

Aumento de Resolução de Imagens de Ressonância
Magnética do Trato Vocal Utilizadas em Modelos de
Síntese Articulatória

Ana Luísa Dine Martins Lemos

Orientadores: Prof. Dr. Nelson Delfino d'Ávila Mascarenhas e
Prof. Dr. Cláudio Alberto Torres Suazo

São Carlos/SP
Outubro/2011

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO**

ANA LUÍSA DINE MARTINS LEMOS

**AUMENTO DE RESOLUÇÃO DE IMAGENS DE RESSONÂNCIA
MAGNÉTICA DO TRATO VOCAL UTILIZADAS EM MODELOS
DE SÍNTESE ARTICULATÓRIA**

**Tese apresentado ao Programa de
Pós-Graduação em Biotecnologia,
para obtenção do título de doutor
em biotecnologia**

Orientação:
Prof. Dr. Nelson D. A. Mascarenhas
Prof. Dr. Cláudio A. T. Suazo

**SÃO CARLOS – SP
2011**

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

M386ar

Martins, Ana Luísa Dine.

Aumento de resolução de imagens de ressonância magnética do trato vocal utilizadas em modelos de síntese articulatória / Ana Luísa Dine Martins. -- São Carlos : UFSCar, 2011.

107 f.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2011.

1. Biotecnologia. 2. Processamento de imagens - técnicas digitais. 3. Restauração de imagens. 4. Reconstrução por super resolução. I. Título.

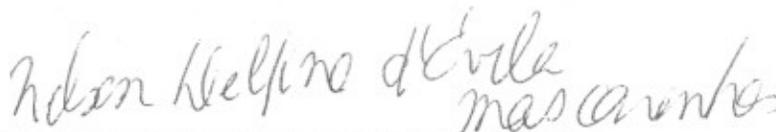
CDD: 660.6 (20^a)

Ana Luísa Dine Martins

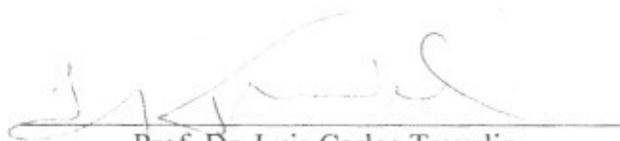
Tese de Doutorado submetida à
Coordenação do Programa de Pós-
Graduação em Biotecnologia, da
Universidade Federal de São
Carlos, como requisito parcial para
a obtenção do título de Doutor em
Biotecnologia

Aprovado em: 31/10/2011

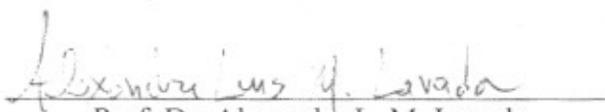
BANCA EXAMINADORA



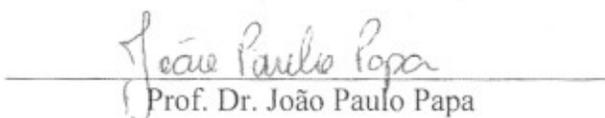
Prof. Dr. Nelson Delfino d'Avila Mascarenhas



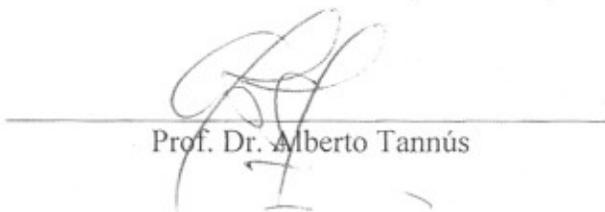
Prof. Dr. Luis Carlos Trevelin



Prof. Dr. Alexandre L. M. Levada



Prof. Dr. João Paulo Papa



Prof. Dr. Alberto Tannús

A Otávio, meus filhos, meus pais, meu irmão e meus avós.

Agradecimentos

À toda a minha família, em especial meu marido, Otávio, meus pais, Walter e Virginia, meu irmão, Raphael, e meus avós, Antonio e Orlanda, pelo amor, apoio e incentivos constantes para que eu pudesse alcançar mais este objetivo.

Ao Prof. Dr. Nelson D. A. Mascarenhas, meu orientador, pessoa por quem tenho uma enorme admiração, por toda a compreensão e paciência nesse período.

Aos meus companheiros de laboratório Michelle, Débora, Alexandre, Denis e Márcia, pelos momentos maravilhosos que compartilhamos durante os anos que residi em São Carlos.

A todos os professores e funcionários do Departamento de Computação da UFSCar, e a todas as pessoas que de alguma maneira contribuíram para a realização deste trabalho.

Agradeço à FAPESP pelo auxílio financeiro.

