlar a similaridade das imagens da seqüência na trajetória de movimento; assim, este coeficiente deve atuar de acordo com o nível de erro de movimento existente, considerando, por exemplo, os *outliers* existentes na seqüência. A detecção de *outliers* não foi considerada para avaliar o desempenho dos métodos na estimação do coeficiente λ_M .

6.4.2 Resultados

Qualidade e Estabilidade dos Métodos

A qualidade em SNR (dB) das seqüências estimadas com os coeficientes encontrados é mostrada nas tabelas 6.10 e 6.11. Nas figuras 6.9 e 6.10 são mostrados alguns resultados visuais.

Tabela 6.10: Qualidade, em SNR (dB), dos Métodos para SR Simultânea com Dois Coeficientes com **Alto Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média	Desvio Padrão
L-TN	20.6	21.6	21.8	20.3	31.8	19.5	22.6	1.1
L-MD	20.8	21.6	21.0	20.4	31.8	18.8	22.4	0.5
KNOWN	21.6	20.8	20.7	19.3	29.4	18.7	21.7	0.4
INTERP	<i>18.4</i>	<i>19.0</i>	<i>17.6</i>	<i>16.6</i>	<i>27.6</i>	<i>16.7</i>	19.3	0.4
PROP-1	22.7	22.7	21.7	20.6	31.9	19.3	23.1	0.4
PROP-2	22.7	22.7	21.7	20.6	31.9	19.3	23.1	0.4

 Tabela 6.11: Qualidade, em SNR (dB), dos Métodos para SR Simultânea com Dois Coeficientes com **Baixo Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Média	Desvio Padrão
L-TN	21.6	24.4	23.8	22.2	35.0	17.6	24.0	1.2
L-MD	21.6	24.9	23.0	22.1	34.8	19.1	24.1	0.7
KNOWN	30.6	32.1	22.3	19.4	33.4	20.1	26.3	0.5
INTERP	<i>18.4</i>	<i>18.6</i>	18.3	<i>14.9</i>	<i>30.2</i>	<i>16.8</i>	19.5	0.6
PROP-1	28.5	31.0	25.3	21.6	34.9	19.8	26.8	0.6
PROP-2	28.5	31.0	25.3	21.6	34.9	19.8	26.8	0.6

Pode-se observar, nas tabelas 6.10 e 6.11, que a qualidade obtida pelo método proposto é superior à obtida pelos demais métodos. Neste caso, o KNOWN, com o coeficiente calculado utilizando as imagens originais, proporcionou resultados muito bons. O desvio-padrão da SNR, observado nas tabelas 6.10 e 6.11, mostra que o método proposto foi mais estável, embora os métodos L-TN e L-MD tenham se mostrado mais estáveis neste algoritmo de SR simultânea que nos outros dois algoritmos de SR tradicionais.

Até o presente momento não existiam métodos de determinação dos coeficiente desenvolvidos para esta aplicação; por isso, não é possível comparar o desempenho com métodos clássicos. Contudo, pode-se notar, comparando estes resultados com os resultados das tabelas 6.2, 6.3, 6.6 e 6.7, que a combinação da super-resolução simultânea com o método dos coeficientes proposto obteve os melhores resultados numéricos deste capítulo.

Pode-se notar, na Figura 6.9, o desempenho dos métodos de determinação dos coeficientes em super-resolução simultânea. Nota-se que o método L-MD falhou neste exemplo, ao passo que o método proposto teve um resultado superior ao KNOWN. Na Figura 6.10 um outro exemplo é apresentado. Neste caso o método L-TN divergiu e teve uma baixa qualidade. Nota-se que o método proposto proporcionou um resultado superior ao dos demais.

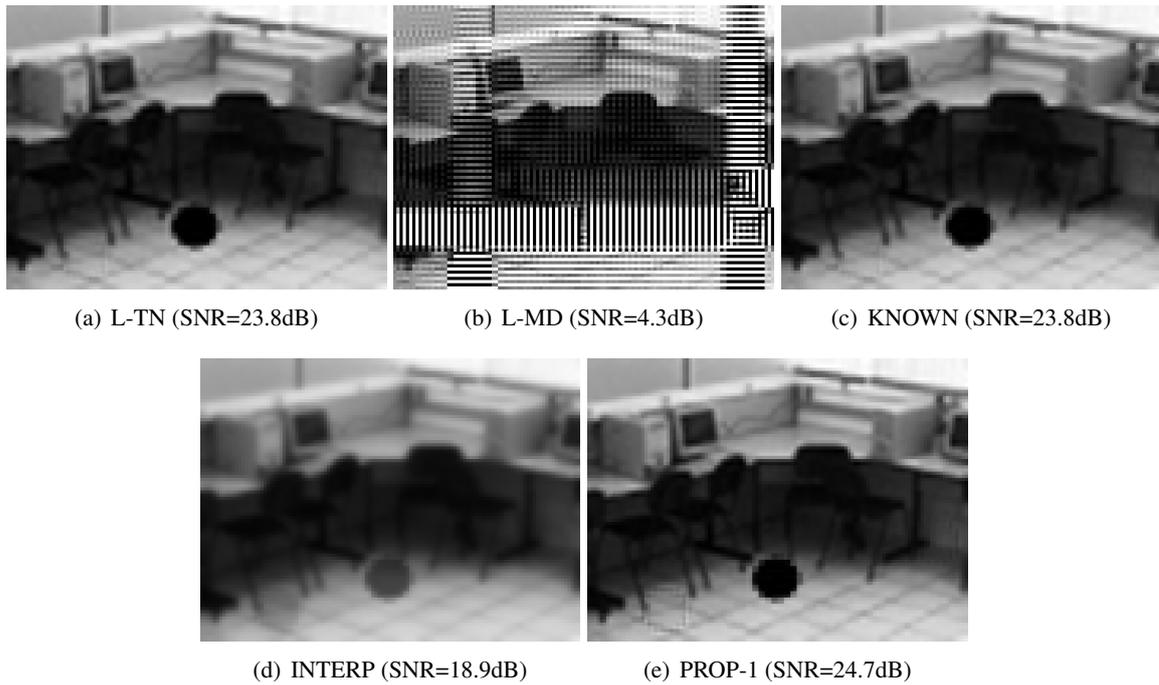


Figura 6.9: Exemplo visual mostrando a imagem 2 da seqüência Sala, com baixo ruído.

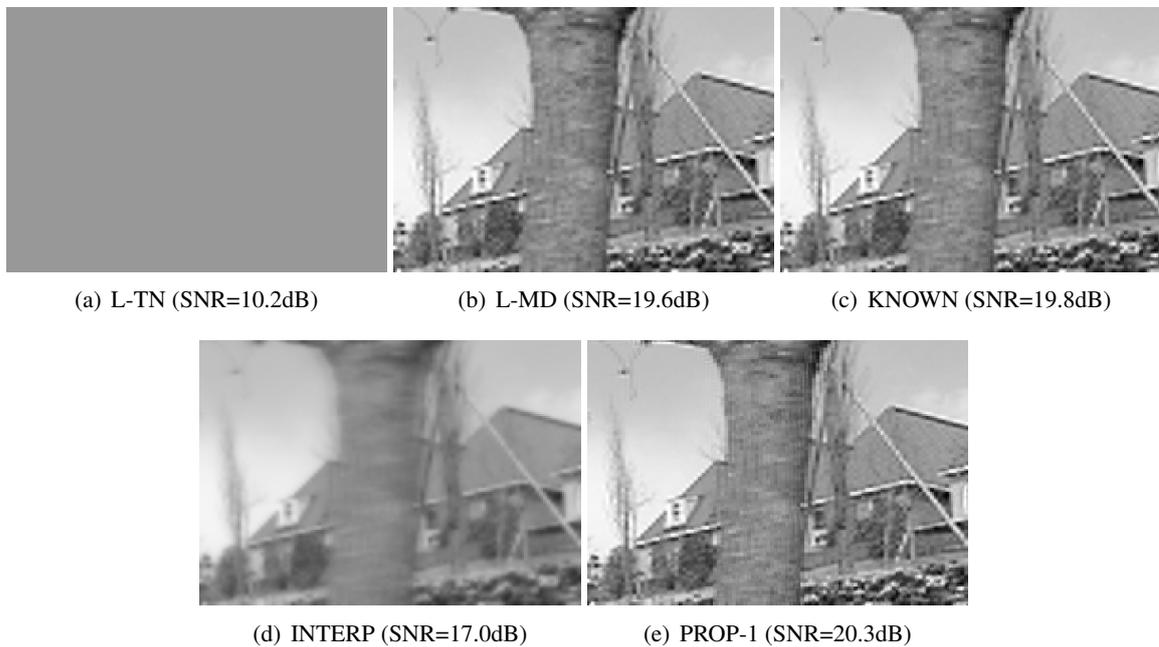


Figura 6.10: Exemplo visual mostrando a imagem 2 da seqüência Jardim, com baixo ruído.

Custo Computacional dos Métodos

Nas tabelas 6.12 e 6.13 o custo computacional dos métodos é contrastado através do tempo, em segundos, necessário para a convergência dos métodos.

Tabela 6.12: Custo computacional, em segundos, dos Métodos para SR Simultânea com Dois Coeficientes com **Alto Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
L-TN	18.6	34.0	89.2	44.4	65.0	73.8	5.8
L-MD	293.1	267.8	2279.0	865.3	996.4	990.5	109.2
KNOWN	6.1	7.9	7.8	8.5	12.8	13.6	1.0
INTERP	5.7	4.3	5.2	6.3	7.5	11.5	0.7
PROP-1	128.4	106.0	156.6	140.1	176.1	178.0	13.0
PROP-2	3.4	17.6	6.5	12.6	18.2	18.4	1.3

 Tabela 6.13: Custo computacional, em segundos, dos Métodos para SR Simultânea com Dois Coeficientes com **Baixo Ruído**

Seqüência Método	Barco	Árvore	Sala	Estrada	Jato	Jardim	Relativo
L-TN	30.2	29.5	47.0	56.3	63.2	84.4	3.9
L-MD	356.6	420.7	1330.0	896.0	934.9	1031.6	67.2
KNOWN	7.9	13.8	8.5	13.6	19.8	17.7	1.0
INTERP	6.9	10.5	4.5	5.7	7.5	10.5	0.6
PROP-1	130.4	128.1	102.1	221.3	119.0	206.8	12.0
PROP-2	31.0	8.9	25.5	29.4	22.0	21.6	2.0

Os resultados numéricos das tabelas 6.12 e 6.13 ilustram o baixo custo computacional do método proposto. Novamente, os resultados mostram que método PROP-2 tem custo computacional similar ao custo da minimização usando os coeficientes fixos, como o KNOWN e INTERP.

Nas tabelas 6.12 e 6.13, pode-se notar que o método L-TN teve baixo custo computacional, bem inferior ao método L-MD e, inclusive, inferior ao PROP-1.

Conclusão do Capítulo

Neste capítulo foram apresentados resultados experimentais comparando os métodos existentes de determinação dos coeficientes de regularização com o método proposto. Estes métodos foram testados em três algoritmos de super-resolução: o tradicional com um único coeficiente, o tradicional com múltiplos coeficientes e o simultâneo com dois coeficientes. Os experimentos mostraram que a qualidade obtida pelo método proposto é tão boa quanto a obtida pelos melhores métodos clássicos. Contudo, o custo computacional do método proposto é bastante inferior ao dos métodos clássicos. Além disso, o método proposto mostrou-se bastante estável.

Capítulo 7

Conclusões e Linhas de Trabalhos Futuros

7.1 Conclusões

Esta tese apresentou duas contribuições importantes. A primeira é uma nova classe de algoritmos de super-resolução simultâneos, cujo desenvolvimento teórico é apresentado no Capítulo 3 e os experimentos são apresentados no Capítulo 4. Esta classe de algoritmos proposta foi desenvolvida baseada nos algoritmos regularizados de super-resolução, fundamentados tanto nos métodos estatísticos MAP quanto nos métodos determinísticos com Regularização Generalizada de Tikhonov.

A segunda contribuição é um novo método de determinação dos coeficientes de regularização, cujo desenvolvimento teórico é apresentado no Capítulo 5 e os experimentos são apresentados no Capítulo 6. Este novo método é fundamentado, principalmente, na teoria de estimação estatística JMAP, uma generalização do MAP. Contudo, o método desenvolvido é bastante similar a alguns métodos determinísticos baseados na Curva-L [77].

7.1.1 Conclusões Sobre o Método de Super-Resolução Simultânea

A principal característica desta nova classe de algoritmos de SR simultânea proposta é a utilização do movimento exclusivamente como informação *a priori*. As análises e os experimentos realizados mostraram que o algoritmo simultâneo proposto mantém a mesma qualidade de estimativa que o algoritmo simultâneo existente [1]. Além disso, o método simultâneo desenvolvido nesta tese proporciona os seguintes avanços em relação ao método simultâneo existente: redução do custo computacional e aumento da robustez aos *outliers*.

Os resultados experimentais apresentados no Capítulo 4 mostram o bom desempenho em termos

de qualidade dos resultados, custo computacional e robustez da estimativa. Em geral, a qualidade dos resultados do método proposto é similar à obtida pelo método simultâneo existente. A complexidade computacional do método proposto é menor que a do método simultâneo existente, ficando próxima da complexidade dos métodos tradicionais. Além disso, o método proposto, utilizando penalidade Huber no termo de movimento, é mais robusto que o método simultâneo existente com as mesmas penalidades no termo de movimento. A robustez do método proposto, neste caso, é similar à robustez do método tradicional robusto proposto em [75]; contudo, a qualidade da estimativa obtida pelo método proposto robusto é maior que a obtida método tradicional robusto de [75].

7.1.2 Conclusões Sobre o Método de Determinação dos Coeficientes de Regularização

O método proposto para determinação dos coeficientes de regularização existente foi desenvolvido fundamentado na teoria JMAP. O método JMAP existente era instável, mas apresentava baixo custo computacional e boa qualidade na estimativa. A partir disso, foi proposta uma versão estável do JMAP, utilizando densidade Gama nos hiperparâmetros, que manteve a mesma qualidade de estimativa e baixo custo computacional do JMAP existente.

Dois métodos rápidos de otimização foram propostos para o problema, usando um método iterativo alternado, que utiliza Gradiente Conjugado e Aproximações Sucessivas, e outro método de minimização direta, que utiliza Gradiente Conjugado Não-Linear. O método foi desenvolvido, inicialmente, para o algoritmo de super-resolução tradicional com um único coeficiente e , posteriormente, estendido para os algoritmos tradicionais de múltiplos coeficientes e para o algoritmo simultâneo proposto nesta tese, com dois parâmetros.

Os experimentos realizados no Capítulo 6, comparando os métodos existentes de determinação dos coeficientes de regularização com o método proposto, demonstraram o bom desempenho do método proposto. Os experimentos realizados mostraram que a qualidade obtida pelo método proposto é tão boa quanto a obtida pelos melhores métodos clássicos; além disso, o custo computacional do método proposto é bastante inferior à maioria dos métodos clássicos. Comparado com outros métodos de baixo custo computacional, o método proposto mostrou-se muito mais estável. O método proposto apresentou um bom desempenho nos três algoritmos de SR testados: o tradicional com um coeficiente, tradicional com múltiplos coeficientes e no simultâneo proposto, com dois coeficientes.

7.2 Linhas de Trabalhos Futuros

Os trabalhos desenvolvidos nesta tese abriram novas questões para futuras investigações. Isso ocorreu tanto na nova classe de algoritmos de super-resolução simultânea desenvolvida quanto no novo método de determinação dos coeficientes de regularização. Assim, as linhas de trabalhos futuros serão listados de acordo com o trabalho precedente.

7.2.1 Linhas de Trabalhos Futuros em Super-Resolução Simultânea

- Estudo da velocidade de convergência dos algoritmos simultâneos, com o objetivo de desenvolver um pré-condicionamento adequado para acelerar a velocidade de convergência.
- Aprimoramento do algoritmo de otimização (Gradiente Conjugado Não-Linear ou outros) com penalidade Huber, com o objetivo de reduzir o custo computacional.
- Estudo de diferentes modelos para a matriz M usada na penalidade na trajetória de movimento, que são vistos no Apêndice B.2. Este estudo teria o objetivo de avaliar o potencial de ganho na utilização de modelos de maior ordem.
- Desenvolvimento e implementação dos algoritmos para ruídos de aquisição Não-Gaussianos, os quais são vistos no Apêndice B.1.
- Estudo e desenvolvimento de métodos seqüenciais de super-resolução, com o objetivo de obter uma qualidade superior ao tradicional e próxima dos métodos simultâneos, mas necessitando de menos memória para implementação.

7.2.2 Linhas de Trabalhos Futuros na Determinação dos Coeficientes

- Desenvolvimento do método proposto para aplicação com normas diferentes da norma ℓ_2 ao quadrado, principalmente, objetivando os algoritmos de SR robustos, com normas Huber e ℓ_1 .
- Análise do erro de estimativa no caso de múltiplos coeficientes, objetivando o aprimoramento da escolha para os μ 's.
- Estudo de outras densidades *a priori* que possam ser assumidas para os hiperparâmetros. Este estudo teria o objetivo de encontrar outras formas estáveis do JMAP.

- Estudo e desenvolvimento de uma extensão do método não apenas para a determinação dos coeficientes de regularização, que estão relacionados aos termos da regularização, mas para a ponderação individual de cada linha das matrizes envolvidas nos termos utilizados na regularização.
- Estudo e desenvolvimento de uma extensão do método não apenas para a determinação das ponderações aplicadas às equações do termo de dados, mas uma ponderação individual de cada linha no termo de dados.

7.2.3 Outras Linhas de Trabalhos Futuros

- Desenvolvimento da estimação conjunta das imagens HR, coeficientes e dos vetores de movimento.