*O que o homem precisa de verdade não é um estado sem tensões
mas sim a luta e o empenho por um objetivo digno dele.*
(V. Frankl)

Sumário

Lista de Figuras	iii
Lista de Tabelas	vi
Resumo	vii
Abstract	viii
1 Introdução	1
1.1 Contextualização e Motivação	1
1.2 Objetivo	4
1.3 Organização	5
2 Revisão Bibliográfica	6
2.1 Síntese Articulatoria	6
2.1.1 Métodos de Imageamento	6
2.1.2 Estudos Dinâmicos baseados em MRI	9
2.2 Reconstrução de Imagens por Super-Resolução	11
2.2.1 Modelo de Formação das Imagens	13
2.2.2 Reconstrução por Super-Resolução aplicada a Vídeos	15
2.3 Métodos Propostos na Literatura	17
2.3.1 Abordagens no Domínio da Frequência	17
2.3.2 <i>Iterative Back Projection</i> - IBP	19
2.3.3 Interpolação Não-Uniforme e Restauração	21
2.3.4 <i>Projections Onto Convex Sets</i> - POCS	21
2.3.5 Modelagem Bayesiana do Problema	26
2.3.6 Interpolação Estatística	30
2.3.7 Filtro de Wiener Adaptativo	32
2.4 Registro das Imagens	34
2.4.1 Modelo de Transformação	36
2.4.2 Medida de Similaridade	42

3	Modelagem por Campos Aleatórios de Markov	44
3.1	Sites e Labels	45
3.2	Sistemas de Vizinhança e Cliques	46
3.3	Campos Aleatórios de Markov	48
3.4	Equivalência entre MRFs e a Distribuição de Gibbs	49
3.5	Algoritmo <i>Iterated Conditional Modes</i>	50
4	Estimador por Mínimo Erro Médio Quadrático	52
4.1	Estimador de Bayes	52
4.2	Princípio da Ortogonalidade	53
4.3	Estimador Linear Não Homogêneo	55
5	Método Proposto	57
5.1	Registro das Imagens	57
5.2	Aumento de Resolução Temporal	58
5.3	Aumento de Resolução Espacial	63
5.3.1	Abordagem Inicial	63
5.3.2	Métodos Baseados no Filtro de Wiener	68
6	Experimentos Comparativos	78
6.1	Procedimento estatístico	78
6.2	Aumento de Resolução Temporal	79
6.3	Aumento de Resolução Espacial	81
7	Conclusões e Trabalhos Futuros	92
7.1	Considerações Finais	92
7.2	Contribuições	96
7.3	Trabalhos Futuros	97
	Referências Bibliográficas	99

Lista de Figuras

1.1	Contornos de interesse para as pesquisas a respeito da produção da fala (adaptado de Bresch e Narayanan (2009)).	2
1.2	Ilustração de três grades de baixa resolução que possuem deslocamentos sub-pixel entre si.	4
2.1	Imageamento por (a) raios-X, (b) tomografia computadorizada, (c) ultrassom, (d) articulografia eletromagnética e (e) ressonância magnética (imagens retiradas de Stone (1997)).	9
2.2	Aquisição, reconstrução e análise utilizando (a) uma amostragem completa do espaço-k seguindo uma trajetória espiral e (b) uma sub-amostragem do espaço-k e a reconstrução utilizando a esparsidade dos dados (adaptado de Bresch <i>et al.</i> (2008)).	10
2.3	Ilustração de duas grades de baixa resolução que possuem deslocamentos inteiros entre si.	12
2.4	Modelo de formação das imagens de baixa resolução (adaptado de Katsaggelos <i>et al.</i> (2007)).	14
2.5	Obtenção de uma sequência de frames de alta resolução a partir de uma sequência de observações de baixa resolução (adaptado de Borman e Stevenson (1998)).	16
2.6	Pixels de baixa resolução que dependem de um pixel de alta resolução de acordo com o processo de aquisição das imagens de baixa resolução.	20
2.7	Interpolação dos pixels de baixa resolução espaçados não linearmente de acordo com o movimento estimado (adaptado de Park <i>et al.</i> (2003)).	22
2.8	Ilustração da convergência das projeções sequenciais para a intersecção dos conjuntos convexos fechados (adaptado de Wheeler <i>et al.</i> (2005)).	23
2.9	Ilustração da geometria que descreve a sobreposição dos pixels de alta resolução sobre um pixel de baixa resolução (adaptado de Stark e Yang (1998)).	24
2.10	Representação esquemática da re-amostragem proposta por Mascarenhas <i>et al.</i> (1996).	31
2.11	Distribuição dos pixels de baixa resolução na grade de alta resolução de acordo com os deslocamentos presentes entre eles (Hardie, 2007).	33
2.12	Exemplo de deformação presente nas imagens de MRI do trato vocal humano. .	35