Apêndice A

Complementos do Capítulo 2

Este apêndice apresenta os complementos do Capítulo 2. No Apêndice A.1 é apresentada a discretização do sistema de aquisição; no Apêndice A.2, os modelos contínuo e discreto de movimento; no Apêndice A.3, as formas comuns da matriz \mathbf{R} e, finalmente, no Apêndice A.4 são mostradas as funções de discrepância mais conhecidas para o problema.

A.1 Discretização do Sistema de Aquisição

Nesta seção será apresentado o processo de discretização dos sistemas de aquisição [8, 63, 94], apresentado na seção 2.1.1. O sistema de aquisição é, originalmente, um operador contínuo-discreto, ou seja, ele converte uma imagem de alta resolução (HR) contínua em uma seqüência de dados discretos, que correspondem a imagem de baixa resolução (LR). O sistema contínuo-discreto, sem ruído, pode ser expresso por:

$$g[n_1, n_2] = \int \int D(n_1, n_2, x_1, x_2) f(x_1, x_2) dx_1 dx_2 \quad (\text{A.1})$$

Em (A.1) $g[n_1, n_2]$ é a imagem LR discreta¹, de tamanho $N_1 \times N_2$, as coordenadas n_1 e n_2 representam as posições discretas do pixel na direção horizontal e na direção vertical, respectivamente. $f(x_1, x_2)$ é a imagem HR contínua², x_1 e x_2 são as coordenadas contínuas na direção horizontal e na direção vertical, respectivamente. A resposta do sistema de aquisição é especificada em $D(n_1, n_2, x_1, x_2)$. Fixando n_1 e n_2 , tem-se a descrição da sensibilidade do elemento do sensor na posição $[n_1, n_2]$ em relação às irradiações vindas da imagem HR contínua. A função resultante é chamada de *Função de Aquisição*.

¹usa-se colchetes [] para representar as funções discretas, sendo $n_1, n_2 \in \mathbb{N}$

²usa-se parênteses () para representar as funções contínuas, onde $x_1, x_2 \in \mathbb{R}$

Fixando x_1 e x_2 , tem-se a descrição do espalhamento da irradiação do ponto (x_1, x_2) nos elementos do sensor, essa função é chamada de *Função de Espalhamento de Ponto*.

Para utilizar processamento digital para recuperar a imagem de alta resolução, é necessário discretizar o sistema (A.1). De acordo com [8, 63, 95], o sistema pode ser discretizado através da utilização de um processo de amostragem-reconstrução da imagem contínua. Nesta tese considerou-se o processo de amostragem-reconstrução descrito em [8, 95], onde a amostragem é especificada por:

$$f[m_1, m_2] = \int \int \psi(m_1, m_2, x_1, x_2) f(x_1, x_2) dx_1 dx_2 \quad (\text{A.2})$$

sendo que $f[m_1, m_2]$ é a imagem discreta HR, composta por M_1 pixels na horizontal e M_2 pixels na vertical. Assume-se $M_1 > N_1$ e $M_2 > N_2$. As funções ψ formam a base de aquisição. A reconstrução é especificada por:

$$\tilde{f}(x_1, x_2) = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} f[m_1, m_2] \phi(m_1, m_2, x_1, x_2) \quad (\text{A.3})$$

na qual $\tilde{f}(x_1, x_2)$ é a imagem contínua reconstruída, a qual pode diferir de $f(x_1, x_2)$ dado que no processo de amostragem parte da informação é perdida. As funções ϕ formam a base contínua de reconstrução.

As bases de aquisição e de reconstrução podem ser escolhidas de diversas formas, entretanto, é importante que o conjunto aquisição-reconstrução não insira novas distorções no sistema. Em outras palavras, dados gerados pela aquisição de $\tilde{f}(x_1, x_2)$ ou de $f(x_1, x_2)$ devem ser os mesmos [63, 96].

Em geral, é preferível a utilização um conjunto de funções espaço-invariante para a composição das bases. Uma outra característica importante é que as funções utilizadas como base de reconstrução devem ser interpolantes [95], como a *sinc* e as *splines* cardinais [95, 97], para que as amplitudes de $f[m_1, m_2]$ representem as amostras da imagem $f(x_1, x_2)$. A escolha típica para imagens é o uso de funções retangulares³, com intensidade unitária na área retangular do pixel e intensidade nula fora da área do pixel. Esta função é preferida porque proporciona uma representação bem localizada.

Sendo assim, podem-se substituir (A.3) em (A.1), obtendo-se:

$$\begin{aligned} g[n_1, n_2] &= \int \int D(n_1, n_2, x_1, x_2) f(x_1, x_2) dx_1 dx_2 \\ g[n_1, n_2] &= \int \int D(n_1, n_2, x_1, x_2) \tilde{f}(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (\text{A.4})$$

³Splines de ordem zero

$$\begin{aligned}
 g[n_1, n_2] &= \int \int D(n_1, n_2, x_1, x_2) \left[\sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} f[m_1, m_2] \phi(m_1, m_2, x_1, x_2) \right] dx_1 dx_2 \\
 g[n_1, n_2] &= \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} f[m_1, m_2] \int \int D(n_1, n_2, x_1, x_2) \phi(m_1, m_2, x_1, x_2) dx_1 dx_2
 \end{aligned} \tag{A.5}$$

Realizando a integral seguinte, que corresponde ao produto interno de funções contínuas [8]:

$$d[n_1, n_2, m_1, m_2] = \int \int D(n_1, n_2, x_1, x_2) \phi(m_1, m_2, x_1, x_2) dx_1 dx_2, \tag{A.6}$$

obtém-se:

$$g[n_1, n_2] = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} f[m_1, m_2] d[n_1, n_2, m_1, m_2] \tag{A.7}$$

que é o sistema discretizado em relação à base ϕ .

Colocando os pixels de $g[n_1, n_2]$ e de $f[m_1, m_2]$ na forma de vetor, através de ordenação lexicográfica, e ordenando os elementos da resposta discreta do sistema de aquisição $d[n_1, n_2, m_1, m_2]$ nas respectivas posições em uma matriz é possível representar o sistema na seguinte forma:

$$\mathbf{g} = \mathbf{D}\mathbf{f} \tag{A.8}$$

sendo \mathbf{g} um vetor de tamanho $N \times 1$ com $N = N_1 N_2$, que representa a imagem discreta LR, \mathbf{f} é um vetor de tamanho $M \times 1$, sendo $M = M_1 M_2$, que representa a imagem discreta HR e \mathbf{D} é a matriz do sistema de aquisição discretizado. O sistema discretizado está pronto para ser utilizado nos algoritmos de super-resolução.

A.1.1 Sistemas Invariantes

Em muitos casos a função de aquisição é a mesma para todos os elementos do sensor, apenas com o centro deslocado em relação à posição espacial do elemento do sensor. Nesses casos, representa-se a resposta do sistema por $D(an_1 - x_1, an_2 - x_2)$ e o sistema contínuo-discreto é chamado de invariante [8, 96].

Quando o sistema de aquisição contínuo-discreto e a base de reconstrução são invariantes, tem-se a

seguinte discretização:

$$d[Rn_1 - m_1, Rn_2 - m_2] = \int \int D(an_1 - x_1, an_2 - x_2) \phi(x_1 - bm_1, x_2 - bm_2) dx_1 dx_2 \quad (\text{A.9})$$

na qual $R = a/b$. Caso R seja um número inteiro, o sistema discreto passa a ser invariante também. Assim, apenas a função de aquisição para um elemento de sensor precisa ser encontrada, pois ela é a mesma função, com deslocamento R , para os demais elementos. Pode-se representar (A.7) através de uma convolução bidimensional (2D) seguida de dizimação por R :

$$g[n_1, n_2] = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} d[Rn_1 - m_1, Rn_2 - m_2] f[m_1, m_2] \quad (\text{A.10})$$

$$g[n_1, n_2] = d * f[Rn_1, Rn_2]$$

sendo que $*$ indica convolução [4, 42].

Nesse caso particular, pode-se observar que a matriz \mathbf{D} terá uma estrutura similar à Bloco-Toeplitz [8, 42]. Uma matriz Bloco-Toeplitz é composta por blocos ordenados com estrutura Toeplitz, onde cada bloco é uma submatriz com estrutura Toeplitz. Na estrutura Toeplitz cada linha é a linha anterior deslocada por um elemento. Entretanto, nessa estrutura similar, cada linha é a linha anterior deslocada em R elementos. Da mesma forma os blocos são deslocados em R blocos. Por exemplo, uma matriz Bloco-Toeplitz é:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \cdots & \mathbf{B}_M \\ \mathbf{B}_{-1} & \mathbf{B}_0 & \cdots & \mathbf{B}_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{-M} & \mathbf{B}_{-M+1} & \cdots & \mathbf{B}_0 \end{bmatrix} \quad (\text{A.11})$$

em que cada bloco \mathbf{B}_x é uma submatriz Toeplitz com a mesma estrutura de (A.11).

Já na estrutura similar, com $R = 2$, tem-se:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_M \\ \mathbf{B}_{-2} & \mathbf{B}_{-1} & \mathbf{B}_0 & \cdots & \mathbf{B}_{M-2} \\ \mathbf{B}_{-4} & \mathbf{B}_{-3} & \mathbf{B}_{-2} & \cdots & \mathbf{B}_{M-4} \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{B}_{-M+2} & \mathbf{B}_{-M+3} & \mathbf{B}_{-M+4} & \cdots & \mathbf{B}_2 \\ \mathbf{B}_{-M} & \mathbf{B}_{-M+1} & \mathbf{B}_{-M+2} & \cdots & \mathbf{B}_0 \end{bmatrix} \quad (\text{A.12})$$

sendo que cada bloco \mathbf{B}_x também tem a mesma estrutura de (A.12).

Essa estrutura similar coincide com a representação do sistema de aquisição pela equação (2.2), onde a aquisição é composta pela multiplicação matriz-vetor por uma matriz Bloco-Toeplitz⁴, que corresponde à convolução com as distorções ópticas, seguido pela multiplicação matriz-vetor por uma matriz de subamostragem, que corresponde à dizimação. O uso de matrizes invariantes proporciona vantagens computacionais e, principalmente, economia de memória no armazenamento de matrizes nos algoritmos de super-resolução.

A.1.2 Exemplo do Sistema de Aquisição

Nesta seção será apresentado um exemplo do sistema de aquisição, o qual utiliza a função de aquisição apresentada no Capítulo 6, e proporciona a captura de uma imagem LR de 2×2 pixels, a partir de uma imagem HR de 4×4 pixels. Na Figura A.1 são mostrados os pixels nas imagens LR e HR. Na equação (A.14) o sistema de aquisição é mostrado.

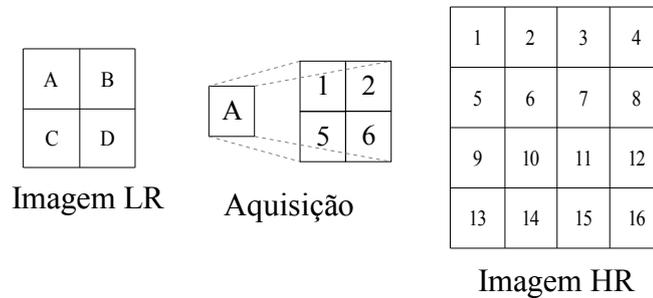


Figura A.1: Ilustração da posição dos pixels nas imagens e da função de aquisição.

$$\mathbf{g}_k = \mathbf{D}_k \mathbf{f}_k \quad (\text{A.13})$$

⁴Em alguns casos, utiliza-se uma aproximação Bloco-Circulante da matriz Bloco-Toeplitz [4, 8, 42]

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ \vdots \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{bmatrix} \quad (\text{A.14})$$

A.2 Modelo Contínuo e Discreto de Movimento

Nesta seção é apresentada a representação do movimento nas formas contínua e discreta. Uma revisão mais completa pode ser encontrada em [73, 98]. O modelo de movimento é apresentado na seção 2.1.2.

O movimento observado em uma seqüência de imagens é a projeção no plano da imagem do movimento tridimensional que ocorre na cena. Esta projeção é chamada de movimento aparente, cuja modelagem é uma forma de representar o efeito causado pelo movimento nas imagens. Isso é feito por meio de uma transformação das coordenadas espaciais da imagem, assumindo que o movimento mantenha as intensidades na seqüência de imagens aproximadamente constante ao longo do tempo.

Modelamos o movimento, nas imagens contínuas, da seguinte forma:

$$f_j(\mathbf{x}) = f_k(\mathbf{x} + \mathbf{v}_{j,k}(\mathbf{x})) + e_{j,k}(\mathbf{x}) \quad (\text{A.15})$$

sendo que $\mathbf{x} = (x_1, x_2)$ são as coordenadas do ponto na imagem contínua no instante temporal j , $\mathbf{v}_{j,k}(\mathbf{x})$ é o vetor de deslocamento e $e_{j,k}(\mathbf{x})$ é o erro de movimento, que modela a nova informação da imagem no instante j , a qual não pode ser obtida da imagem no instante k . O erro de movimento inclui desde pequenas correções na amplitude, ou *inliers*, até o preenchimento com informações completamente novas, ou *outliers*. O conjunto de vetores de deslocamento, para todos os \mathbf{x} , é chamado de campo de vetores.

O comportamento do campo de vetores é dependente dos objetos presentes na cena e do tipo de mo-

vimento envolvido. Entretanto, assume-se que um modelo adequado seja capaz de representar o comportamento dos vetores de deslocamento. Através desse modelo é possível representar o campo de vetores contínuos por um número finito de parâmetros. Mais informações sobre os modelos de representação do campo de vetores são encontradas em [98, 73]. Precisa-se definir o modelo adequado para o campo de vetores e estimar seus parâmetros para utilizar a informação fornecida por (A.15).

Contudo, para que a transformação de movimento seja aplicada, através de processamento digital, é necessário discretizar a equação (A.15). Utiliza-se o mesmo esquema de amostragem-reconstrução adotado na seção A.1 e em [8, 63]. Assim, o sistema discretizado pode ser representado na forma:

$$\mathbf{f}_j = \mathbf{M}_{j,k} \mathbf{f}_k + \mathbf{e}_{j,k} \quad (\text{A.16})$$

que é o modelo adotado neste trabalho, descrito na seção 2.1.2.

Note que a matriz $\mathbf{M}_{j,k}$ depende:

- dos modelo assumido para descrever o comportamento do campo de vetores contínuos;
- dos parâmetros de movimento;
- das funções escolhidas para a amostragem e a reconstrução das imagens.

Em alguns casos particulares, a escolha das funções de amostragem e de reconstrução levam à seguinte representação discreta:

$$f_j[\mathbf{m}] = f_k[\mathbf{m} + \mathbf{v}_{j,k}[\mathbf{m}]] + e_{j,k}[\mathbf{m}] \quad (\text{A.17})$$

sendo que $\mathbf{m} = [m_1, m_2]$ são as coordenadas da imagem discreta, $f_j[\mathbf{m}]$ e $f_k[\mathbf{m}]$ são as imagens discretas nos instantes temporais j e k , respectivamente, e $\mathbf{v}_{j,k}[\mathbf{m}]$ é o vetor de movimento, cujos valores são inteiros. Nesse caso, as linhas da matriz $\mathbf{M}_{j,k}$ são compostas por elementos nulos com exceção de um elemento não-nulo, com valor unitário, que representa o deslocamento entre posições inteiras na grade da imagem. Embora (A.17) seja amplamente utilizada para representar o movimento, vamos utilizar o modelo em (A.16) que é mais geral.

A.2.1 Exemplo do Transformação de Movimento

Nesta seção será apresentado um exemplo da matriz de transformação de movimento para caso de movimento translacional global. Na Figura A.2 são mostrados os pixels nas imagens HR. Na equação

(A.19), a transformação é mostrada. Neste exemplo, assume-se periodicidade nas bordas.

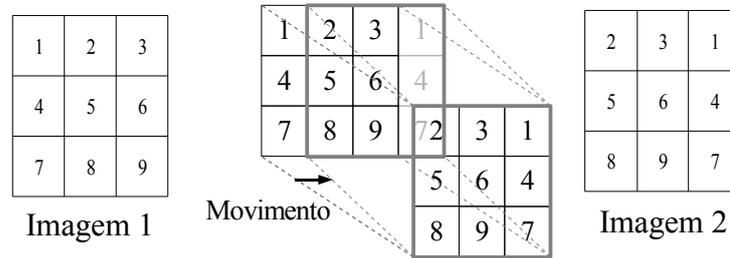


Figura A.2: Ilustração da posição dos pixels nas imagens e da mudança de posição ocorrida com o movimento.

$$\mathbf{f}_1 = \mathbf{M}_{1,2}\mathbf{f}_2 \tag{A.18}$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \\ 5 \\ 6 \\ 4 \\ 8 \\ 9 \\ 7 \end{bmatrix} \tag{A.19}$$

A.3 Formas Comuns da Matriz \mathbf{R}

Nesta seção, são apresentadas as formas mais comuns das matrizes \mathbf{R} , introduzidas na seção 2.1.4. Essas matrizes são utilizadas para determinar as variações das intensidades das imagem. Exemplo de operadores utilizados são: Laplaciano, Gradiente, combinações de derivadas direcionais, entre outros [8, 40]. As matrizes \mathbf{R} representam versões discretas destes operadores [4, 42], que são, geralmente, invariantes. Nesses casos, a matriz \mathbf{R} passa a ser uma matriz Bloco-Toeplitz ou uma concatenação de matrizes Bloco-Toeplitz.

As formas mais comuns são:

- Gradiente Discreto:

$$\begin{aligned} \mathbf{Rf} \equiv \nabla f[m_1, m_2] = & \{f[m_1, m_2] - f[m_1 - 1, m_2]\} \\ & \cup \{f[m_1, m_2] - f[m_1, m_2 - 1]\} \end{aligned} \quad (\text{A.20})$$

Neste caso \mathbf{R} é uma concatenação de duas matrizes Bloco-Toeplitz, uma para derivada discreta na direção horizontal outra para derivada discreta da direção vertical. Essa forma é considerada em [8, 40].