2.13	Deformação da imagem Lena, utilizando FFD baseada em funções base B-spline.	39
3.1	Sistemas de vizinhança (a) de primeira ordem, (b) de segunda ordem e (c) de ordens mais altas.	47
3.2	Sistemas de vizinhança (a) de primeira ordem, (b) de segunda ordem e (c) de ordens mais altas.	47
4.1	Interpretação geométrica da ortogonalidade (adaptado de Papoulis (1984)).	54
5.1	Erro entre duas imagens após o registro utilizando o método proposto por Ruckert <i>et al.</i> (1999): (a) imagem de referência; (b) malha de pontos de controle uniformemente espaçados; (c) imagem a ser registrada de acordo com a imagem de referência; (d) malha de pontos de controle identifica após a aplicação do registro; (e) erro entre as imagens antes da aplicação do registro; (f) erro entre as imagens após a aplicação do registro.	59
5.2	(a) e (b) Frames adjacentes adquiridos durante a emissão de uma palavra. (c) Combinação das intensidades de mesma localização espacial desses dois frames, a fim de gerar um frame intermediário.	60
5.3	Ilustração da interpolação entre fatias adjacentes (adaptado de Penney <i>et al.</i> (2004)).	60
5.4	Ilustração da interpolação entre duas malhas de pontos de controle por meio de funções spline cúbicas.	61
5.5	Ilustração da aplicação do método de registro, sempre registrando a primeira imagem com relação às demais.	62
5.6	Movimento de um ponto de controle correspondente em 20 imagens consecutivas em uma sequência observada.	62
5.7	Aumento de resolução temporal através da soma ponderada das imagens vizinhas transformadas.	63
5.8	Ilustração de duas sub-amostragens de uma grade de alta resolução, provocando o deslocamento de ordem sub-pixel entre as grades de baixa resolução.	65
5.9	Ilustração de dois pixels observados que sobrepõem um pixel de alta resolução no alinhamento das imagens de baixa resolução com a grade de alta resolução.	66
5.10	(a) Sistema de vizinhança bidimensional de segunda ordem; (b) Dois sistemas de vizinhança tridimensionais.	68
5.11	Pixels observados dispostos na grade de alta resolução de acordo com os deslocamentos presentes entre eles.	70
5.12	Estimação de blocos de pixels de alta resolução.	71
5.13	Estimação de quatro imagens de alta resolução a partir do mesmo número de imagens de baixa resolução observadas.	73
5.14	Estimação de cada banda multitemporal.	73
5.15	Estimação de uma imagem de alta resolução a partir de um subconjunto das imagens de baixa resolução observadas, considerando o modelo de formação das imagens mostrado na Equação (5.23).	76
5.16	Ilustração das áreas dos pixels de baixa resolução utilizados na construção das matriz de covariância.	77

6.1	Recorte da região que concentra as deformações das primeiras quatro imagens de uma sequência observada.	80
6.2	(a) e (c) Detalhe das segunda e terceira imagens de uma sequência observada. (e) e (g) Malhas de pontos de controle correspondentes a (a) e (c), respectivamente. (b) e (d) Imagens geradas por interpolação linear e utilizando splines cúbicas na direção do movimento, respectivamente. (f) e (h) Malhas de pontos de controle correspondentes a (b) e (d), respectivamente.	81
6.3	Aumento de Resolução Temporal - Experimento 2 - Remoção das imagens duas a duas consecutivamente.	82
6.4	Experimento 3 - Sequência sub-amostrada temporalmente.	82
6.5	Boxplot dos dados obtidos nos experimentos (a) 1, (b) 2 e (c) 3.	83
6.6	Ilustração do processo de geração das imagens de baixa resolução simuladas.	83
6.7	Forma qualitativa das funções potenciais utilizadas na comparação. (a) Funções baseadas na diferença quadrática (GIMLL e DAMRF). (b) Modelo TV. (c) Modelo de Potts.	84
6.8	(a) Detalhe da imagem original utilizada nesse experimento. (b) Detalhe da interpolação bilinear da imagem de referência. Imagens reconstruídas pelos modelos (c) GIMLL; (d) DAMRF; (e) TV; e (f) Potts.	85
6.9	(a) Detalhe das imagens reconstruídas utilizando as abordagens (a) WI, (b) WM, (c) IEI, (d) IEM, (e) MTI, (f) MTM.	87
6.10	<i>Boxplot</i> das três propostas baseadas no filtro de Wiener.	88
6.11	<i>Boxplot</i> das três propostas baseadas no filtro de Wiener considerando os dois modelos utilizados para caracterizar as estruturas de correlação espacial: (a) filtro de Wiener adaptativo, (b) interpolação estatística e (c) abordagem multi-temporal.	88
6.12	Detalhe das imagens reconstruídas considerando que o registro não é perfeito: (a) GIMLL, (b) WM, (c) WI, (d) IEM, (e) IEI, (f) MTM, (g) MTI, e o filtro de Wiener adaptativo considerando a variância do ruído proporcional ao erro do registro (h) WM e (i) WI.	90

Lista de Tabelas

2.1	Métodos de imageamento utilizados para adquirir dados a respeito do trato vocal (adaptado de Bresch <i>et al.</i> (2008)).	8
6.1	Médias dos MSDs dos 3 experimentos.	80
6.2	NMSE das 7 imagens simuladas, reconstruídas utilizando o modelo GIMLL em comparação com a interpolação bilinear das imagens de baixa resolução simuladas e as imagens reconstruídas utilizando os outros modelos.	85
6.3	NMSE das 7 imagens simuladas, reconstruídas utilizando o modelo GIMLL em comparação com as propostas baseados no filtro de Wiener.	87
6.4	Tempo de Execução da reconstrução das 7 imagens simuladas, utilizando o modelo GIMLL em comparação com as propostas baseados no filtro de Wiener.	88
6.5	NMSE de imagens reconstruídas utilizando o modelo GIMLL e as abordagens baseadas no filtro de Wiener, considerando de 7 a apenas uma observação de baixa resolução.	89
6.6	NMSE de imagens reconstruídas considerando que o registro não é perfeito.	91

Resumo

A síntese articulatória procura produzir a fala através de modelos do trato vocal e dos processos articulatórios envolvidos. Os avanços no imageamento por ressonância magnética, permitiram que resultados importantes fossem alcançados com relação à fala e à forma do trato vocal. Entretanto um dos principais desafios ainda é a aquisição rápida e de alta qualidade das sequências de imagens. Além da opção de se utilizar meios de aquisição cada vez mais potentes, o que pode ser financeiramente inviável, abordagens propostas na literatura procuram aumentar a resolução modificando o processo de aquisição. Este trabalho propõe o aumento de resolução espaço-temporal das sequências adquiridas utilizando apenas técnicas de processamento de imagens digitais. A abordagem proposta é formada por duas etapas: o aumento de resolução temporal por meio de uma técnica de interpolação por compensação de movimento; e o aumento de resolução espacial por meio de uma técnica de reconstrução de imagens por super resolução. Com relação ao aumento de resolução temporal, dois métodos de interpolação são comparados: interpolação linear considerando duas imagens adjacentes e interpolação por splines cúbicas considerando quatro imagens consecutivas. Como, de acordo com os experimentos desenvolvidos, não existe diferença significativa entre esses dois métodos, a interpolação linear foi adotada por ser um procedimento mais simples e, conseqüentemente, apresentar menor custo computacional. O objetivo inicial para o aumento de resolução espacial das imagens observadas foi a extensão da abordagem proposta pela aluna em seu projeto de mestrado. Adotando uma abordagem de máxima probabilidade *a posteriori* (MAP), as imagens de alta resolução foram modeladas utilizando o modelo de campos aleatórios de Markov (MRF) *Generalized Isotropic Multi-Level Logistic* (GIMLL) e o algoritmo *Iterated Conditional Modes* (ICM) foi utilizado para maximizar as probabilidades condicionais locais sequencialmente. Entretanto, apesar de ter apresentado resultados promissores, devido à dimensão do problema tratado, o algoritmo ICM apresentou alto custo computacional. Considerando as limitações de performance desse algoritmo, decidiu-se adaptar o filtro de Wiener para o problema da reconstrução por super resolução. Utilizando dois trabalhos encontrados na literatura como inspiração, foram desenvolvidas três abordagens denominadas interpolação estatística, abordagem multitemporal e filtro de Wiener adaptativo. Em todos os casos, um modelo Markoviano separável e um modelo isotrópico foram comparados na caracterização das estruturas de correlação espacial. No caso da interpolação estatística e da abordagem multitemporal esses modelos foram utilizados para caracterizar as estruturas de correlação das observações e cruzada. Por outro lado, no caso da abordagem denominada filtro de Wiener adaptativo, esses modelos foram utilizados para caracterizar as estruturas de correlação espaciais *a priori*. De acordo com os experimentos desenvolvidos, o modelo isotrópico apresentou desempenho superior quando comparado ao modelo Markoviano separável. Além disso, considerando todas as propostas baseadas no filtro de Wiener e a proposta inicial baseada no modelo de Markov GIMLL, o filtro de Wiener adaptativo apresentou os melhores resultados e se mostrou mais rápido do que apenas uma iteração da abordagem baseada no modelo GIMLL.

Abstract

Articulatory Synthesis consists in reproducing speech by means of models of the vocal tract and of articulatory processes. Recent advances in Magnetic Resonance Imaging (MRI) allowed for important improvements with respect to the speech comprehension and the forms taken by the vocal tract. However, one of the main challenges in the field is the *fast* and at the same time *high-quality* acquisition of image sequences. Since adopting more powerful acquisition devices might be financially infeasible, a more feasible solution proposed in the literature is the resolution enhancement of the images by changes introduced in the acquisition model. This dissertation proposes a method for the spatio-temporal resolution enhancement of the obtained sequences using only digital image processing techniques. The approach involves two stages: (1) the temporal resolution enhancement by means of a motion compensated interpolation technique; and (2) the spatial resolution enhancement by means of a super resolution image reconstruction technique. With respect to the temporal resolution enhancement, two interpolation models are compared: linear interpolation considering two adjacent images and cubic splines interpolation considering four contiguous images. Since both models performed equally in the experiments, the linear interpolation was adopted, for its simplicity and lower computational cost. The initial goal of the spatial resolution enhancement was an extension of the candidate's approach proposed in her master's thesis. Adopting a maximum *a posteriori* probability approach (MAP), the high-resolution images were modeled using the Markov Random Fields (MRF) Generalized Isotropic Multi-Level Logistic (GIMLL) model and the Iterated Conditional Modes (ICM) algorithm. However, even though the approach has presented promising results, due to the dimension of the target problem, the algorithm presented high computational cost. Considering this limitation, an adaptation of the Wiener filter for the super-resolution reconstruction problem was considered. Inspired by two methods available in the literature, three approaches were proposed: the statistical interpolation, the multi-temporal approach, and the adaptive Wiener filter. In all cases, a separable Markovian model and an isotropic model were compared in the characterization of the spatial correlation structures. These models were used to characterize the correlation and cross correlation of observations for the statistical interpolation and the multi-temporal approach. On the other hand, for the adaptive Wiener filter, these models were used to characterize the *a priori* spatial correlation. According to the conducted experiments, the isotropic model outperformed the separable Markovian model. Besides, considering all Wiener filter-based approaches and the initial approach based on the GIMLL model, the adaptive Wiener filter outperformed all other approaches and was also faster than a single iteration of the GIMLL-based approach.