- Diferenças Finitas de Primeira Ordem:

$$\begin{aligned} \mathbf{Rf} \equiv \Delta^1 f[m_1, m_2] = & \{f[m_1, m_2] - f[m_1 - 1, m_2]\} \\ & \cup \{f[m_1, m_2] - f[m_1, m_2 - 1]\} \\ & \cup \left\{ \frac{1}{\sqrt{2}}(f[m_1, m_2] - f[m_1 - 1, m_2 - 1]) \right\} \\ & \cup \left\{ \frac{1}{\sqrt{2}}(f[m_1, m_2] - f[m_1 + 1, m_2 - 1]) \right\} \end{aligned} \quad (\text{A.21})$$

Neste caso \mathbf{R} é uma concatenação de quatro matrizes Bloco-Toeplitz, que representam a operação de diferença finita de primeira ordem, uma para cada direção, sendo: horizontal, vertical, diagonal ascendente e diagonal descendente. Essa forma é considerada nos trabalhos [22, 99].

- Diferenças Finitas de Segunda Ordem:

$$\begin{aligned} \mathbf{Rf} \equiv \Delta^2 f[m_1, m_2] = & \{f[m_1 - 1, m_2] - 2f[m_1, m_2] + f[m_1 + 1, m_2]\} \\ & \cup \{f[m_1, m_2 - 1] - 2f[m_1, m_2] + f[m_1, m_2 + 1]\} \\ & \cup \left\{ \frac{1}{\sqrt{2}}(f[m_1 - 1, m_2 - 1] - 2f[m_1, m_2] + f[m_1 + 1, m_2 + 1]) \right\} \\ & \cup \left\{ \frac{1}{\sqrt{2}}(f[m_1 + 1, m_2 - 1] - 2f[m_1, m_2] + f[m_1 - 1, m_2 + 1]) \right\} \end{aligned} \quad (\text{A.22})$$

Neste caso \mathbf{R} é uma concatenação de quatro matrizes Bloco-Toeplitz, que representam a operação de diferença finita de segunda ordem, uma para cada direção, sendo: horizontal, vertical, diagonal ascendente e diagonal descendente. Essa forma é considerada nos trabalhos [21, 46, 100].

- Laplaciano Discreto A:

$$\begin{aligned} \mathbf{Rf} \equiv \nabla^2 f[m_1, m_2] = & \{f[m_1 - 1, m_2] + f[m_1, m_2 - 1] \\ & - 4f[m_1, m_2] + f[m_1 + 1, m_2] + f[m_1, m_2 + 1]\} \end{aligned} \quad (\text{A.23})$$

Neste caso \mathbf{R} é uma única matriz Bloco-Toeplitz. Essa forma é considerada nos trabalhos [42, 73,

101].

- Laplaciano Discreto B:

$$\begin{aligned}
 \mathbf{Rf} \equiv \nabla^2 f[m_1, m_2] = & \{f[m_1 - 1, m_2] + f[m_1, m_2 - 1] \\
 & + f[m_1 - 1, m_2 - 1] + f[m_1 + 1, m_2 - 1] - 8f[m_1, m_2] \\
 & + f[m_1 + 1, m_2] + f[m_1, m_2 + 1] \\
 & + f[m_1 + 1, m_2 + 1] + f[m_1 - 1, m_2 + 1]\}
 \end{aligned} \tag{A.24}$$

Neste outro caso \mathbf{R} também é uma única matriz Bloco-Toeplitz. Essa forma é considerada nos trabalhos [2, 73].

A.4 Funções de Discrepância mais Conhecidas

Nesta seção as funções de discrepância mais comuns [48] são listadas, são elas:

- Mínimos Quadrados (LS - *Least Squares*):

$$J(\mathbf{g}, \mathbf{x}) = \|\mathbf{g} - \mathbf{x}\|_2^2 = \sum_{i=1}^N |[\mathbf{g}]_i - [\mathbf{x}]_i|^2 \tag{A.25}$$

- Mínimos Quadrados Generalizados ou Ponderados (WLS - *Weighted Least Squares*):

$$\begin{aligned}
 J(\mathbf{g}, \mathbf{x}) &= \|\mathbf{g} - \mathbf{x}\|_{\mathbf{W}}^2 = (\mathbf{g} - \mathbf{x})^T \mathbf{W} (\mathbf{g} - \mathbf{x}) \\
 J(\mathbf{g}, \mathbf{x}) &= (\mathbf{g} - \mathbf{x})^T \mathbf{W} (\mathbf{g} - \mathbf{x}) = \sum_{i=1}^N \sum_{j=1}^N [\mathbf{W}]_{ij} ([\mathbf{g}]_i - [\mathbf{x}]_i) ([\mathbf{g}]_j - [\mathbf{x}]_j)
 \end{aligned} \tag{A.26}$$

- Norma L_p , na potência p :

$$J(\mathbf{g}, \mathbf{x}) = \|\mathbf{g} - \mathbf{x}\|_p^p = \sum_{i=1}^N |[\mathbf{g}]_i - [\mathbf{x}]_i|^p \tag{A.27}$$

- Norma L_p :

$$J(\mathbf{g}, \mathbf{x}) = \|\mathbf{g} - \mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^N |[\mathbf{g}]_i - [\mathbf{x}]_i|^p} \tag{A.28}$$

- Norma Huber:

$$J(\mathbf{g}, \mathbf{x}) = \mathcal{H}_T(\mathbf{g} - \mathbf{x}) = \sum_{i=1}^N H_T([\mathbf{g}]_i - [\mathbf{x}]_i) \tag{A.29}$$

onde

$$H_T(x) = \begin{cases} |x|^2 & \text{se } |x| \leq T \\ 2T|x| - T^2 & \text{se } |x| > T \end{cases} \quad (\text{A.30})$$

sendo que T é o parâmetro que determina o limiar entre a região de ponderação quadrática e a região de ponderação linear.

- Discrepância de Kullback-Leibler:

$$J(\mathbf{g}, \mathbf{x}) = \sum_{i=1}^N [\mathbf{g}]_i \ln \left(\frac{[\mathbf{g}]_i}{[\mathbf{x}]_i} \right) \quad (\text{A.31})$$

- Discrepância Simétrica de Kullback-Leibler:

$$J(\mathbf{g}, \mathbf{x}) = \sum_{i=1}^N \left[[\mathbf{g}]_i \ln \left(\frac{[\mathbf{g}]_i}{[\mathbf{x}]_i} \right) + [\mathbf{x}]_i \ln \left(\frac{[\mathbf{x}]_i}{[\mathbf{g}]_i} \right) \right] \quad (\text{A.32})$$

- Outra variante da Discrepância de Kullback-Leibler:

$$J(\mathbf{g}, \mathbf{x}) = \sum_{i=1}^N \left[[\mathbf{x}]_i - [\mathbf{g}]_i + [\mathbf{g}]_i \ln \left(\frac{[\mathbf{g}]_i}{[\mathbf{x}]_i} \right) \right] \quad (\text{A.33})$$

Apêndice B

Complementos do Capítulo 3

Este apêndice apresenta os complementos do Capítulo 3. No Apêndice B.1 são apresentados os termos de dados considerados na tese; no Apêndice B.2, os exemplos construtivos de \mathbf{M} ; no Apêndice B.3, os métodos de otimização do gradiente conjugado e, finalmente, no Apêndice B.4 são mostrados os gradientes das funções de discrepância mais conhecidas.

B.1 Termos de Dados Considerados

Nesta seção são apresentadas as três distribuições consideradas nesta tese, as quais são comumente encontradas em problemas de aquisição de imagem. São elas: Gaussiana, Laplaciana, Poisson. As três distribuições são assumidas com independentes e identicamente distribuídas [8].

- **Gaussiana:**

Esta distribuição é a mais utilizada para modelar o ruído de aquisição [8], sendo considerada em situações onde o ruído é proveniente de diversas fontes, como ruído eletrônico e excitação aleatória dos elétrons no sensor.

A densidade de probabilidade é:

$$\rho(\mathbf{g}|\mathbf{f}) = \frac{1}{(2\pi\sigma_{\eta}^2)^{\frac{LN}{2}}} e^{-\frac{1}{2\sigma_{\eta}^2}\|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2^2} \quad (\text{B.1})$$

sendo que σ_{η} é o desvio-padrão do ruído de aquisição.

A medida de discrepância equivalente é:

$$\begin{aligned}
 J(\mathbf{g}, \mathbf{D}\mathbf{f}) &= \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2^2 \\
 J(\mathbf{g}, \mathbf{D}\mathbf{f}) &= \sum_{k=1}^L \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_2^2 \\
 J(\mathbf{g}, \mathbf{D}\mathbf{f}) &= \sum_{k=1}^L \sum_{n=1}^N \left| [\mathbf{g}_k]_n - [\mathbf{D}_k \mathbf{f}_k]_n \right|^2
 \end{aligned} \tag{B.2}$$

O hiperparâmetro é: $\theta_\eta = 2\sigma_\eta^2$

- **Laplaciana:**

Esta distribuição é considerada em situações onde o ruído possui picos de grande amplitude em meio a erros de pequena amplitude [55].

A densidade de probabilidade é:

$$\rho(\mathbf{g}|\mathbf{f}) = \frac{1}{(2\theta_\eta)^{LN}} e^{-\frac{1}{\theta_\eta} \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_1} \tag{B.3}$$

sendo que $\sqrt{2}\theta_\eta$ é o desvio-padrão do ruído de aquisição. A medida de discrepância equivalente é:

$$\begin{aligned}
 J(\mathbf{g}, \mathbf{D}\mathbf{f}) &= \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_1 \\
 J(\mathbf{g}, \mathbf{D}\mathbf{f}) &= \sum_{k=1}^L \|\mathbf{g}_k - \mathbf{D}_k \mathbf{f}_k\|_1 \\
 J(\mathbf{g}, \mathbf{D}\mathbf{f}) &= \sum_{k=1}^L \sum_{n=1}^N \left| [\mathbf{g}_k]_n - [\mathbf{D}_k \mathbf{f}_k]_n \right|
 \end{aligned} \tag{B.4}$$

O hiperparâmetro é o próprio θ_η .

- **Poisson:**

Esta distribuição é considerada em aplicações onde a aquisição ocorre com pouca luz. Neste caso, o processo de aquisição da imagem é equivalente a um processo de contagem de fótons [8, 42], na qual a densidade de probabilidade é:

$$\rho(\mathbf{g}|\mathbf{f}) = \prod_{k=1}^L \prod_{n=1}^N \frac{[\mathbf{D}_k \mathbf{f}_k]_n^{[\mathbf{g}_k]_n}}{[\mathbf{g}_k]_n!} e^{-[\mathbf{D}_k \mathbf{f}_k]_n} \tag{B.5}$$

Assumindo que a amplitude de \mathbf{g} equivale à quantidade de fótons que chega no sensor dividida

pela unidade de tempo T , tem-se a seguinte aproximação para a densidade de Poisson:

$$\rho(\mathbf{g}|\mathbf{f}) \approx \prod_{k=1}^L \prod_{n=1}^N \left(\frac{[\mathbf{D}_k \mathbf{f}_k]_n}{[\mathbf{g}_k]_n} \right)^{[T \mathbf{g}_k]_n} e^{-T([\mathbf{D}_k \mathbf{f}_k]_n - [\mathbf{g}_k]_n)} \quad (\text{B.6})$$

sendo que a medida de discrepância equivalente é:

$$J(\mathbf{g}, \mathbf{D}\mathbf{f}) = \sum_{k=1}^L \sum_{n=1}^N \left[[\mathbf{D}_k \mathbf{f}_k]_n - [\mathbf{g}_k]_n + [\mathbf{g}_k]_n \ln \left(\frac{[\mathbf{g}_k]_n}{[\mathbf{D}_k \mathbf{f}_k]_n} \right) \right] \quad (\text{B.7})$$

e o parâmetro de controle é $1/T$.

B.2 Exemplos Construtivos de \mathbf{M}

Nesta seção os exemplos construtivos da matriz \mathbf{M} , usada no termo *a priori* do algoritmo de super-resolução proposto, são apresentados. A matriz \mathbf{M} deve determinar as variações de intensidade ao longo da trajetória de movimento. A melhor forma para sua generalização ainda é um problema em aberto, mas dois casos importantes podem ser considerados.

No primeiro caso, pode-se criar \mathbf{M} para fornecer medidas de diferença entre quadros em instantes temporais diferentes, segundo:

$$\mathcal{V}(\mathbf{M}\mathbf{f}) = \sum_k \sum_j \mathcal{V}(\mathbf{I}\mathbf{f}_k - \mathbf{M}_{k,j}\mathbf{f}_j) \quad (\text{B.8})$$

Sendo que a matriz \mathbf{M} seria:

$$\mathbf{M} = \begin{bmatrix} -\mathbf{M}_{2,1} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{M}_{3,2} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{M}_{L,L-1} & \mathbf{I} \\ \hline -\mathbf{M}_{3,1} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{M}_{4,2} & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{M}_{L,L-1} & \mathbf{0} & \mathbf{I} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline -\mathbf{M}_{L,1} & \mathbf{0} & \cdots & & \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (\text{B.9})$$

No segundo, pode-se criar \mathbf{M} para medidas de ordem arbitrária ao longo da trajetória de movimento:

$$\mathcal{V}(\mathbf{M}\mathbf{f}) = \sum_k \mathcal{V} \left(\sum_j \alpha_j (\mathbf{I}\mathbf{f}_k - \mathbf{M}_{k,j}\mathbf{f}_j) \right) \quad (\text{B.10})$$

Sendo que a matriz \mathbf{M} seria:

$$\mathbf{M} = \begin{bmatrix} \sum_j \alpha_{1,j}\mathbf{I} & -\alpha_{1,2}\mathbf{M}_{1,2} & -\alpha_{1,3}\mathbf{M}_{1,3} & \cdots & -\alpha_{1,L}\mathbf{M}_{1,L} \\ -\alpha_{2,1}\mathbf{M}_{2,1} & \sum_j \alpha_{2,j}\mathbf{I} & -\alpha_{2,3}\mathbf{M}_{2,3} & \cdots & -\alpha_{2,L}\mathbf{M}_{2,L} \\ -\alpha_{3,1}\mathbf{M}_{3,1} & -\alpha_{3,2}\mathbf{M}_{3,2} & \sum_j \alpha_{3,j}\mathbf{I} & \cdots & -\alpha_{3,L}\mathbf{M}_{3,L} \\ \vdots & \vdots & & \ddots & \vdots \\ -\alpha_{L,1}\mathbf{M}_{L,1} & -\alpha_{L,2}\mathbf{M}_{L,2} & \cdots & -\alpha_{L,L-1}\mathbf{M}_{L,L-1} & \sum_j \alpha_{L,j}\mathbf{I} \end{bmatrix}, \quad (\text{B.11})$$

B.3 Métodos de Otimização do Gradiente Conjugado

B.3.1 Gradiente Conjugado Linear

O Gradiente Conjugado (GC) Linear [40] é utilizado para otimização de sistemas cujo ponto de mínimo equivale à solução de um sistema linear de equações. No caso dos algoritmos simultâneos propostos neste trabalho, isso ocorre quando a medida de discrepância é o quadrado da norma ℓ_2 e a penalidade possui potencial quadrático, por exemplo:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{D}\mathbf{f}\|_2^2 + \|\mathbf{R}\mathbf{f}\|_2^2 + \|\mathbf{M}\mathbf{f}\|_2^2 \quad (\text{B.12})$$

Igualando o gradiente a zero, tem-se que o mínimo ocorre quando:

$$\begin{aligned} [\mathbf{D}^T\mathbf{D} + \mathbf{R}^T\mathbf{R} + \mathbf{M}^T\mathbf{M}] \mathbf{f} &= \mathbf{D}^T\mathbf{g} \\ \mathbf{A}\mathbf{f} &= \mathbf{b} \end{aligned} \quad (\text{B.13})$$

Assim, o mínimo de (B.12) é a solução do sistema linear (B.13) e o método do Gradiente Conjugado pode ser aplicado para solução de (B.13). O GC Linear é descrito na Tabela B.1.

Tabela B.1: Gradiente Conjugado Linear

$n := 0;$	
$\mathbf{f}_0 :=$ estimativa inicial;	
$\mathbf{r}_0 := \mathbf{A}\mathbf{f}_0 - \mathbf{b};$	gradiente inicial
$\mathbf{p}_0 := -\mathbf{r}_0;$	direção de busca inicial
$e_0 := \ \mathbf{r}_0\ ^2;$	
Começa iterações GC	
$\mathbf{h}_n := \mathbf{A}\mathbf{p}_n;$	
$\tau_n := e_n/\mathbf{p}_n^T\mathbf{h}_n;$	otimização unidimensional
$\mathbf{f}_{n+1} := \mathbf{f}_n + \tau_n\mathbf{p}_n;$	atualização da solução
$\mathbf{r}_{n+1} := \mathbf{r}_n + \tau_n\mathbf{h}_n;$	atualização do gradiente
$e_{n+1} := \ \mathbf{r}_{n+1}\ ^2;$	
$\beta_n := e_{n+1}/e_n;$	
$\mathbf{p}_{n+1} := -\mathbf{r}_{n+1} + \beta_n\mathbf{p}_n;$	atualização da direção de busca
$n := n + 1;$	
Termina iterações GC	

B.3.2 Gradiente Conjugado Não-Linear

O Gradiente Conjugado Não-Linear é uma variação do Gradiente Conjugado para minimização de funções custo não-quadráticas. Nos algoritmos de super-resolução propostos neste trabalho isso vai ocorrer sempre que a medida de discrepância no termo de dados ou o potencial no termo de informações *a priori* for não-quadrático.