Introdução

1.1 Contextualização e Motivação

A síntese articulatória procura produzir a fala através de modelos do trato vocal e dos processos articulatórios envolvidos. Isso é feito por meio da modelagem da forma do trato vocal e do fluxo do ar que faz vibrar as cordas vocais. A Figura 1.1 evidencia os contornos de interesse para as pesquisas a respeito da produção da fala. Estão em evidência os seguintes componentes do aparelho fonador: laringe, epiglote, língua, lábios, paredes da faringe, glote, véu palatino e palato duro (céu da boca). De acordo com Bresch e Narayanan (2009), esses oito componentes anatômicos (com exceção do palato duro) são chamados articuladores da fala, os quais são controlados durante a produção da fala. O conhecimento a respeito da posição e movimentos desses articuladores é fundamental para as pesquisas a respeito da produção da fala.

Os avanços no imageamento por ressonância magnética (MRI - *Magnetic Resonance Imaging*), permitiram que resultados importantes fossem alcançados com relação à fala e à forma do trato vocal. Essa técnica de imageamento permite que todo o trato vocal seja representado em qualquer orientação, de forma estática ou dinâmica, sem nenhum tipo de efeito danoso conhecido à pessoa imageada. Desde o primeiro estudo proposto por Baer *et al.* (1991), muitas pesquisas foram conduzidas utilizando imagens adquiridas por MRI: estudos sobre a produção das vogais (Badin *et al.*, 1998; Demolin *et al.*, 2000); sobre a produção de consoantes (Narayanan *et al.*, 2004, 1995); e em diferentes línguas tais como francês (Badin e Serrurier, 2006; Demolin

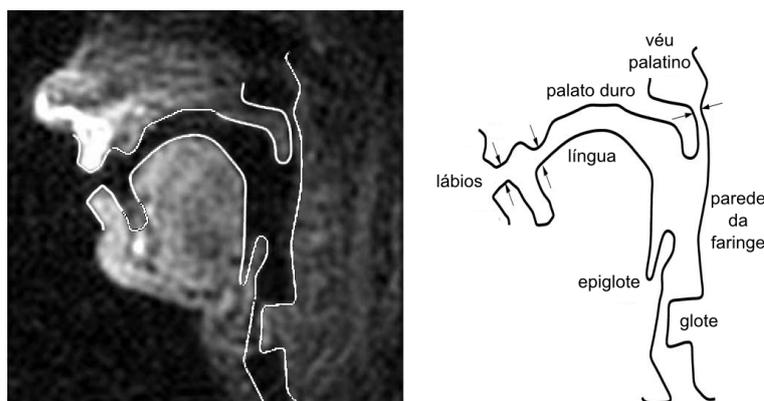


Figura 1.1: Contornos de interesse para as pesquisas a respeito da produção da fala (adaptado de Bresch e Narayanan (2009)).

et al., 1996), alemão (Behrends e Wismuller, 2001; Mády *et al.*, 2001), japonês (Kitamura *et al.*, 2005; Takemoto *et al.*, 2006), e português europeu (Martins *et al.*, 2008; Vasconcelos *et al.*, 2010).

Considerando os estudos dinâmicos, as imagens podem ser adquiridas de duas formas. Na técnica cine-MRI, várias aquisições do mesmo evento são adquiridas e utilizadas para reconstruir uma sequência de imagens de boa qualidade. Alguns trabalhos alcançaram resultados interessantes utilizando essa técnica (Stone *et al.*, 2001; Takemoto *et al.*, 2006). Entretanto, para alcançar uma boa qualidade em toda a sequência, a cine-MRI necessita de várias aquisições idênticas do mesmo evento, o que nem sempre é viável. A MRI de tempo real (RT-MRI - *real time MRI*) se refere à contínua aquisição de imagens com taxa de amostragem suficiente para capturar o movimento ou fisiologia de interesse. Entretanto, segundo Bresch *et al.* (2008), a RT-MRI da produção da fala humana enfrenta várias restrições importantes. Uma das principais é que a resolução espaço-temporal necessária para a identificação do movimento dos articuladores da fala varia de acordo com a velocidade e localidade desse movimento, e tal informação não é conhecida *a priori*. Assim, um dos principais desafios relacionados ao processamento do sinal adquirido é a aquisição rápida e de alta qualidade de sequências de imagens, de forma que seja possível detectar características relevantes em cada uma das imagens, além da variação da forma do trato vocal em toda a sequência.

Além da opção de se utilizar meios de aquisição cada vez mais potentes, o que pode ser financeiramente inviável, abordagens propostas na literatura procuram aumentar a resolução das sequências de imagens melhorando o processo de aquisição. Utilizando um sistema de ressonância magnética de 1.5T equipado com gradientes rápidos (CompactPlus, PowerTrak 6000, Philips Medical Systems, Best, The Netherlands), Demolin *et al.* (2000) adaptaram uma implementação rápida da técnica Turbo Spin Echo (TSE) a fim de alcançar monitoramento contínuo

e dinâmico do trato vocal com resolução temporal de 4 imagens por segundo. Entretanto, de acordo com Santa-Marta *et al.* (2004), a estratégia de imageamento dessa técnica influencia a qualidade das imagens. O borramento provocado pode fazer com que pequenas estruturas cheguem a desaparecer. Além disso, a resolução temporal alcançada é muito inferior à necessária para capturar as características dinâmicas da língua. Narayanan *et al.* (2004) propuseram uma abordagem na qual o espaço k é amostrado seguindo uma trajetória espiral. Essa estratégia permite taxas de aquisição de 8-9 imagens por segundo e taxas de reconstrução de 20-24 imagens por segundo. Porém, existem algumas limitações inerentes a essa estratégia: borramento devido ao off-resonance e reconstrução não trivial já que a grade é não-Cartesiana. Bresch *et al.* (2008) executam a detecção dos contornos de interesse mostrados na Figura 1.1, utilizando uma amostragem esparsa do espaço k seguindo a trajetória espiral. Segundo os autores, abordagens convencionais incluem os seguintes estágios: reconstrução de imagens livres de artefatos; cálculo da magnitude das imagens reconstruídas; detecção dos contornos de interesse; e cálculo dos parâmetros que descrevem o trato vocal. Essas abordagens produzem dados intermediários de alta qualidade, para só então atingir seu objetivo. Dessa forma, os autores propõem que os contornos de interesse sejam identificados com base nas amostras do espaço k.

No contexto das imagens utilizadas em estudos de síntese articulatória, não foi encontrado nenhum trabalho na literatura que proponha o aumento de resolução espaço-temporal das sequências adquiridas utilizando os meios de aquisição existentes. A reconstrução de imagens por super-resolução (SRIR - *Super Resolution Image Reconstruction*) é uma alternativa interessante para esse problema. Trata-se de uma metodologia que utiliza apenas técnicas de processamento de imagens digitais aplicadas às imagens observadas a fim de reconstruir sequências de maior resolução (Park *et al.*, 2003). Para que isso seja possível, as imagens observadas devem possuir deslocamentos de ordem sub-pixel (deslocamentos por uma fração da dimensão do pixel observado). Essa restrição permite que existam informações adicionais em cada uma das imagens observadas. Essas informações podem ser utilizadas para aumentar a resolução como ilustrado na Figura 1.2. Imagens que possuem essa característica podem ser adquiridas por meio de várias aquisições da mesma câmera, utilizando múltiplas câmeras localizadas em diferente posições, por meio de movimento da cena ou dos objetos de interesse, por sistemas de imageamento vibratórios ou utilizando frames de um vídeo. As imagens de ressonância magnética adquiridas durante a emissão da fala são semelhantes a frames de um vídeo. Como comprovado empiricamente, elas possuem deslocamentos de ordem sub-pixel entre si, o que viabiliza o aumento de resolução por meio de técnicas de SRIR.

Com isso, considerando que esse problema não foi explorado e a continuação do trabalho de mestrado realizado (Martins *et al.*, 2009a, 2007, 2009b), abre-se um nicho de pesquisa interessante e importante a ser explorado para as pesquisas em síntese articulatória. De fato,

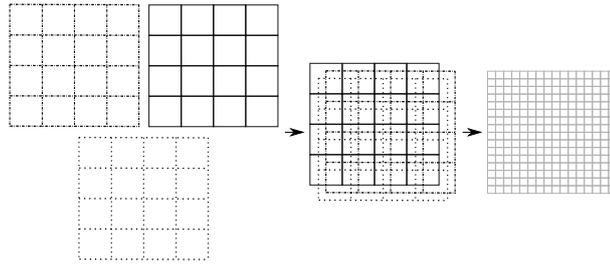


Figura 1.2: Ilustração de três grades de baixa resolução que possuem deslocamentos sub-pixel entre si.

uma abordagem de SRIR aplicada a esse contexto pode diminuir as exigências a respeito da resolução alcançada durante a aquisição das sequências de imagens. No melhor caso, não são necessários meios de aquisição mais potentes ou intervenções no processo de aquisição.

1.2 Objetivo

O objetivo desta tese é propor uma abordagem de SRIR computacionalmente eficiente, para o aumento de resolução espaço-temporal de sequências de imagens de ressonância magnética do trato vocal utilizadas em estudos de síntese articulatória. No projeto de mestrado, adotando um framework Bayesiano, as imagens de alta resolução foram modeladas utilizando campos aleatórios de Markov (MRF - *Markov Random Field*), especificamente o modelo de Potts (Martins *et al.*, 2007). O algoritmo *Iterated Conditional Modes* (ICM) foi utilizado para maximizar as funções densidade de probabilidades condicionais locais. No contexto das imagens do trato vocal, a mesma abordagem, porém utilizando o modelo de Markov *Generalized Isotropic Multi-Level Logistic* (GIMLL), alcançou resultados promissores (Martins *et al.*, 2011). Entretanto, devido ao número de imagens e à dimensão de cada uma das imagens, o algoritmo ICM apresentou custo computacional excessivamente elevado.

No contexto das imagens de MRI do trato vocal, um dos principais desafios é a aquisição rápida e de alta qualidade de sequências de imagens (Bresch e Narayanan, 2009). Uma abordagem capaz de melhorar a qualidade das imagens, porém com alto custo computacional, não seria a mais adequada. Considerando as limitações de performance do algoritmo ICM, iniciou-se a investigação de uma abordagem não iterativa que fornecesse uma solução de mínimo erro médio quadrático. A matriz que caracteriza o filtro de Wiener é escolhida de forma que o erro médio quadrático seja minimizado e trata-se de uma abordagem não iterativa. Dessa forma, decidiu-se adaptar o filtro de Wiener para o problema da reconstrução por super-resolução. Foram encontrados dois trabalhos na literatura que propõem abordagens para o aumento de resolução espacial baseadas no filtro de Wiener (Hardie, 2007; Mascarenhas *et al.*, 1996). Variações des-

ses dois trabalhos considerando o contexto das imagens do trato vocal foram investigadas e, em comparação com a abordagem baseada no algoritmo ICM, chegou-se a resultados superiores com custo computacional bastante reduzido.