O Gradiente Conjugado Não-Linear tem se mostrado mais eficiente que o método do gradiente *Steepest Descent* [40]; além disso, requer menos processamento que os métodos de Newton para sistemas de grandes dimensões.

O problema de super-resolução pode ser representado da seguinte forma:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} J(\mathbf{f}) \quad (\text{B.14})$$

sendo que $J(\mathbf{f})$ é a função custo não-quadrática a ser minimizada.

O método do Gradiente Conjugado Não-Linear, na versão de Fletcher e Reeves [40], para realização de (B.14) está descrito na Tabela B.2.

Tabela B.2: Gradiente Conjugado Não-Linear

$n := 0;$	
$\mathbf{f}_0 :=$ estimativa inicial;	
$\mathbf{r}_0 := \nabla J(\mathbf{f}_0);$	gradiente inicial
$\mathbf{p}_0 := -\mathbf{r}_0;$	direção de busca inicial
$e_0 := \ \mathbf{r}_0\ ^2;$	
Começa iterações GC	
$\tau_n := \arg \min_{\tau > 0} J(\mathbf{f}_n + \tau \mathbf{p}_n);$	otimização unidimensional
$\mathbf{f}_{n+1} := \mathbf{f}_n + \tau_n \mathbf{p}_n;$	atualização da solução
$\mathbf{r}_{n+1} := \nabla J(\mathbf{f}_{n+1});$	atualização do gradiente
$e_{n+1} := \ \mathbf{r}_{n+1}\ ^2;$	
$\beta_n := e_{n+1}/e_n;$	
$\mathbf{p}_{n+1} := -\mathbf{r}_{n+1} + \beta_n \mathbf{p}_n;$	atualização da direção de busca
$n := n + 1;$	
Termina iterações GC	

Uma variante deste método, desenvolvida por Polak e Ribiere [40], utiliza o seguinte β_n :

$$\beta_n := \frac{(\mathbf{r}_{n+1} - \mathbf{r}_n)^T \mathbf{r}_{n+1}}{(\mathbf{r}_{n+1} - \mathbf{r}_n)^T \mathbf{p}_n} \quad (\text{B.15})$$

B.3.3 Cálculo do Gradiente para o Gradiente Conjugado Não-Linear

O cálculo do gradiente, $\nabla J(\mathbf{f})$, utilizado do Gradiente Conjugado Não-Linear da Tabela B.2 pode ser expresso na seguinte forma:

$$\nabla J(\mathbf{f}_n) = \mathbf{D}^T \mathbf{W}_{(\mathbf{D}\mathbf{f}_n - \mathbf{g})} [\mathbf{D}\mathbf{f}_n - \mathbf{g}] + \lambda_R \mathbf{R}^T \mathbf{W}_{(\mathbf{R}\mathbf{f}_n)} \mathbf{R}\mathbf{f}_n + \lambda_M \mathbf{M}^T \mathbf{W}_{(\mathbf{M}\mathbf{f}_n)} \mathbf{M}\mathbf{f}_n \quad (\text{B.16})$$

sendo que $\mathbf{W}_{(\mathbf{x})}$ é uma matriz diagonal, na qual o elemento i contém a ponderação w_i . Esta ponderação é definida de acordo com a norma escolhida, estas ponderações são apresentadas no Apêndice B.4.

A primeira tentativa para o passo inicial pode ser expressa como:

$$\tau_n = \frac{[\nabla J(\mathbf{f}_n)]^T \mathbf{p}_n}{\mathbf{p}_n^T [\nabla^2 J(\mathbf{f}_n)] \mathbf{p}_n} \quad (\text{B.17})$$

em que τ_n é o passo, e $\nabla^2 J(\mathbf{f}_n) = \mathbf{D}^T \mathbf{W}_{(\mathbf{D}\mathbf{f}_n - \mathbf{g})} \mathbf{D} + \lambda_R \mathbf{R}^T \mathbf{W}_{(\mathbf{R}\mathbf{f}_n)} \mathbf{R} + \lambda_M \mathbf{M}^T \mathbf{W}_{(\mathbf{M}\mathbf{f}_n)} \mathbf{M}$ é a matriz hessiana.

Contudo, se o valor inicial do passo não satisfizer a condição forte de Wolfe [102], a otimização unidimensional precisará ser melhorada. Isso pode ser feito utilizando o método da secante [83], por

exemplo.

B.4 Gradientes das Funções de Discrepância mais Conhecidas

Nesta seção são apresentados os gradientes das Funções de Discrepância mais Conhecidas, e sua ponderação w_i , na forma $\mathbf{A}^T \mathbf{W}_{(\mathbf{Ax})} \mathbf{Ax}$, apresentada no Apêndice B.3

- Mínimos Quadrados (LS - *Least Squares*):

$$J(\mathbf{Ax}) = \|\mathbf{Ax}\|_2^2 \quad (\text{B.18})$$

$$\nabla J(\mathbf{Ax}) = 2\mathbf{A}^T \mathbf{Ax} \quad (\text{B.19})$$

$$w_i = 2 \quad (\text{B.20})$$

- Norma L_p , na potência p :

$$J(\mathbf{Ax}) = \|\mathbf{Ax}\|_p^p \quad (\text{B.21})$$

$$\nabla J(\mathbf{Ax}) = p\mathbf{A}^T \Phi(\mathbf{Ax}) \quad (\text{B.22})$$

sendo $\Phi(\mathbf{x}) = [\phi(x_1) \dots \phi(x_N)]^T$, e:

$$\phi(x) = \text{sign}(x)|x|^{p-1} \quad (\text{B.23})$$

Logo:

$$w_i = \frac{p}{|x_i|^{2-p}} \quad (\text{B.24})$$

- Norma Huber:

$$J(\mathbf{Ax}) = \mathcal{H}_T(\mathbf{Ax}) = \sum_{i=1}^N H_T([\mathbf{Ax}]_i) \quad (\text{B.25})$$

onde

$$H_T(x) = \begin{cases} |x|^2 & \text{se } |x| \leq T \\ 2T|x| - T^2 & \text{se } |x| > T \end{cases} \quad (\text{B.26})$$

sendo que T é o parâmetro que determina a o limiar entre a região de ponderação quadrática e a região de ponderação linear.

$$\nabla J(\mathbf{Ax}) = 2\mathbf{A}^T \mathcal{T}_T(\mathbf{Ax}) \quad (\text{B.27})$$

sendo $\mathcal{T}_T(\mathbf{x}) = [T_T(x_1) \dots T_T(x_N)]^T$, e:

$$T_T(x) = \begin{cases} x & \text{se } |x| \leq T \\ T \text{sign}(x) & \text{se } |x| > T \end{cases} \quad (\text{B.28})$$

Logo:

$$w_i = \begin{cases} 2 & \text{se } |x_i| \leq T \\ \frac{2T}{|x_i|} & \text{se } |x_i| > T \end{cases} \quad (\text{B.29})$$

Apêndice C

Complementos do Capítulo 5

Este apêndice apresenta os complementos do Capítulo 5. No Apêndice C.1 é apresentado o método de gradiente conjugado não-linear usado na determinação dos coeficientes; no Apêndice C.2, a análise de convexidade do método proposto com minimização direta; no Apêndice C.3, a análise de convergência do critério iterativo alternado de aproximação sucessiva e, finalmente, no Apêndice C.4 é mostrada a dedução da escolha alternativa para o μ , necessário no método proposto para determinação dos coeficientes de regularização.

C.1 Método de Gradiente Conjugado Não Linear (GC-NL)

Esta seção apresenta o método do gradiente conjugado não linear usado no método proposto para determinação dos coeficientes. O GC-NL é usado na minimização direta da equação (5.28), no Capítulo 5. O algoritmo GC-NL é apresentado na Tabela C.1:

Tabela C.1: Gradiente Conjugado Não-Linear para Solução da Imagem HR e do Coeficiente

$n := 0;$	
$\mathbf{f}_0 :=$ estimativa inicial;	
$\lambda_0 :=$ coeficiente inicial;	
$\mathbf{r}_0 := \mathbf{C}^T(\mathbf{C}\mathbf{f}_0 - \mathbf{g}) + \lambda_0 \mathbf{R}^T \mathbf{R}\mathbf{f}_0;$	gradiente inicial
$\mathbf{p}_0 := -\mathbf{r}_0;$	direção de busca inicial
$e_0 := \ \mathbf{r}_0\ ^2;$	
Começa iterações GC	
$\mathbf{h}_n := \mathbf{C}^T \mathbf{C}\mathbf{p}_n + \lambda_n \mathbf{R}^T \mathbf{R}\mathbf{p}_n;$	otimização unidimensional
$\tau_n := \mathbf{p}_n^T \mathbf{r}_n / \mathbf{p}_n^T \mathbf{h}_n;$	otimização unidimensional
$\mathbf{f}_{n+1} := \mathbf{f}_n + \tau_n \mathbf{p}_n;$	atualização da solução
$\mathbf{r}_{n+1}^C := \mathbf{C}\mathbf{f}_{n+1} - \mathbf{g}$	parte atualização do gradiente
$\mathbf{r}_{n+1}^R := \mathbf{R}\mathbf{f}_{n+1}$	parte atualização do gradiente
$\lambda_{n+1} := \mu \ \mathbf{r}_{n+1}^C\ _2 / \ \mathbf{r}_{n+1}^R\ _2$	novos λ
$\mathbf{r}_{n+1} := \mathbf{C}^T \mathbf{r}_{n+1}^C + \lambda_{n+1} \mathbf{R}^T \mathbf{r}_{n+1}^R$	atualização do gradiente
$e_{n+1} := \ \mathbf{r}_{n+1}\ ^2;$	
$\beta_n := e_{n+1} / e_n;$	
$\mathbf{p}_{n+1} := -\mathbf{r}_{n+1} + \beta_n \mathbf{p}_n;$	atualização da direção de busca
$n := n + 1;$	
Termina iterações GC	

C.2 Análise de Convexidade do Método Proposto com Minimização Direta

Esta seção tem o objetivo de mostrar que a função custo global usada no critério proposto para determinação dos coeficientes é convexa. Para mostrar isso, é necessário demonstrar que a matriz Hessiana é positiva semi-definida, ou seja, tem todos autovalores maiores ou iguais a zero. O Critério proposto é:

$$\begin{aligned} \hat{\mathbf{f}} &= \arg \min_{\mathbf{f}} J(\mathbf{f}) \\ \hat{\mathbf{f}} &= \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2 + \mu \|\mathbf{R}\mathbf{f}\|_2 \end{aligned} \quad (\text{C.1})$$

O Gradiente é:

$$\nabla_{\mathbf{f}} [J(\mathbf{f})] = \frac{\mathbf{C}^T \mathbf{C}\mathbf{f} - \mathbf{C}^T \mathbf{g}}{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2} + \mu \frac{\mathbf{R}^T \mathbf{R}\mathbf{f}}{\|\mathbf{R}\mathbf{f}\|_2} \quad (\text{C.2})$$

E a matriz Hessiana é:

$$H_{\mathbf{f}} [J(\mathbf{f})] = \frac{1}{\|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2} \mathbf{C}^T [\mathbf{I} - (\mathbf{d}\mathbf{d}^T)] \mathbf{C} + \mu \frac{1}{\|\mathbf{R}\mathbf{f}\|_2} \mathbf{R}^T [\mathbf{I} - (\mathbf{r}\mathbf{r}^T)] \mathbf{R} \quad (\text{C.3})$$

sendo que $\mathbf{d} = (\mathbf{g} - \mathbf{Cf})/\|\mathbf{g} - \mathbf{Cf}\|_2$ e $\mathbf{r} = \mathbf{Rf}/\|\mathbf{Rf}\|_2$ são vetores unitários, \mathbf{I} é a matriz Identidade, e $\mathbf{d}\mathbf{d}^T$ e $\mathbf{r}\mathbf{r}^T$ são operações de produto externo (resultando em uma matriz). Então a Hessiana é positiva semi-definida.

Comprovando que a matriz Hessiana é positiva semi-definida, então a função custo em (C.1) é convexa. Conforme [83, 40], o Gradiente Conjugado Não Linear irá convergir se a função custo for convexa.

C.2.1 Cálculo do Gradiente

Nesta subseção é apresentado o cálculo do Gradiente de:

$$\begin{aligned} J(\mathbf{f}) &= \|\mathbf{b} - \mathbf{Af}\|_2 \\ J(\mathbf{f}) &= \sqrt{\sum_i (b_i - \sum_j a_{ij}f_j)^2} \end{aligned} \quad (\text{C.4})$$

Sendo:

$$\nabla_{\mathbf{f}} [J(\mathbf{f})] = \begin{bmatrix} \frac{\partial}{\partial f_1} \sqrt{\sum_i (b_i - \sum_j a_{ij}f_j)^2} \\ \frac{\partial}{\partial f_2} \sqrt{\sum_i (b_i - \sum_j a_{ij}f_j)^2} \\ \vdots \\ \frac{\partial}{\partial f_M} \sqrt{\sum_i (b_i - \sum_j a_{ij}f_j)^2} \end{bmatrix} \quad (\text{C.5})$$

Sendo que a derivada em uma posição k é:

$$\begin{aligned} \frac{\partial}{\partial f_k} J(\mathbf{f}) &= \left(\frac{\partial}{\partial f_k} \left[\sum_i (b_i - \sum_j a_{ij}f_j)^2 \right]^{1/2} \right) \\ \frac{\partial}{\partial f_k} J(\mathbf{f}) &= \left(\frac{1}{2} \left[\sum_i (b_i - \sum_j a_{ij}f_j)^2 \right]^{-1/2} \right) \left(\frac{\partial}{\partial f_k} \left[\sum_i (b_i - \sum_j a_{ij}f_j)^2 \right] \right) \\ \frac{\partial}{\partial f_k} J(\mathbf{f}) &= \left(\frac{1}{2} \left[\sum_i (b_i - \sum_j a_{ij}f_j)^2 \right]^{-1/2} \right) \left(\sum_i \frac{\partial}{\partial f_k} \left[(b_i - \sum_j a_{ij}f_j)^2 \right] \right) \\ \frac{\partial}{\partial f_k} J(\mathbf{f}) &= \left(\frac{1}{2} \left[\sum_i (b_i - \sum_j a_{ij}f_j)^2 \right]^{-1/2} \right) \left(\sum_i 2(b_i - \sum_j a_{ij}f_j) \left[\frac{\partial}{\partial f_k} (b_i - \sum_j a_{ij}f_j) \right] \right) \end{aligned} \quad (\text{C.6})$$

$$\frac{\partial}{\partial f_k} J(\mathbf{f}) = \left(\frac{1}{2} \left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]^{-1/2} \right) \left(\sum_i 2(b_i - \sum_j a_{ij} f_j)(-a_{ik}) \right) \quad (\text{C.7})$$

$$\frac{\partial}{\partial f_k} J(\mathbf{f}) = (\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^{-1}) \left(\sum_i -a_{ik}(b_i - \sum_j a_{ij} f_j) \right)$$

Juntando todas as derivadas parciais tem-se:

$$\nabla_{\mathbf{f}} [J(\mathbf{f})] = \frac{\mathbf{A}^T (\mathbf{A}\mathbf{f} - \mathbf{b})}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2} \quad (\text{C.8})$$

C.2.2 Cálculo da Hessiana

Nesta subsecção é apresentado o cálculo da Hessiana de:

$$\begin{aligned} J(\mathbf{f}) &= \|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2 \\ J(\mathbf{f}) &= \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} \end{aligned} \quad (\text{C.9})$$

Para a matriz Hessiana tem-se:

$$H_{\mathbf{f}} [J(\mathbf{f})] = \begin{bmatrix} \frac{\partial^2}{\partial f_1^2} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} & \frac{\partial^2}{\partial f_1 \partial f_2} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} & \cdots & \frac{\partial^2}{\partial f_1 \partial f_M} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} \\ \frac{\partial^2}{\partial f_2 \partial f_1} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} & \frac{\partial^2}{\partial f_2^2} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} & \cdots & \frac{\partial^2}{\partial f_2 \partial f_M} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial f_M \partial f_1} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} & \frac{\partial^2}{\partial f_M \partial f_2} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} & \cdots & \frac{\partial^2}{\partial f_M^2} \sqrt{\sum_i (b_i - \sum_j a_{ij} f_j)^2} \end{bmatrix} \quad (\text{C.10})$$

Sendo que a derivada segunda em uma posição kl é:

$$\begin{aligned} \frac{\partial^2}{\partial f_k \partial f_l} J(\mathbf{f}) &= \frac{\partial^2}{\partial f_k \partial f_l} \left(\left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]^{1/2} \right) \\ \frac{\partial^2}{\partial f_k \partial f_l} J(\mathbf{f}) &= \frac{\partial}{\partial f_l} \left(\frac{\partial}{\partial f_k} \left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]^{1/2} \right) \\ \frac{\partial^2}{\partial f_k \partial f_l} J(\mathbf{f}) &= \frac{\partial}{\partial f_l} \left[\left(\left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]^{-1/2} \right) \left(\sum_i -a_{ik}(b_i - \sum_j a_{ij} f_j) \right) \right] \end{aligned} \quad (\text{C.11})$$

$$\begin{aligned}
 \frac{\partial^2}{\partial f_k \partial f_i} J(\mathbf{f}) &= \left[\frac{\partial}{\partial f_i} \left(\left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]^{-1/2} \right) \right] \left(\sum_i -a_{ik} (b_i - \sum_j a_{ij} f_j) \right) \\
 &\quad + (\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^{-1}) \left[\frac{\partial}{\partial f_i} \left(\sum_i -a_{ik} (b_i - \sum_j a_{ij} f_j) \right) \right] \\
 \frac{\partial^2}{\partial f_k \partial f_i} J(\mathbf{f}) &= -1/2 \frac{\frac{\partial}{\partial f_i} \left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]}{\left[\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right]^{3/2}} \left(\sum_i -a_{ik} (b_i - \sum_j a_{ij} f_j) \right) \\
 &\quad + (\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^{-1}) \left(\sum_i a_{ik} a_{il} \right) \\
 \frac{\partial^2}{\partial f_k \partial f_i} J(\mathbf{f}) &= -1/2 \frac{\sum_i \left[2(b_i - \sum_j a_{ij} f_j) \left(\frac{\partial}{\partial f_i} (b_i - \sum_j a_{ij} f_j) \right) \right]}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \left(\sum_i -a_{ik} (b_i - \sum_j a_{ij} f_j) \right) \\
 &\quad + (\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^{-1}) \left(\sum_i a_{ik} a_{il} \right) \\
 \frac{\partial^2}{\partial f_k \partial f_i} J(\mathbf{f}) &= - \frac{\left(\sum_i (b_i - \sum_j a_{ij} f_j) (-a_{il}) \right)}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \left(\sum_i -a_{ik} (b_i - \sum_j a_{ij} f_j) \right) \\
 &\quad + \frac{(\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^2) \left(\sum_i a_{ik} a_{il} \right)}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \\
 \frac{\partial^2}{\partial f_k \partial f_i} J(\mathbf{f}) &= - \frac{\left(\sum_i a_{il} (b_i - \sum_j a_{ij} f_j) \right) \left(\sum_i a_{ik} (b_i - \sum_j a_{ij} f_j) \right)}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \\
 &\quad + \frac{\left(\sum_i (b_i - \sum_j a_{ij} f_j)^2 \right) \left(\sum_i a_{ik} a_{il} \right)}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \\
 \frac{\partial^2}{\partial f_k \partial f_i} J(\mathbf{f}) &= \frac{1}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \left[\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^2 \left(\sum_i a_{ik} a_{il} \right) \right. \\
 &\quad \left. - \left(\sum_i a_{il} (b_i - \sum_j a_{ij} f_j) \right) \left(\sum_i a_{ik} (b_i - \sum_j a_{ij} f_j) \right) \right]
 \end{aligned} \tag{C.12}$$

A matriz Hessiana resultante é:

$$H_{\mathbf{f}} [J(\mathbf{f})] = \frac{1}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^3} \left(\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2^2 \mathbf{A}^T \mathbf{A} - \mathbf{A}^T (\mathbf{b} - \mathbf{A}\mathbf{f}) (\mathbf{b} - \mathbf{A}\mathbf{f})^T \mathbf{A} \right) \tag{C.13}$$

Simplificando:

$$H_{\mathbf{f}} [J(\mathbf{f})] = \frac{1}{\|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2} \left(\mathbf{A}^T [\mathbf{I} - \mathbf{r}\mathbf{r}^T] \mathbf{A} \right) \tag{C.14}$$

onde $\mathbf{r} = (\mathbf{b} - \mathbf{A}\mathbf{f}) / \|\mathbf{b} - \mathbf{A}\mathbf{f}\|_2$.

Observe que é possível decompor a matriz identidade em uma soma de vetores unitários ortogonais

quaisquer, ou seja:

$$\mathbf{I} = \sum_i \mathbf{u}_i \mathbf{u}_i^T \quad (\text{C.15})$$

Suponha que o vetor unitário \mathbf{r} seja um dos vetores \mathbf{u}_i . Assim, a matriz $\mathbf{I} - \mathbf{r}\mathbf{r}^T$ deve ter um autovalor nulo e todos os demais autovalores igual a um. Logo, a Hessiana é positiva semi-definida.

C.3 Análise de Convergência do Critério Iterativo Alternado de Aproximação Sucessiva

O objetivo desta seção é mostrar que o critério iterativo alternado, apresentado na seção 5.2, converge para a solução global. Para esta análise será preciso, inicialmente, apresentar a decomposição generalizada de valores singulares (GSVD) e, em seguida, a decomposição das expressões utilizadas no método de aproximações sucessivas em função da GSVD. Logo após, será apresentada a análise do critério provando a convergência do método proposto.

C.3.1 GSVD - Decomposição Generalizada de Valores Singulares

Dadas duas matrizes \mathbf{C} , de tamanho $N \times M$, e \mathbf{R} , de tamanho $P \times M$, considera-se $N \geq M \geq P$. Tem-se a seguinte GSVD:

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{X}$$

$$\mathbf{R} = \mathbf{V}\mathbf{M}\mathbf{X}$$

$$\mathbf{S}^T\mathbf{S} + \mathbf{M}^T\mathbf{M} = \mathbf{I} \quad (\text{C.16})$$

$$\mathbf{C}^T\mathbf{C} = \mathbf{X}^T\mathbf{S}^T\mathbf{S}\mathbf{X}$$

$$\mathbf{R}^T\mathbf{R} = \mathbf{X}^T\mathbf{M}^T\mathbf{M}\mathbf{X}$$

Em (C.16), \mathbf{I} é a matriz identidade. As matrizes resultantes são: \mathbf{X} , de tamanho $M \times M$, \mathbf{U} , de tamanho $N \times M$, e \mathbf{V} , de tamanho $P \times P$, \mathbf{S} , de tamanho $M \times M$, e \mathbf{M} , de tamanho $P \times M$.

Assume-se $(N + P) \geq M$ e que a matriz $\begin{bmatrix} \mathbf{D} \\ \mathbf{R} \end{bmatrix}$, de tamanho $(N + P) \times M$, tem posto M , então a matriz \mathbf{X} será inversível [89], sendo $\mathbf{X}^{-1} = \mathbf{Y}$ sua inversa. Define-se $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]^T$, sendo

\mathbf{x}_i^T a linha i da matriz, e $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M]$, na qual \mathbf{y}_i é a coluna i matriz inversa. Observe que:

$$\mathbf{x}_i^T \mathbf{y}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (\text{C.17})$$

As colunas das matrizes \mathbf{U} e \mathbf{V} são ortonormais, ou seja:

$$\mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (\text{C.18})$$

e

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (\text{C.19})$$

Sendo \mathbf{v}_i e \mathbf{u}_i as colunas i de \mathbf{V} e \mathbf{U} , respectivamente.

As matrizes \mathbf{M} e \mathbf{S} , são diagonais. Os elementos da diagonal de $\mathbf{M}^T \mathbf{M}$ são: $1 \geq \mu_1^2 \geq \mu_2^2 \geq \dots \geq \mu_M^2 \geq 0$. Considerando que o posto de \mathbf{R} é P , tem-se que $1 \geq \mu_1 \geq \mu_2 \geq \dots \geq \mu_P > 0$, sendo que os elementos são decrescentes. Os elementos da diagonal de $\mathbf{S}^T \mathbf{S}$ são: $0 \leq \sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_M^2 \leq 1$. Considerando que o posto de \mathbf{C} é M , tem-se que $0 < \sigma_1 \leq \dots \leq \sigma_M \leq 1$. Estes elementos são crescentes. Os valores singulares generalizados são $\gamma_i = \sigma_i / \mu_i$.

Logo, tem-se a seguinte GSVD, na forma expandida:

$$\mathbf{C} = \sum_{i=1}^M \mathbf{u}_i \sigma_i \mathbf{x}_i^T \quad (\text{C.20})$$

$$\mathbf{R} = \sum_{i=1}^P \mathbf{v}_i \mu_i \mathbf{x}_i^T$$

A GSVD será utilizada na decomposição dos termos utilizados no método de aproximação sucessiva.

C.3.2 Decomposição dos Termos

Para a análise do método proposto é necessário apresentar as seguintes expressões na forma decomposta. Abaixo tem-se:

$$\mathbf{f}_\lambda = (\mathbf{C}^T \mathbf{C} + \lambda \mathbf{R}^T \mathbf{R})^{-1} \mathbf{C}^T \mathbf{g}$$

$$\mathbf{f}_\lambda = (\mathbf{X}^T (\mathbf{S}^T \mathbf{S} + \lambda \mathbf{M}^T \mathbf{M}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}^T \mathbf{U}^T \mathbf{g} \quad (\text{C.21})$$

$$\mathbf{f}_\lambda = \mathbf{Y} (\mathbf{S}^T \mathbf{S} + \lambda \mathbf{M}^T \mathbf{M})^{-1} \mathbf{S}^T \mathbf{U}^T \mathbf{g}$$

Decompondo em seus componentes:

$$\begin{aligned} \mathbf{f}_\lambda &= \sum_{i=1}^M \left(\frac{\sigma_i}{\sigma_i^2 + \lambda \mu_i^2} (\mathbf{u}_i^T \mathbf{g}) \right) \mathbf{y}_i \\ \mathbf{f}_\lambda &= \sum_{i=1}^P \left(\frac{\sigma_i^2 / \mu_i^2}{\sigma_i^2 / \mu_i^2 + \lambda} \left(\frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \right) \mathbf{y}_i + \sum_{i=P+1}^M \left(\frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \mathbf{y}_i \\ \mathbf{f}_\lambda &= \sum_{i=1}^P \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \left(\frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \right) \mathbf{y}_i + \sum_{i=P+1}^M \left(\frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \mathbf{y}_i \end{aligned} \quad (\text{C.22})$$

sendo que no primeiro somatório, estão os componentes que sofrem influência da regularização, no segundo somatório, estão os componente que não sofrem influência da regularização. Lembrando que no caso, a matriz \mathbf{C} é assumida sobredeterminada e de posto completo.

Decompondo o vetor de dados capturados:

$$\begin{aligned} \mathbf{g} &= \mathbf{C} \mathbf{f} + \eta \\ \mathbf{g} &= \sum_{i=1}^M \sigma_i (\mathbf{x}_i^T \mathbf{f}) \mathbf{u}_i + \eta \end{aligned} \quad (\text{C.23})$$

Decompondo o termo de ruído:

$$\mathbf{g} = \sum_{i=1}^M \sigma_i (\mathbf{x}_i^T \mathbf{f}) \mathbf{u}_i + \sum_{i=1}^M (\mathbf{u}_i^T \eta) \mathbf{u}_i + \eta_\perp \quad (\text{C.24})$$

Observe o vetor η_\perp representa a parte composta pelos componentes fora do “range” de \mathbf{C} . Logo, não há presença da imagem \mathbf{f} nestes componentes, apenas o ruído.

A decomposição do termo de dados é:

$$\begin{aligned}
 \|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2 &= \left\| \sum_{i=1}^M (\mathbf{u}_i^T \mathbf{g}) \mathbf{u}_i + \eta_\perp - \sum_{i=1}^P \left(\sigma_i \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda \mu_i^2} \right) \frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \mathbf{u}_i - \sum_{i=P+1}^M \left(\sigma_i \frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \mathbf{u}_i \right\|_2^2 \\
 \|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2 &= \|\eta_\perp\|_2^2 + \sum_{i=1}^P \left\| \left(\frac{(\sigma_i^2 + \lambda \mu_i^2) - \sigma_i^2}{\sigma_i^2 + \lambda \mu_i^2} (\mathbf{u}_i^T \mathbf{g}) \right) \mathbf{u}_i \right\|_2^2 \\
 \|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2 &= \left\| \sum_{i=1}^P \left(\frac{\lambda}{\gamma_i^2 + \lambda} (\mathbf{u}_i^T \mathbf{g}) \right) \mathbf{u}_i \right\|_2^2 + \|\eta_\perp\|_2^2 \\
 \|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2 &= \sum_{i=1}^P \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + \|\eta_\perp\|_2^2
 \end{aligned} \tag{C.25}$$

Sendo que sua derivada é:

$$\begin{aligned}
 [\|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2]' &= \frac{d[\|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2]}{d\lambda} \\
 [\|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2]' &= \sum_{i=1}^P \frac{2\lambda |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} - \sum_{i=1}^P \frac{2\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \\
 [\|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2]' &= \sum_{i=1}^P \frac{(2\lambda(\gamma_i^2 + \lambda) - 2\lambda^2) |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \\
 [\|\mathbf{g} - \mathbf{Cf}_\lambda\|_2^2]' &= \lambda \sum_{i=1}^P \frac{2\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3}
 \end{aligned} \tag{C.26}$$

A decomposição do termo *a priori* é:

$$\begin{aligned}
 \|\mathbf{Rf}_\lambda\|_2^2 &= \left\| \sum_{i=1}^P \left(\mu_i \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \right) \frac{\mathbf{u}_i^T \mathbf{g}}{\sigma_i} \right) \mathbf{v}_i \right\|_2^2 \\
 \|\mathbf{Rf}_\lambda\|_2^2 &= \left\| \sum_{i=1}^P \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \frac{\mathbf{u}_i^T \mathbf{g}}{\gamma_i} \right) \mathbf{v}_i \right\|_2^2 \\
 \|\mathbf{Rf}_\lambda\|_2^2 &= \sum_{i=1}^P \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2}
 \end{aligned} \tag{C.27}$$

Sendo que sua derivada é:

$$\begin{aligned}
 [\|\mathbf{Rf}_\lambda\|_2^2]' &= \frac{d[\|\mathbf{Rf}_\lambda\|_2^2]}{d\lambda} \\
 [\|\mathbf{Rf}_\lambda\|_2^2]' &= - \sum_{i=1}^P \frac{2\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3}
 \end{aligned} \tag{C.28}$$

Essas decomposições serão utilizadas na análise.

C.3.3 Análise da Convergência da Implementação por Aproximações Sucessivas

Nesta seção, a convergência do algoritmo proposto é demonstrada:

$$\lambda_{k+1} = \mu \frac{\|\mathbf{g} - \mathbf{C}\mathbf{f}_{\lambda_k}\|_2}{\|\mathbf{R}\mathbf{f}_{\lambda_k}\|_2} \triangleq g(\lambda_k) \quad (\text{C.29})$$

Para provar que o algoritmo converge para uma solução λ^* , para qualquer λ_0 inicial, na faixa $0 < \lambda < \infty$, deve-se mostrar que:

$$\left| \frac{dg(\lambda)}{d\lambda} \right| < 1 \quad (\text{C.30})$$

Sendo que esta solução é única se:

$$\frac{df(\lambda)}{d\lambda} = \frac{d}{d\lambda} [\lambda - g(\lambda)] \quad (\text{C.31})$$

não mudar de sinal.

Entretanto, devido a dificuldades em verificar as relações (C.30) e (C.31), será provada uma relação alternativa, igualmente válida. Observe que, como λ é positivo maior que zero, então é possível executar o seguinte algoritmo alternativo:

$$\ln(\lambda_{k+1}) = \ln(g(\lambda_k)) \quad (\text{C.32})$$

Considerando $y_{k+1} = \ln(\lambda_{k+1})$ e $\lambda_k = e^{y_k}$, tem-se:

$$\lambda_{k+1} = g(\lambda_k)$$

$$\ln(\lambda_{k+1}) = \ln(g(\lambda_k)) \quad (\text{C.33})$$

$$y_{k+1} = \ln(g(e^{y_k}))$$

$$y_{k+1} = h(y_k)$$

sendo:

$$h(y_k) = \ln(g(e^{y_k})) \quad (\text{C.34})$$

Logo, se o algoritmo de iterações sucessivas de (C.33) convergir para y^* , então, automaticamente,

$\lambda^* = e^{y^*}$ será encontrado. Para provar a convergência de (C.33), basta mostrar que:

$$\left| \frac{dh(y)}{dy} \right| < 1 \quad (\text{C.35})$$

e para mostrar que a solução é única, é necessário mostrar que a expressão seguinte tem sinal único:

$$\frac{d}{dy} [y - h(y)] \quad (\text{C.36})$$

Observe que y é crescente com derivada igual a 1. Logo, para que (C.36) tenha sinal único, basta que a relação (C.35) seja satisfeita.

Manipulando (C.35), substituindo $\lambda = e^y$ e $\ln(\lambda) = y$ e aplicando a regra de cadeia, tem-se:

$$\begin{aligned} \frac{dh(y)}{dy} &= \frac{d \ln(g(e^y))}{dy} \\ \frac{dh(y)}{dy} &= \frac{d \ln(g(e^y))}{dy} \frac{d\lambda}{d\lambda} \\ \frac{dh(y)}{dy} &= \left(\frac{d \ln(g(\lambda))}{d\lambda} \right) \left(\frac{d\lambda}{dy} \right) \\ \frac{dh(y)}{dy} &= \left(\frac{d \ln(g(\lambda))}{d\lambda} \right) \left(\frac{d\lambda}{dy} \right) \end{aligned} \quad (\text{C.37})$$

Deixando todas variáveis do lado direito em função de λ :

$$\begin{aligned} \frac{dh(y)}{dy} &= \left(\frac{d \ln(g(\lambda))}{d\lambda} \right) \left(\frac{d\lambda}{dy} \right) \\ \frac{dh(y)}{dy} &= \left(\frac{d \ln(g(\lambda))}{d\lambda} \right) \left(\frac{d\lambda}{d \ln(\lambda)} \right) \\ \frac{dh(y)}{dy} &= \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} \end{aligned} \quad (\text{C.38})$$

Assim, é necessário demonstrar que:

$$\left| \frac{dh(y)}{dy} \right| < 1 \quad (C.39)$$

$$\left| \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} \right| < 1$$

Se esta relação for satisfeita, então a iteração (C.29) irá convergir para uma solução única para $0 < \lambda < \infty$.