1.3 Organização

O Capítulo 2 apresenta a revisão bibliográfica a respeito das pesquisas em síntese articulatória, discutindo os meios de aquisição de imagens utilizados, em especial MRI e os estudos dinâmicos desenvolvidos. Além disso, é apresentada uma descrição do problema da SRIR, assim como as principais abordagens propostas na literatura. Finalmente, o problema do registro das imagens é descrito, e seus principais componentes são apresentados.

Os Capítulos 3 e 4 fornecem a fundamentação teórica para o entendimento das abordagens de SRIR investigadas. Primeiramente o Capítulo 3 discute a modelagem por MRFs seguida da utilização do algoritmo ICM como método de otimização. Esse procedimento foi adotado na proposta inicial para o aumento de resolução espacial das imagens. Por fim, o Capítulo 4 discute a estimação por mínimo erro médio quadrático, critério buscado pelo filtro de Wiener. As três propostas finais para o aumento de resolução espacial das sequências se baseiam nessa estimação.

O Capítulo 5 descreve em detalhes a metodologia proposta para o aumento de resolução espaço-temporal das sequências de imagens do trato vocal. A abordagem proposta se divide em duas etapas: o aumento de resolução temporal por meio de uma técnica de interpolação por compensação de movimento, e o aumento de resolução espacial por meio de uma técnica de SRIR. Ambas as etapas dependem da estimação do movimento presente entre as observações. Dessa forma primeiramente a aplicação do método de registro adotado é detalhada. Em seguida é apresentado o método adotado para a geração das imagens intermediárias às imagens observadas, em função das transformações estimadas pelo método de registro. A abordagem inicialmente explorada para o aumento de resolução espacial das imagens é apresentada e suas principais deficiências são discutidas. Finalmente são descritas as três variações baseadas no filtro de Wiener, exploradas como soluções mais eficientes para o problema de aumento de resolução espacial.

O Capítulo 6 descreve os experimentos desenvolvidos, os resultados alcançados, e os procedimentos estatísticos utilizados para avaliar esses resultados. Finalmente, o Capítulo 7 apresenta as conclusões e considerações finais, bem como as perspectivas futuras para a continuidade das pesquisas relacionadas ao projeto.

Revisão Bibliográfica

2.1 Síntese Articulatória

Pesquisas a respeito da produção da fala sob o aspecto articulatório tem por objetivo explicar como as pessoas utilizam os órgãos do aparelho fonador para produzir os diferentes sons que compõem a fala. A produção da fala é resultante da combinação de uma fonte de energia sonora modulada por uma função de transferência (filtro) determinada pela forma do trato vocal. De fato, de acordo com a teoria acústica da produção da fala, o trato vocal pode ser visto como um tubo acústico com área da seção transversal variável (Fant, 1960). Assim, informação espacial e temporal a respeito da forma e dimensões tomadas pelo trato vocal é essencial para entender e modelar os processos articulatórios envolvidos na produção da fala.

2.1.1 Métodos de Imageamento

Diferentes tipos de dados estão disponíveis para as pesquisas em síntese articulatória. Algumas pesquisas incluem informações adquiridas por meio de tomografia computadorizada (CT - *Computed Tomography*), articulografia eletromagnética (EMA - *Electromagnetic Articulography*), raios-X, ultrassom, e MRI. A Tabela 2.1 mostra algumas vantagens e desvantagens inerentes a cada um desses métodos de imageamento. Inicialmente, o imageamento por raios-X foi bastante utilizado a fim de validar as primeiras teorias a respeito da produção da fala (Chiba

e Kajiyama, 1941; Fant, 1960; Heinz e Stevens, 1964). As áreas das seções transversais eram aproximadas com base nas projeções laterais, o que é uma das debilidades dessa técnica de imageamento. Além disso, estruturas formadas por tecidos não rígidos, como a língua, são identificadas com dificuldade. Isso ocorre porque todas as estruturas rígidas que estão no caminho do feixe de raios, incluindo dentes, mandíbula e vértebras, são imageadas de forma marcante (Figura 2.1(a)). De qualquer forma, seu uso decaiu consideravelmente principalmente devido aos riscos de radiação ionizante envolvidos. Mesmo assim, como as sequências possuem uma boa informação dinâmica a respeito da produção da fala (resolução temporal por volta de 60 imagens por segundo), os bancos de sequências de imagens construídos ainda são utilizados (Iskarous, 2005; Jallon e Berthommier, 2009; Munhall *et al.*, 1995). A CT utiliza raios-X para adquirir imagens de fatias do corpo imageado. Imagens adquiridas no plano coronal são formadas por meio da projeção de feixes de raios-X muito finos a partir de muitas origens. O scanner rotaciona ao redor do objeto adquirindo essas imagens, e um computador cria uma composição de todas elas. Dessa forma, em comparação com a imageamento por raios-X, a CT é capaz de imagear tecidos não rígidos de forma mais clara. Porém, essa técnica também fornece riscos de radiação, o que fez com que esse método raramente fosse utilizado nas pesquisas a respeito da articulação da fala.