Continuando a dedução, substituindo $g(\lambda)$ e $[\ln x]' = x'/x$:

$$\begin{aligned} \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \frac{\left[\ln \mu \frac{\|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2}{\|\mathbf{Rf}_{\lambda_k}\|_2} \right]'}{[\ln \lambda]'} \\ \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \lambda [\ln \mu + \ln \|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2 - \ln \|\mathbf{Rf}_{\lambda_k}\|_2]' \\ \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \lambda [[\ln \mu]' + [\ln \|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2]' - [\ln \|\mathbf{Rf}_{\lambda_k}\|_2]'] \\ \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \lambda \left[[\ln \sqrt{\|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2}]' - [\ln \sqrt{\|\mathbf{Rf}_{\lambda_k}\|_2^2}]' \right] \end{aligned} \quad (C.40)$$

$$\begin{aligned} \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \lambda \left[\left[\frac{1}{2} \ln \|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2 \right]' - \left[\frac{1}{2} \ln \|\mathbf{Rf}_{\lambda_k}\|_2^2 \right]' \right] \\ \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \frac{\lambda}{2} \left[[\ln \|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2]' - [\ln \|\mathbf{Rf}_{\lambda_k}\|_2^2]' \right] \\ \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \frac{\lambda}{2} \left[\frac{[\|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2]'}{\|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2} - \frac{[\|\mathbf{Rf}_{\lambda_k}\|_2^2]'}{\|\mathbf{Rf}_{\lambda_k}\|_2^2} \right] \\ \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} &= \frac{\lambda}{2} \left[\frac{[\|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2]' \|\mathbf{Rf}_{\lambda_k}\|_2^2 - [\|\mathbf{Rf}_{\lambda_k}\|_2^2]' \|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2}{\|\mathbf{g} - \mathbf{Cf}_{\lambda_k}\|_2^2 \|\mathbf{Rf}_{\lambda_k}\|_2^2} \right] \end{aligned} \quad (C.41)$$

Substituindo as decomposições da seção C.3.2, na qual a expressão do termo de dados desconsidera

o termo dependente só do ruído, tem-se:

$$\frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} = \frac{\lambda}{2} \left[\frac{[\|\mathbf{g} - \mathbf{C}\mathbf{f}_{\lambda_k}\|_2^2]' \|\mathbf{R}\mathbf{f}_{\lambda_k}\|_2^2 - [\|\mathbf{R}\mathbf{f}_{\lambda_k}\|_2^2]' \|\mathbf{g} - \mathbf{C}\mathbf{f}_{\lambda_k}\|_2^2}{\|\mathbf{g} - \mathbf{C}\mathbf{f}_{\lambda_k}\|_2^2 \|\mathbf{R}\mathbf{f}_{\lambda_k}\|_2^2} \right]$$

$$\frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} = \frac{\lambda}{2} \left[\frac{\left(\lambda \sum_{i=1}^P \frac{2\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_{i=1}^P \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) - \left(- \sum_{i=1}^P \frac{2\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_{i=1}^P \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right)}{\left(\sum_{i=1}^P \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_{i=1}^P \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right)} \right] \quad (\text{C.42})$$

O limites nos índices serão omitidos, dado que são todos de 1 à P . Então:

$$\left| \frac{[\ln(g(\lambda))]' }{[\ln \lambda]'} \right| < 1$$

$$\left| \frac{\lambda}{2} \left[\frac{\left(\sum_i \frac{2\lambda\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) - \left(- \sum_i \frac{2\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right)}{\left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right)} \right] \right| < 1 \quad (\text{C.43})$$

Pode-se retirar o operador de valor absoluto, ele não é mais necessário dado que todos os termos são *sempre positivos ou nulos*:

$$\frac{\lambda}{2} \left[\frac{\left(\sum_i \frac{2\lambda\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) + \left(\sum_i \frac{2\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right)}{\left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right)} \right] < 1$$

$$\lambda \left[\left(\sum_i \frac{\lambda\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) + \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \right] < \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right)$$

$$\left(\sum_i \frac{\lambda^2 \gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) < \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) - \left(\sum_i \frac{\lambda\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right)$$

$$\left(\sum_i \frac{\lambda^2 \gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) < \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2 (\gamma_i^2 + \lambda)}{(\gamma_i^2 + \lambda)^3} - \sum_i \frac{\lambda\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right)$$

$$\left(\sum_i \frac{\lambda^2 \gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_i \frac{\gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} \right) < \left(\sum_i \frac{\lambda^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \quad (\text{C.44})$$

Juntando os somatórios, em que um dos somatórios passará a ser função de j e outro de i :

$$\begin{aligned}
 & \left(\sum_i \frac{\lambda^2 \gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \left(\sum_j \frac{\gamma_j^2 |\mathbf{u}_j^T \mathbf{g}|^2}{(\gamma_j^2 + \lambda)^2} \right) < \left(\sum_j \frac{\lambda^2 |\mathbf{u}_j^T \mathbf{g}|^2}{(\gamma_j^2 + \lambda)^2} + cte \right) \left(\sum_i \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \right) \\
 & \sum_i \sum_j \frac{\lambda^2 \gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2 \gamma_j^2 |\mathbf{u}_j^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^2} < \sum_i \sum_j \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2 \lambda^2 |\mathbf{u}_j^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^2} + cte \sum_i \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \\
 & \sum_i \sum_j \frac{\lambda^2 \gamma_i^2 |\mathbf{u}_i^T \mathbf{g}|^2 \gamma_j^2 |\mathbf{u}_j^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^2} - \sum_i \sum_j \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2 \lambda^2 |\mathbf{u}_j^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^2} < cte \sum_i \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \quad (C.45) \\
 & \sum_i \sum_j \frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2 \gamma_i^2 [\gamma_j^2 - \gamma_i^2]}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^2} < \mathcal{K} \\
 & \sum_i \sum_j \frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2 \gamma_i^2 [\gamma_j^2 - \gamma_i^2]}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2 (\gamma_i^2 + \lambda)} < \mathcal{K}
 \end{aligned}$$

sendo $\mathcal{K} = cte \sum_i \frac{\gamma_i^4 |\mathbf{u}_i^T \mathbf{g}|^2}{(\gamma_i^2 + \lambda)^3} \geq 0$.

Para provar que a iteração (C.29) converge, basta provar que:

$$\sum_{i=1}^P \sum_{j=1}^P \frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2 \gamma_i^2 [\gamma_j^2 - \gamma_i^2]}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2 (\gamma_i^2 + \lambda)} < \mathcal{K} \quad (C.46)$$

Duas situações precisam ser consideradas: primeiro o caso mais extremo, quando todos valores singulares generalizados são iguais e $\mathcal{K} = 0$; segundo, quando não são todos iguais. No primeiro caso, no qual todos valores singulares generalizados, γ_i , são iguais e $\mathcal{K} = 0$, a relação (C.46) é de igualdade com zero, ou seja, o algoritmo (C.29) não converge, nem diverge. No segundo caso, no qual os valores singulares generalizados podem ser diferentes, observa-se que cada elemento do somatório é composto por apenas dois valores singulares distintos, γ_i e γ_j . Caso $i = j$, o elemento é nulo, caso $i \neq j$ tem-se que cada elemento do somatório será:

$$E_{ij} = \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2} \right] \left[\frac{\gamma_i^2}{(\gamma_i^2 + \lambda)} \right] [\gamma_j^2 - \gamma_i^2] = [C_{ij}][a_i][d_{ij}] \quad (C.47)$$

Assim, se a soma do par $E_{ij} + E_{ji}$ for negativa, então o somatório será negativo, satisfazendo a

relação (C.46) para qualquer $\mathcal{K} \geq 0$. Assim, tem-se:

$$\begin{aligned}
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2} \right] \left[\frac{\gamma_i^2}{(\gamma_i^2 + \lambda)} \right] \left[\gamma_j^2 - \gamma_i^2 \right] + \left[\frac{|\mathbf{u}_j^T \mathbf{g}|^2 |\mathbf{u}_i^T \mathbf{g}|^2 \lambda^2}{(\gamma_j^2 + \lambda)^2 (\gamma_i^2 + \lambda)^2} \right] \left[\frac{\gamma_j^2}{(\gamma_j^2 + \lambda)} \right] \left[\gamma_i^2 - \gamma_j^2 \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2} \right] \left[\frac{\gamma_i^2 (\gamma_j^2 + \lambda)}{(\gamma_i^2 + \lambda) (\gamma_j^2 + \lambda)} (\gamma_j^2 - \gamma_i^2) + \frac{\gamma_j^2 (\gamma_i^2 + \lambda)}{(\gamma_j^2 + \lambda) (\gamma_i^2 + \lambda)} (\gamma_i^2 - \gamma_j^2) \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2} \right] \left[\frac{\gamma_i^2 (\gamma_j^2 + \lambda)}{(\gamma_i^2 + \lambda) (\gamma_j^2 + \lambda)} (\gamma_j^2 - \gamma_i^2) - \frac{\gamma_j^2 (\gamma_i^2 + \lambda)}{(\gamma_j^2 + \lambda) (\gamma_i^2 + \lambda)} (\gamma_j^2 - \gamma_i^2) \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^2 (\gamma_j^2 + \lambda)^2} \right] \left[\frac{\gamma_i^2 (\gamma_j^2 + \lambda) (\gamma_j^2 - \gamma_i^2) - \gamma_j^2 (\gamma_i^2 + \lambda) (\gamma_j^2 - \gamma_i^2)}{(\gamma_i^2 + \lambda) (\gamma_j^2 + \lambda)} \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^3} \right] \left[\gamma_i^2 (\gamma_j^2 + \lambda) (\gamma_j^2 - \gamma_i^2) - \gamma_j^2 (\gamma_i^2 + \lambda) (\gamma_j^2 - \gamma_i^2) \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^3} \right] \left[(\gamma_i^2 \gamma_j^2 + \gamma_i^2 \lambda) (\gamma_j^2 - \gamma_i^2) - (\gamma_j^2 \gamma_i^2 + \gamma_j^2 \lambda) (\gamma_j^2 - \gamma_i^2) \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^3} \right] \left[\gamma_i^2 \gamma_j^4 + \gamma_j^2 \gamma_i^2 \lambda - \gamma_i^4 \gamma_j^2 - \gamma_i^4 \lambda - \gamma_i^2 \gamma_j^4 + \gamma_j^2 \gamma_i^2 \lambda + \gamma_i^4 \gamma_j^2 - \gamma_j^4 \lambda \right] \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^3} \right] \left[\gamma_j^2 \gamma_i^2 - \gamma_i^4 + \gamma_j^2 \gamma_i^2 - \gamma_j^4 \right] \lambda \\
 E_{ij} + E_{ji} &= \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^3} \right] \left[-(\gamma_j^2 - \gamma_i^2)^2 \right] \lambda
 \end{aligned} \tag{C.48}$$

Ou seja:

$$E_{ij} + E_{ji} = - \left[\frac{|\mathbf{u}_i^T \mathbf{g}|^2 |\mathbf{u}_j^T \mathbf{g}|^2 \lambda^2}{(\gamma_i^2 + \lambda)^3 (\gamma_j^2 + \lambda)^3} \right] \left[(\gamma_j^2 - \gamma_i^2)^2 \right] \lambda \leq 0 \tag{C.49}$$

A equação (C.49) mostra que a soma do par $E_{ij} + E_{ji}$ será sempre negativa ou nula, então o somatório em (C.46) será negativo, satisfazendo a condição de convergência. Conclui-se que a iteração (C.29) irá convergir para todo $0 < \lambda < \infty$ inicial, se os valores singulares generalizados não forem todos iguais. Entretanto, não haverá convergência, nem divergência, para problemas nos quais todos os valores singulares generalizados são iguais. Por princípio, como o problema em questão é mal condicionado, então os valores singulares generalizados não devem ser todos iguais.

C.4 Escolha Alternativa para o μ

O objetivo desta seção é mostrar uma escolha alternativa para o μ utilizado no método proposto dos coeficientes. Esta escolha é baseada na análise do erro de estimativa $\mathbf{f} - \mathbf{f}_\lambda$. A curva de desempenho do critério proposto, em função de λ , é analisada e comparada com a curva do erro de estimativa. Através desta análise, é encontrada uma expressão para o μ , que faz com que a curva do critério proposto seja o mais similar possível da curva do erro de estimativa. Logo, o ponto mínimo do critério proposto estará bem próximo do λ_{opt} , que é o ponto mínimo do erro de estimativa.

Abaixo, tem-se a equação do critério proposto:

$$\hat{\lambda} = \arg \min_{\lambda} \|\mathbf{g} - \mathbf{C}\mathbf{f}_\lambda\|_2 + \mu \|\mathbf{R}\mathbf{f}_\lambda\|_2 \quad (\text{C.50})$$

sendo que o ponto mínimo depende, dentre outros fatores, do μ escolhido.

O objetivo desta análise é encontrar o μ que faça com que o $\hat{\lambda}$ fique o mais próximo possível do λ_{opt} . Define-se o λ_{opt} como o λ que minimiza a norma quadrática do erro de estimativa, dado pela equação:

$$\|\mathbf{e}_\lambda\|_2 = \|\mathbf{f} - \mathbf{f}_\lambda\|_2 \quad (\text{C.51})$$

O vetor *erro de estimativa*, \mathbf{e}_λ , pode ser decomposto em três componentes:

$$\mathbf{e}_\lambda = \mathbf{f} - \mathbf{f}_\lambda = \mathbf{e}_{cte} + \mathbf{e}_\eta(\lambda) + \mathbf{e}_s(\lambda) \quad (\text{C.52})$$

na qual \mathbf{e}_{cte} é a parte constante, que não depende de λ , $\mathbf{e}_\eta(\lambda)$ é o erro causado pela amplificação do ruído, que é significativo quando o λ é muito pequeno e decresce com o λ , e $\mathbf{e}_s(\lambda)$ que é o erro causado pela regularização excessiva, que é significativo quando λ é muito grande e cresce com o λ .

Duas hipóteses principais serão consideradas nesta busca pelo μ :

- o ponto de mínimo de $\|\mathbf{e}_\lambda\|_2$ é bem próximo do ponto de mínimo de $\|\mathbf{e}_s(\lambda)\|_2 + \|\mathbf{e}_\eta(\lambda)\|_2$.
- existe uma constante a que pode aproximar o comportamento de $\|\mathbf{e}_\eta(\lambda)\|_2$ por $a\|\mathbf{R}\mathbf{f}_\lambda\|_2$, quando λ for pequeno, e existe uma outra constante b que pode aproximar o comportamento de $\|\mathbf{e}_s(\lambda)\|_2$ por $b\|\mathbf{g} - \mathbf{C}\mathbf{f}_\lambda\|_2$, quando λ for grande.

Se essas hipóteses forem satisfeitas, então pode-se definir $\mu = a/b$.

C.4.1 Decomposição dos Termos Através da GSVD

Utilizando a GSVD, apresentada na seção C.3, tem-se a seguinte decomposição para os vetores do erro de estimação:

$$\mathbf{e}_s(\lambda) = \sum_{i=1}^P \left(\frac{\lambda}{\gamma_i^2 + \lambda} \right) (\mathbf{x}_i^T \mathbf{f}) \mathbf{y}_i \quad (\text{C.53})$$

para o erro causado pela regularização excessiva, e:

$$\mathbf{e}_\eta(\lambda) = - \sum_{i=1}^P \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \right) \left(\frac{\eta_i}{\sigma_i} \right) \mathbf{y}_i \quad (\text{C.54})$$

para o erro causado pelo ruído amplificado.

Para os vetores do termo de dado e do termo de regularização tem-se:

$$\mathbf{d}(\lambda) = \mathbf{g} - \mathbf{C}\mathbf{f}_\lambda \quad (\text{C.55})$$

$$\mathbf{d}(\lambda) = \sum_{i=1}^P \left(\frac{\lambda}{\gamma_i^2 + \lambda} \right) (\sigma_i (\mathbf{x}_i^T \mathbf{f}) + \eta_i) \mathbf{u}_i + \eta_\perp$$

para o termo de dados, sendo que $\mathbf{d}(\lambda)$ é o vetor de resíduo em função de λ , e:

$$\begin{aligned} \mathbf{w}(\lambda) &= \mathbf{R}\mathbf{f}_\lambda \\ \mathbf{w}(\lambda) &= \sum_{i=1}^P \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \right) \left(\mu_i (\mathbf{x}_i^T \mathbf{f}) + \frac{\mu_i \eta_i}{\sigma_i} \right) \mathbf{v}_i \end{aligned} \quad (\text{C.56})$$

para o termo de regularização, sendo que $\mathbf{w}(\lambda)$ é o vetor que representa as variações da imagem restaurada com o respectivo λ .

Define-se as seguintes expressões para relacionar os termos da função custo proposta com os termos do erro de estimativa:

$$\mathbf{d}(\lambda) = \mathbf{C}\mathbf{e}_s(\lambda) + \eta_{FILT}(\lambda) + \eta_\perp \quad (\text{C.57})$$

sendo que $\eta_{FILT}(\lambda) = \sum_{i=1}^P \left(\frac{\lambda}{\gamma_i^2 + \lambda} \right) (\eta_i) \mathbf{u}_i$ é a parte do ruído que sofre o efeito do λ , e:

$$\mathbf{w}(\lambda) = -\mathbf{R}\mathbf{e}_\eta(\lambda) + \mathbf{R}\mathbf{f}_{FILT}(\lambda) \quad (\text{C.58})$$

sendo que $\mathbf{f}_{FILT}(\lambda) = \sum_{i=1}^P \left(\frac{\gamma_i^2}{\gamma_i^2 + \lambda} \right) (\mathbf{x}_i^T \mathbf{f}) \mathbf{y}_i$ é a imagem suavizada pelo λ utilizado. Estas igualdades são exatas.

Observe que o vetor usado no termos de dados, $\mathbf{d}(\lambda)$, está relacionado com o erro de regularização excessiva, $\mathbf{e}_s(\lambda)$, através da multiplicação pela matriz \mathbf{C} , mais um componente secundário proveniente do ruído, $\eta_{FILT}(\lambda) + \eta_{\perp}$. Já o termo de regularização, $\mathbf{w}(\lambda)$, está relacionado com o erro de ruído amplificado, $\mathbf{e}_{\eta}(\lambda)$, através da multiplicação pela matriz \mathbf{R} , mais um componente secundário proveniente da imagem regularizada, $\mathbf{Rf}_{FILT}(\lambda)$.

C.4.2 Análise da Similaridade dos Termos

Para encontrar as constantes a e b , que aproximam o tamanho das normas em (C.50) do tamanho dos componentes do erro em (C.52), é necessário entender dois efeitos importantes: o efeito da multiplicação da matriz, \mathbf{R} ou \mathbf{C} , e o efeito da soma do componente secundário.

Simplificaremos a análise assumindo as seguintes aproximações:

$$\frac{\|\mathbf{e}_s(\lambda)\|_2^2}{b(\lambda)} = \|\mathbf{d}(\lambda)\|_2^2 \approx \|\mathbf{C}\mathbf{e}_s(\lambda)\|_2^2 + \|\eta_{FILT}(\lambda)\|_2^2 + \|\eta_{\perp}\|_2^2 \quad (\text{C.59})$$

e

$$\frac{\|\mathbf{e}_{\eta}(\lambda)\|_2^2}{a(\lambda)} = \|\mathbf{w}(\lambda)\|_2^2 \approx \|\mathbf{R}\mathbf{e}_{\eta}(\lambda)\|_2^2 + \|\mathbf{Rf}_{FILT}(\lambda)\|_2^2 \quad (\text{C.60})$$

Por enquanto, a e b serão consideradas como variáveis em função de λ . Estas aproximações não são válidas no caso geral, contudo, no problema de super-resolução elas são bem razoáveis, conforme será observado mais a frente.

Similaridade do Termo de Dados e do Erro de Suavização Excessiva

Considerando o efeito da multiplicação da matriz no erro de suavização excessiva, tem-se:

$$\|\mathbf{C}\mathbf{e}_s(\lambda)\|_2^2 = c(\lambda)^2 \|\mathbf{e}_s(\lambda)\|_2^2 \quad (\text{C.61})$$

sendo:

$$c(\lambda)^2 = \frac{\|\mathbf{C}\mathbf{e}_s(\lambda)\|_2^2}{\|\mathbf{e}_s(\lambda)\|_2^2} \quad (\text{C.62})$$

na qual $\sigma_{min}^2 \leq c(\lambda)^2 \leq \sigma_{max}^2$, ou seja, o coeficiente de multiplicação pode mudar ao longo de λ (porque $\mathbf{e}_s(\lambda)$ muda ao longo de λ) e está entre o maior e o menor valor singular da matriz \mathbf{C} .

Considerando os dois efeitos, tem-se:

$$\|\mathbf{e}_s(\lambda)\|_2 \approx b(\lambda) \sqrt{c(\lambda)^2 \|\mathbf{e}_s(\lambda)\|_2^2 + \|\eta_{FILT}(\lambda)\|_2^2 + \|\eta_{\perp}\|_2^2} \quad (\text{C.63})$$

assim:

$$\frac{1}{b(\lambda)} \approx \sqrt{c(\lambda)^2 + \frac{\|\eta_{FILT}(\lambda)\|_2^2 + \|\eta_{\perp}\|_2^2}{\|\mathbf{e}_s(\lambda)\|_2^2}} \quad (\text{C.64})$$

Logo, o coeficiente $b(\lambda)$, que ajusta a norma $\|\mathbf{d}(\lambda)\|_2$ para que esta se torne igual a $\|\mathbf{e}_s(\lambda)\|_2$ se modifica ao longo de λ e depende do efeito da matriz \mathbf{C} e do componente secundário. O componente secundário pode ser definido como:

$$NS(\lambda)^2 = \frac{\|\eta_{FILT}(\lambda)\|_2^2 + \|\eta_{\perp}\|_2^2}{\|\mathbf{e}_s(\lambda)\|_2^2} \quad (\text{C.65})$$

sendo que $NS(\lambda)$ expressa a razão entre a norma do ruído pela norma da regularização excessiva. O resultado é a seguinte relação para $b(\lambda)$:

$$\frac{1}{b(\lambda)} \approx \sqrt{c(\lambda)^2 + NS(\lambda)^2} \quad (\text{C.66})$$

Esta formulação é válida para diferentes tipos de problemas inversos. Contudo, considerando o problema de super-resolução, a equação (C.66) pode ser bastante simplificada.

A correção da magnitude do termo de dados é importante quando o λ é grande, pois a magnitude do termo de dados é significativamente maior que a do termo de regularização. Quando o λ é grande, a imagem estimada está excessivamente regularizada. Logo, o componente $NS(\lambda)$, que é a razão entre a norma do ruído pela norma da regularização excessiva, é bem pequeno e pode ser desprezado. Além disso, uma análise experimental mostrou que nos problemas de super-resolução o $c(\lambda)$ é, em geral, menor que o valor singular máximo e maior que a média dos valores singulares de \mathbf{C} . Assim, as seguintes expressões simplificadas podem ser utilizadas para representar o $\frac{1}{b}$:

$$\frac{1}{b} \approx \sqrt{\text{tr}(\mathbf{C}^T \mathbf{C})/M} \quad (\text{C.67})$$

que aproxima os valores singulares médios de \mathbf{C} , na qual $\text{tr}(\mathbf{C})$ especifica o traço da matriz \mathbf{C} , ou:

$$\frac{1}{b} \approx \|\mathbf{C}\|_2 \quad (\text{C.68})$$

que é o valor singular máximo de \mathbf{C} .

A Figura C.1 ilustra o comportamento das curvas $\frac{1}{b(\lambda)}$, $c(\lambda)$ e $NS(\lambda)$ e a aproximação pelas constantes sugeridas nas equações (C.67) e (C.68). Pode-se observar que os valores sugeridos estão próximos do valor ideal para $\frac{1}{b}$.

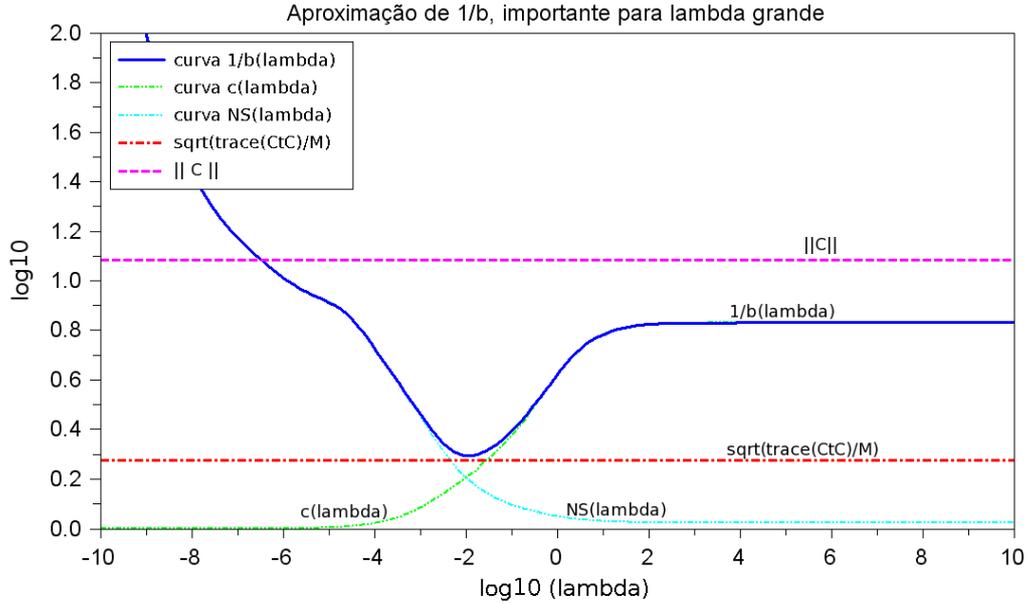


Figura C.1: Comportamento das curvas $\frac{1}{b(\lambda)}$, $c(\lambda)$ e $NS(\lambda)$.

Similaridade do Termo de Regularização e do Erro de Ruído Amplificado

Considerando o efeito da multiplicação da matriz \mathbf{R} no erro de ruído amplificado, tem-se:

$$\|\mathbf{R}\mathbf{e}_\eta(\lambda)\|_2^2 = r(\lambda)^2 \|\mathbf{e}_\eta(\lambda)\|_2^2 \quad (\text{C.69})$$

$$r(\lambda)^2 = \frac{\|\mathbf{R}\mathbf{e}_\eta(\lambda)\|_2^2}{\|\mathbf{e}_\eta(\lambda)\|_2^2} \quad (\text{C.70})$$

sendo que $\sigma_{min}^2 \leq r(\lambda)^2 \leq \sigma_{max}^2$, ou seja, o coeficiente de multiplicação pode mudar ao longo de λ e está entre o maior e o menor valor singular da matriz \mathbf{R} .

Considerando os dois efeitos temos:

$$\|\mathbf{e}_\eta(\lambda)\|_2 \approx a(\lambda) \sqrt{r(\lambda)^2 \|\mathbf{e}_\eta(\lambda)\|_2^2 + \|\mathbf{R}\mathbf{f}_{FILT}(\lambda)\|_2^2} \quad (\text{C.71})$$

assim:

$$\frac{1}{a(\lambda)} \approx \sqrt{r(\lambda)^2 + \frac{\|\mathbf{R}\mathbf{f}_{FILT}(\lambda)\|_2^2}{\|\mathbf{e}_\eta(\lambda)\|_2^2}} \quad (\text{C.72})$$

Logo, o coeficiente $a(\lambda)$, que ajusta a norma $\|\mathbf{w}(\lambda)\|_2$ para que esta se torne igual à $\|\mathbf{e}_\eta(\lambda)\|_2$, se modifica ao longo de λ e depende do efeito da matriz \mathbf{R} e do termo secundário. Esta relação pode ser definida como:

$$DA(\lambda)^2 = \frac{\|\mathbf{R}\mathbf{f}_{FILT}(\lambda)\|_2^2}{\|\mathbf{e}_\eta(\lambda)\|_2^2} \quad (\text{C.73})$$

na qual $DA(\lambda)$ expressa a razão entre a norma da derivada da imagem pela norma do ruído amplificado. O resultado é a seguinte função $a(\lambda)$ em:

$$\frac{1}{a(\lambda)} \approx \sqrt{r(\lambda)^2 + DA(\lambda)^2} \quad (\text{C.74})$$

Assim como (C.66), esta formulação também é válida para diferentes tipos de problemas inversos. Também, considerando o problema de super-resolução, a equação (C.74) pode ser bastante simplificada.

A correção da magnitude do termo de regularização é mais importante quando o λ é pequeno, pois a magnitude do termo de regularização é predominante no critério. Quando o λ é pequeno, a imagem estimada está bastante corrompida pelo ruído amplificado. Logo, o componente $DA(\lambda)$, que expressa a razão entre a norma da derivada da imagem pela norma do ruído amplificado é bem pequena e pode ser desprezada. Além disso, uma análise experimental mostrou que nos problemas de super-resolução o $r(\lambda)$ é, em geral, menor que o valor singular máximo e maior que a média dos valores singulares de \mathbf{R} . As seguintes expressões podem ser utilizadas para substituir o $\frac{1}{a}$:

$$\frac{1}{a} \approx \sqrt{\text{tr}(\mathbf{R}^T \mathbf{R})/M} \quad (\text{C.75})$$

que aproxima a média dos valores singulares de \mathbf{R} , e:

$$\frac{1}{a} \approx \|\mathbf{R}\|_2 \quad (\text{C.76})$$

que é o valor singular máximo de \mathbf{R} .

A Figura C.2 ilustra o comportamento das curvas $\frac{1}{a(\lambda)}$, $r(\lambda)$ e $DA(\lambda)$ e a aproximação pelas constantes sugeridas nas equações (C.75) e (C.76). Pode-se observar que os valores sugeridos estão bem próximos do valor ideal para $\frac{1}{a}$.

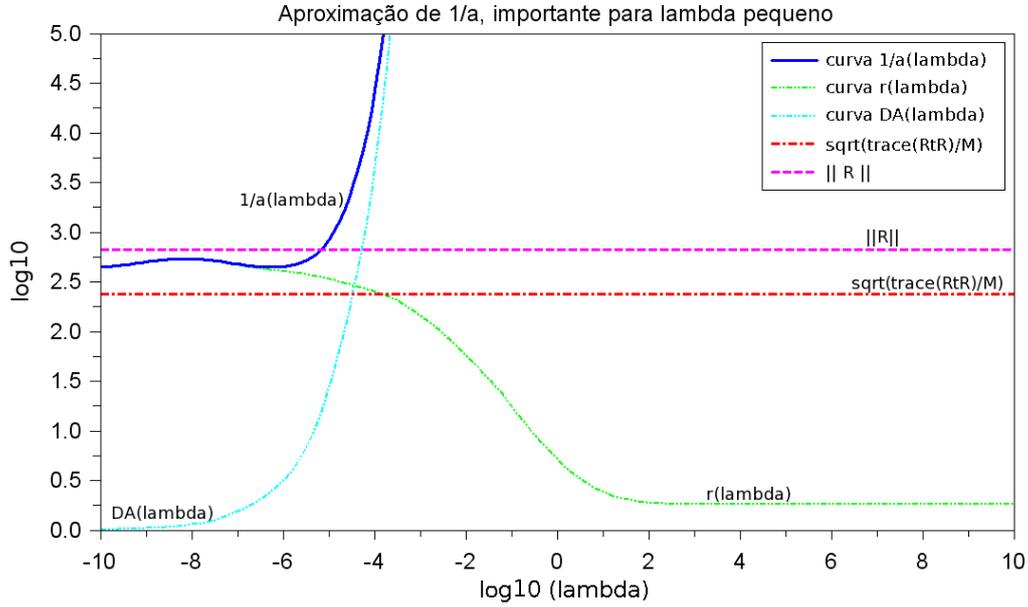


Figura C.2: Comportamento das curvas $\frac{1}{a(\lambda)}$, $r(\lambda)$ e $DA(\lambda)$.

C.4.3 Escolha do μ

Assim, sob as condições especificadas, é possível escolher um μ constante que aproxime a curva do critério proposto, em função de λ , da curva do erro de estimativa. Expressões simplificadas para μ que podem ser utilizadas nestas situações são:

Correção tipo 1:

$$\mu = \frac{a}{b} = \frac{\sqrt{\text{tr}(\mathbf{C}^T \mathbf{C})/M}}{\sqrt{\text{tr}(\mathbf{R}^T \mathbf{R})/M}} \quad (\text{C.77})$$

Correção tipo 2:

$$\mu = \frac{a}{b} = \frac{\|\mathbf{C}\|_2}{\|\mathbf{R}\|_2} \quad (\text{C.78})$$

Estas alternativas foram propostas devido à simplicidade em sua obtenção. As figuras C.1 e C.2 mostram que estas opções não estão longe do valor ideal. A Figura C.3 ilustra o desempenho das escolhas dadas pelas equações (C.77) e (C.78). Pode-se observar na Figura C.3 que as duas alternativas são suficientemente boas para o propósito. Esta figura mostra que os pontos de mínimo dos critérios corrigidos são bem mais próximos do ponto mínimo do erro de estimativa que o critério com $\mu = 1$. Nos experimentos realizados no Capítulo 6 desta tese optou-se pela escolha da equação (C.77), que mostrou-se ligeiramente melhor nos experimentos.

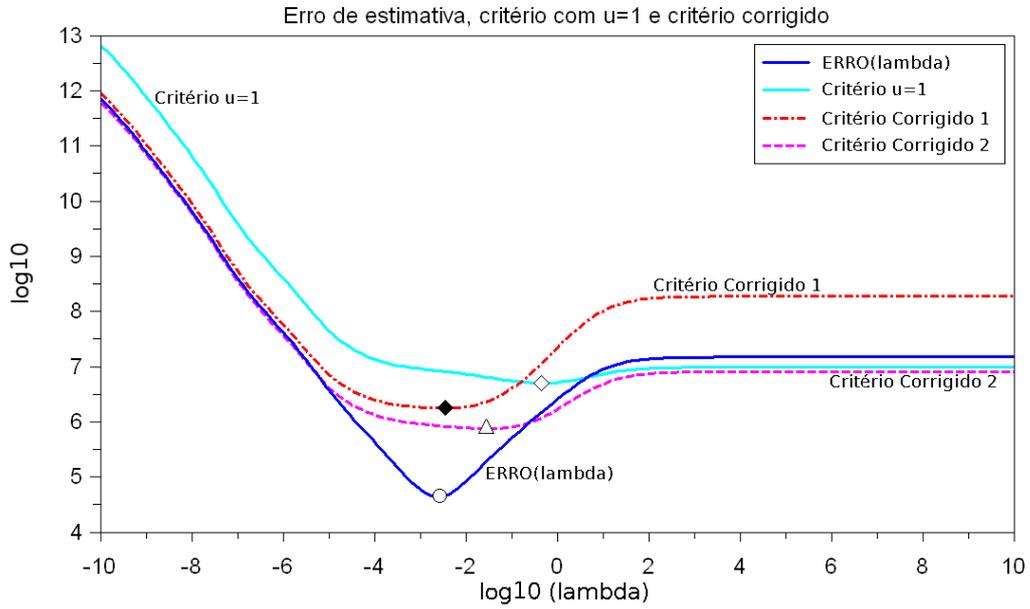


Figura C.3: Erro de estimativa, critério com $\mu = 1$ e critério corrigido.

A grande vantagem destas escolhas é que elas não dependem do conhecimento específico sobre o ruído ou sobre a derivada da imagem. Contudo, Elas só são válidas se a razão entre a norma da derivada da imagem pela norma do ruído amplificado, $DA(\lambda)$, for bem pequena quando λ for pequeno e se a razão entre a norma do ruído pela norma da regularização excessiva, $NS(\lambda)$, for bem pequena quando λ for grande.

Referências Bibliográficas

- [1] BORMAN, S.; STEVENSON, R. L. Simultaneous multi-frame map super-resolution video enhancement using spatio-temporal priors. In: *IEEE International Conference on Image Processing*. 1999. v. 3, p. 469–473.
- [2] PARK, S. C.; PARK, M. K.; KANG, M. G. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, v. 20, n. 3, p. 21–36, maio 2003.
- [3] CHAUDHURI, S. *Super-Resolution Imaging*. : Kluwer, 2001. (The Kluwer International Series in Engineering and Computer Science).
- [4] GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. : Addison-Wesley, 1992.
- [5] QUISTGAARD, J. U. Signal acquisition and processing in medical diagnostic ultrasound. *IEEE Signal Processing Magazine*, v. 14, n. 1, p. 67–74, jan. 1997.
- [6] OLLINGER, J. M.; FESSLER, J. A. Positron-emission tomography. *IEEE Signal Processing Magazine*, v. 14, n. 1, p. 43–55, jan. 1997.
- [7] MOLINA, R. et al. Image restoration in astronomy. *IEEE Signal Processing Magazine*, v. 18, n. 2, p. 11–29, mar. 2001.
- [8] BARRETT, H. H.; MYERS, K. J. *Foundations of Image Science*. : John Wiley & Sons, 2004. (Wiley Series in Pure and Applied Optics).
- [9] BERTERO, M.; BOCCACCI, P. Super-resolution in computational imaging. *Micron*, v. 34, n. 6, p. 265–273, out. 2003.
- [10] BORMAN, S.; STEVENSON, R. L. *Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research*. 1998.
- [11] BORMAN, S.; STEVENSON, R. L. Super-resolution from image sequences - a review. In: *Proc. 1998 Midwest Symp. Circuits and Systems*. 1999. p. 374–378.
- [12] NGUYEN, N. X. *Numerical Algorithms for Image Superresolution*. Tese (Doutorado) — Stanford University, jul. 2000.
- [13] LUKOSZ, W. Optical systems with resolving power exceeding the classical limit. *Jornal of the Optical Society of America*, v. 56, n. 11, p. 1463–1472, 1966.

- [14] LUKOSZ, W. Optical systems with resolving power exceeding the classical limit ii. *Journal of the Optical Society of America*, v. 57, n. 7, p. 932–941, 1967.
- [15] TSAI, R. Y.; HUANG, T. S. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, v. 1, p. 317–339, 1984.
- [16] ELAD, M.; FEUER, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, v. 6, n. 12, p. 1646–1658, dez. 1997.
- [17] RAJAN, D.; CHAUDHURI, S.; JOSHI, M. V. Multi-objective super resolution: concepts and examples. *IEEE Signal Processing Magazine*, v. 20, n. 3, p. 49–61, maio 2003.
- [18] THÉVENAZ, P.; BLU, T.; UNSER, M. Handbook of medical imaging, processing and analysis. In: _____. San Diego Ca, USA: Academic Press, 2000. cap. Image Interpolation and Resampling, p. 393–420.
- [19] THÉVENAZ, P.; BLU, T.; UNSER, M. Interpolation revisited. *IEEE Transactions on Medical Imaging*, v. 19, n. 7, p. 739–758, jul. 2000.
- [20] BANHAM, M. R.; KATSAGGELOS, A. K. Digital image restoration. *IEEE Signal Processing Magazine*, v. 14, n. 2, p. 24–41, mar. 1997.
- [21] SCHULTZ, R. R.; STEVENSON, R. L. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, v. 5, n. 6, p. 996–1011, jun. 1996.
- [22] CAPEL, D.; ZISSERMAN, A. Computer vision applied to super resolution. *IEEE Signal Processing Magazine*, v. 20, n. 3, p. 75–86, maio 2003.
- [23] SEGALL, C. A.; MOLINA, R.; KATSAGGELOS, A. K. High-resolution imagens from low-resolution compressed video. *IEEE Signal Processing Magazine*, v. 20, n. 8, p. 37–48, maio 2003.
- [24] ELAD, M.; FEUER, A. Super-resolution reconstruction of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 21, n. 9, p. 817–834, set. 1999.
- [25] MATEOS, J.; KATSAGGELOS, A. K.; MOLINA, R. Resolution enhancement of compressed low resolution video. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. 2000. v. 4, p. 1919–1922.
- [26] ZIBETTI, M. V. W.; MAYER, J. Simultaneous super-resolution for video sequences. In: *IEEE International Conference on Image Processing*. 2005. v. 1, p. 877–880.
- [27] ZIBETTI, M. V. W.; MAYER, J. Adaptive regularization for resolution restoration of multiframe considering subpixel contributions. In: *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*. : IEEE, 2003. v. 2, p. 952–956.
- [28] ZIBETTI, M. V. W.; MAYER, J. Compensation of the effects caused by the outliers in multiframe iterative techniques. In: *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*. : IEEE, 2003. v. 2, p. 957–960.

- [29] ZIBETTI, M. V. W.; MAYER, J. Um modelo de aquisição de seqüência de imagens com melhoria de resolução através de regularização adaptativa. In: *Anais Simpósio Brasileiro de Sensoriamento Remoto*. Belo Horizonte, Brazil: , 2003. p. 2161–2168.
- [30] ZIBETTI, M. V. W. *Restauração da resolução de seqüência de Imagens usando regularização adaptativa*. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, fev. 2003.
- [31] ZIBETTI, M. V. W.; MAYER, J. Outlier robust and edge-preserving simultaneous super-resolution. In: *IEEE International Conference on Image Processing*. 2006. v. 1, p. 1741–1744.
- [32] ZIBETTI, M. V. W.; MAYER, J. Super-resolução regularizada e simultânea de seqüências de imagens. *Revista da Sociedade Brasileira de Telecomunicações*, v. 21, n. 2, p. 45–57, ago. 2006.
- [33] ZIBETTI, M. V. W.; MAYER, J. A robust and computationally efficient simultaneous super-resolution scheme for image sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007. Submetida para publicação.
- [34] BOSE, N. K.; LERTRATTANAPANICH, S.; KOO, J. Advances in superresolution using l-curve. In: *The 2001 IEEE International Symposium on Circuits and Systems*. 2001. v. 2, p. 433–436.
- [35] NGUYEN, N.; MILANFAR, P.; GOLUB, G. A computationally efficient super-resolution image reconstruction algorithm. *IEEE Transactions on Image Processing*, v. 10, n. 3, p. 573–583, abr. 2001.
- [36] MOLINA, R. et al. Parameter estimation in bayesian high-resolution image reconstruction with multisensors. *IEEE Transactions on Image Processing*, v. 12, n. 12, p. 1655–1667, dez. 2003.
- [37] LEE, E. S.; KANG, M. G. Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration. *IEEE Transactions on Image Processing*, v. 12, n. 7, p. 826–837, jul. 2003.
- [38] HE, H.; KONDI, L. P. An image super-resolution algorithm for different error levels per frame. *IEEE Transactions on Image Processing*, v. 15, n. 3, p. 592–603, mar. 2006.
- [39] HANSEN, P. C. (Ed.). *Rank-Deficient and Discrete Ill-Posed Problems*. : SIAM, 1998. (SIAM Monographs an Matematical Modeling and Computation).
- [40] VOGEL, C. R. *Computational Methods for Inverse Problems*. : SIAM, 2002. (Frontiers in Applied Mathematics).
- [41] ZIBETTI, M. V. W.; BAZÁN, F. S. V.; MAYER, J. Determining the regularization parameters for super-resolution problems. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007. Submetida para publicação.
- [42] BOVIK, A. (Ed.). *Handbook of Image and Video Processing*. : Academic Press, 2000.
- [43] SAAD, Y. *Iterative methods for sparse linear systems*. : International Thompson Publishing, 1995.
- [44] HUBER, P. J. *Robust Statistics*. : John Wiley & Sons, 1981.

- [45] HASLER, D. et al. Outlier modeling in image matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 25, n. 3, p. 301–315, mar. 2003.
- [46] SCHULTZ, R. R.; MENG, L.; STEVENSON, R. L. Subpixel motion estimation for super-resolution image sequence enhancement. *Journal of Visual Communication and Image Representation*, v. 9, n. 1, p. 38–50, mar. 1998.
- [47] WINKLER, G. *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*. 2nd edition. ed. : Springer-Verlag, 2003. (Applications of mathematics).
- [48] MOHAMMAD-DJAFARI, A. From deterministic to probabilistic approaches to solve inverse problems. In: *Bayesian Inference for Inverse Problems, SPIE 98*. 1998. p. 2–11.
- [49] BAZARAA, M. S.; SHERALI, H. D.; SHETTY, C. M. *Nonlinear Programming: Theory and Algorithms*. 2. ed. : John Wiley & Sons, 1993.
- [50] KATSAGGELOS, A. K. Iterative image restoration algorithms. *Optical Engineering, special issue on Visual Communications and Image Processing*, v. 28, n. 7, p. 735–748, jul. 1989.
- [51] SRIVASTAVA, A. Stochastic models for capturing image variability. *IEEE Signal Processing Magazine*, v. 19, n. 5, p. 63–76, set. 2002.
- [52] GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, p. 721–741, 1984.
- [53] STEVENSON, R. L.; SCHMITZ, B. E.; DELP, E. J. Discontinuity preserving regularization of inverse visual problems. *IEEE Transactions on Systems, Man and Cybernetics*, v. 24, n. 3, p. 455–469, mar. 1994.
- [54] BOUMAN, C.; SAUER, K. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on Image Processing*, v. 2, n. 3, p. 269–310, jul. 1993.
- [55] HAMZA, A. B.; KRIM, H.; UNAL, G. B. Unifying probabilistic and variational estimation. *IEEE Signal Processing Magazine*, v. 19, n. 5, p. 37–47, set. 2002.
- [56] AIZAWA, K.; KOMATSU, T.; SAITO, T. A scheme for acquiring very high resolution images using multiplecameras. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992*. 1992. v. 3, p. 289–292.
- [57] PAIK, J. et al. Image sequence interpolation for improving the resolution of the magnified image. In: *IEEE Asia Pacific Conference on Circuits and Systems*. 1996. p. 544–547.
- [58] SHIN, J.; CHOUNG, Y. C.; PAIK, J. High-resolution image sequence interpolation. In: *IEEE Region 10 Annual Conference on Speech and image technologies for computing and telecommunications*. 1997. v. 2, p. 781–784.
- [59] HONG, M.-C.; KANG, M. G.; KATSAGGELOS, A. K. An iterative weighted regularized algorithm for improving the resolution of video sequences. In: *IEEE International Conference on Image Processing*. 1997. v. 2, p. 474–477.

- [60] WANG, Z.; QI, F. On ambiguities in super-resolution modeling. *IEEE Signal Processing Letters*, v. 11, n. 8, p. 678–681, ago. 2004.
- [61] KYBIC, J.; BLU, T.; UNSER, M. Generalized sampling: a variational approach .i. theory. *IEEE Transactions on Signal Processing*, v. 50, n. 8, p. 1965–1976, ago. 2002.
- [62] KYBIC, J.; BLU, T.; UNSER, M. Generalized sampling: a variational approach .ii. applications. *IEEE Transactions on Signal Processing*, v. 50, n. 8, p. 1977–1985, ago. 2002.
- [63] BENEDETTO, J.; ZAYED, A. *Sampling, Wavelets, and Tomography*. : Birkhäuser, 2004. (Applied and Numerical Harmonic Analysis Series).
- [64] KOO, Y.; KIN, W. An image resolution enhancing technique using adaptive sub-pixel interpolation for digital still camera system. *IEEE Transactions on Consumer Electronics*, v. 45, n. 1, p. 118–123, fev. 1999.
- [65] NATTERER, F.; WÜBBELING, F. *Mathematical methods in image reconstruction*. : SIAM, 2001.
- [66] IRANI, M.; PELEG, S. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, v. 53, p. 231–239, maio 1991.
- [67] IRANI, M.; PELEG, S. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, v. 4, n. 4, p. 324–335, dez. 1993.
- [68] ZOMET, A.; RAV-ACHA, A.; PELEG, S. Robust super-resolution. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2001. v. 1, p. 645–650.
- [69] JIANG, Z.; WONG, T.-T.; BAO, H. Practical super-resolution from dynamic video sequences. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2003. v. 2, p. 549–554.
- [70] GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations*. : The Johns Hopkins University Press, 1996. (Johns Hopkins Series in the Mathematical Sciences).
- [71] COSTA, G. H.; BERMUDEZ, J. C. M. Statistical analysis of the lms algorithm applied to super-resolution image reconstruction. *IEEE Transactions on Signal Processing*, 2007. Aceito para publicação.
- [72] TEKALP, A.; OZKAN, M.; SEZAN, M. High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. 1992. v. 3, p. 169–172.
- [73] TEKALP, A. M. (Ed.). *Digital Video Processing*. : Prentice Hall, 1995. (Prentice Hall signal processing series).
- [74] PATTI, A. J.; SEZAN, M. I.; TEKALP, A. M. Robust methods for high-quality stills from interlaced video in the presence of dominant motion. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 7, n. 2, p. 328–342, abr. 1997.

- [75] FARSIU, S. et al. Fast and robust multi-frame super-resolution. *IEEE Transactions on Image Processing*, v. 13, n. 10, p. 1327–1344, out. 2004.
- [76] EREN, P. E.; SEZAN, M. I.; TEKALP, A. M. Robust, object-based high-resolution image reconstruction from low-resolution video. *IEEE Transactions on Image Processing*, v. 6, n. 10, p. 1446–1451, out. 1997.
- [77] REGINSKA, T. A regularization parameter in discrete ill-posed problems. *SIAM Journal on Scientific Computing*, v. 17, n. 3, p. 740–749, maio 1996.
- [78] BELGE, M.; KILMER, M. E.; MILLER, E. L. Efficient determination of multiple regularization parameters in a generalized l-curve framework. *Inverse Problems*, v. 18, n. 4, p. 1161–1183, ago. 2002.
- [79] MOLINA, R.; KATSAGGELOS, A. K.; MATEOS, J. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Transactions on Image Processing*, v. 8, n. 2, p. 231–246, fev. 1999.
- [80] MOHAMMAD-DJAFARI, A. Joint estimation of parameters and hyperparameters in a bayesian approach of solving inverse problems. In: *IEEE International Conference on Image Processing*. 1996. p. 473–472.
- [81] FARSIU, S. et al. Fast and robust super-resolution. In: *IEEE International Conference on Image Processing*. 2003. v. 2, p. 291–294.
- [82] THERRIEN, C. W. *Discrete Random Signals and Statistical Signal Processing*. : Prentice Hall, 1992. (Prentice Hall signal processing series).
- [83] SHEWCHUK, J. R. *An Introduction to Conjugate Gradient Method Without the Agoning Pain*. 1994.
- [84] SCHULTZ, R. R.; STEVENSON, R. L. A bayesian approach to image expansion for improved definition. *IEEE Transactions on Image Processing*, v. 3, n. 3, p. 233–242, maio 1994.
- [85] GAMAL, A. E.; ELTOUKHY, H. Cmos image sensors. *IEEE Circuits and Devices Magazine*, v. 21, n. 3, p. 6–20, maio 2005.
- [86] EL-KHAMY, S. et al. Efficient implementation of image interpolation as an inverse problem. *Digital Signal Processing*, v. 15, n. 2, p. 137–152, mar. 2005.
- [87] HORN, B. K. P.; SCHUNCK, B. G. Determining optical flow. *Artificial Intelligence*, v. 17, p. 185–203, ago. 1981.
- [88] ORAINTARA, S. et al. A method for choosing the regularization parameter in generalized tikhonov regularized linear inverse problems. In: *IEEE International Conference on Image Processing*. 2000.
- [89] GOLUB, G. H.; MATT, U. V. Generalized cross-validation for large-scale problems. *Journal of Computational and Graphical Statistics*, v. 6, n. 1, p. 1–34, 1997.

- [90] KANG, M. G.; KATSAGGELOS, A. K. General choice of the regularization functional in regularized image restoration. *IEEE Transactions on Image Processing*, v. 4, n. 5, p. 594–602, maio 1995.
- [91] HARDIE, R. C.; BARNARD, K. J.; ARMSTRONG, E. E. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, v. 6, n. 12, p. 1621–1633, dez. 1997.
- [92] MOHAMMAD-DJAFARI, A. A full bayesian approach for inverse problems. In: *15th International Workshop on Maximum Entropy and Bayesian Methods (MaxEnt95)*. 1995.
- [93] PAPOULIS, A. *Probability, Random Variables, and stochastic Processes*. : McGraw-Hill, 1991.
- [94] UNSER, M. A general hilbert space framework for the discretization of continuous signal processing operators. In: *SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing*. 1995. v. 2569-I, p. 51–61.
- [95] UNSER, M. Sampling-50 years after shannon. *Proceedings of the IEEE*, v. 88, n. 4, p. 569–587, abr. 2000.
- [96] UNSER, M.; ALDROUBI, A. A general sampling theory for nonideal acquisition devices. *IEEE Transactions on Signal Processing*, v. 42, n. 11, p. 2915–2925, nov. 1994.
- [97] UNSER, M. Splines: a perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, v. 16, n. 6, p. 22–38, nov. 1999.
- [98] STILLER, C.; KONRAD, J. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, v. 16, n. 4, p. 70–91, jul. 1999.
- [99] CAPEL, D. *Image Mosaicing and Super-resolution*. Tese (Doutorado) — University of Oxford, 2001.
- [100] BORMAN, S. *Topics in Multiframe Superresolution Restoration*. Tese (Doutorado) — University of Notre Dame, 2004.
- [101] HE, H.; KONDI, L. P. Map based resolution enhancement of video sequences using a huber-markov random field image prior model. In: *IEEE International Conference on Image Processing*. 2003. v. 2, p. 933–936.
- [102] DAI, Y. et al. Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal on Optimization*, v. 10, n. 2, p. 345–348, 1999.