O ultrassom produz uma imagem utilizando as propriedades de reflexão das ondas de som. Elas são refletidas ao atingir uma interface formada por um tecido de densidade diferente, como ossos ou passagens de ar. Essa técnica de imageamento normalmente é utilizada para estudar o movimento da língua (Akgul *et al.*, 1999; Shawker *et al.*, 1983; Stone e Lundberg, 1996). A Figura 2.1(c) mostra uma imagem de ultrassom da língua no plano sagital. Como é possível notar, quando as ondas de som atingem o ar na superfície da língua, elas refletem de volta formando uma linha branca. A área mais escura imediatamente abaixo é o corpo da língua. Uma grande limitação do ultrassom é a impossibilidade de imagear além de interfaces como tecido-ar ou tecido-osso. De acordo com Stone (1997), como a interface tecido-ar na superfície da língua reflete a onda de som, estruturas que estão atrás não podem ser imageadas. Fato semelhante ocorre quando o ultrassom atinge um osso. Áreas escuras são criadas na imagem, impossibilitando a localização exata desse osso.

A EMA é um dispositivo de rastreamento de pontos dentro e ao redor da cavidade oral por meio de campos magnéticos alternados. A Figura 2.1(d) mostra um sistema de EMA tridimensional. Um cubo acrílico é posicionado ao redor da cabeça do locutor. Esse cubo possui seis transmissores que produzem correntes magnéticas em diferentes frequências. Na cavidade oral são fixados sensores nos articuladores de interesse. Cada sensor produz uma corrente alternada que varia de acordo com sua distância aos seis transmissores. Duas limitações de sistemas de rastreamento de pontos são: sensores precisam ser fixados às estruturas, possivelmente inter-

ferindo na produção da fala; e apenas pontos são rastreados, assim o comportamento de todos os articuladores da fala é apenas inferido. Essa última limitação é especialmente problemática para as estruturas formadas por tecidos não rígidos, como os lábios, a língua, etc. A maior vantagem desse tipo de sistema é a alta taxa de rastreamento e sua habilidade em rastrear várias articuladores ao mesmo tempo.

Tabela 2.1: Métodos de imageamento utilizados para adquirir dados a respeito do trato vocal (adaptado de Bresch *et al.* (2008)).

Método	Vantagens	Desvantagens	Comentários
CT	Alta resolução espacial e temporal; capaz de capturar as estruturas da faringe; possibilidade de imageamento 3D.	Exposição à radiação.	Método raramente utilizado nesse contexto.
EMA	Alta resolução espacial e temporal; possibilidade de imageamento 3D.	Fornecer dados espacialmente esparsos; não é capaz de capturar as estruturas da faringe.	Método frequentemente utilizado nesse contexto.
Raio-X	Alta resolução espacial e temporal.	Exposição à radiação; imagens mostram apenas uma projeção do volume imageado, o que dificulta a extração de contornos; fornece dados espacialmente esparsos.	Método raramente utilizado nesse contexto; bancos de dados existentes ainda são utilizados nas pesquisas.
Ultrassom	Alta resolução temporal; método não invasivo e seguro; áudio de boa qualidade pode ser adquirido durante o imageamento.	Fornecer imagens ruidosas; detecta apenas o primeiro limite entre tecido e ar; e o detector fica em contato com a mandíbula, o que pode afetar a articulação da fala.	Utilizado à princípio para imagear a língua.
MRI	Método não invasivo e seguro; captura estruturas da faringe; possibilidade de imageamento 3D.	Resolução espacial e temporal relativamente pobre; alto custo financeiro; limitado a pessoas que não possuem muitos dos tratamentos dentários ou implantes; requer que a pessoa imageada se mantenha em posição supina; dificuldade na aquisição de áudio de boa qualidade durante o imageamento.	Método emergente nesse contexto.

O MRI é uma ferramenta poderosa para a obtenção de dados a respeito da geometria do trato vocal. Essa técnica de imageamento não acarreta em nenhum risco conhecido de radiação e as imagens possuem uma boa relação sinal/ruído. Elas são próprias para a modelagem tridimensional e, como essa técnica permite uma boa separação entre diferentes estruturas, a área e volume do caminho percorrido pelo ar podem ser calculados diretamente (Figura 2.1(e)). Entretanto, um dos grandes desafios nesse contexto é a possibilidade de examinar as mudanças na forma do trato vocal durante a emissão da fala. Devido à baixa resolução temporal e ao longo tempo de exposição requerido, inicialmente as pesquisas se limitaram a estudos meramente estáticos (Baer *et al.*, 1991; Dang *et al.*, 1993; Narayanan *et al.*, 1999, 1995; Ong e Stone, 1998). Essas pesquisas foram muito importantes já que forneceram informações desconhecidas a respeito da morfologia tridimensional do trato vocal, mas ainda era necessário visualizar as mudanças na forma do trato vocal durante a fala de palavras e fonemas em tempo

real. Nos últimos anos, o aumento da resolução temporal dos meios de aquisição em conjunto com estratégias rápidas de imageamento viabilizaram a realização de vários estudos dinâmicos.

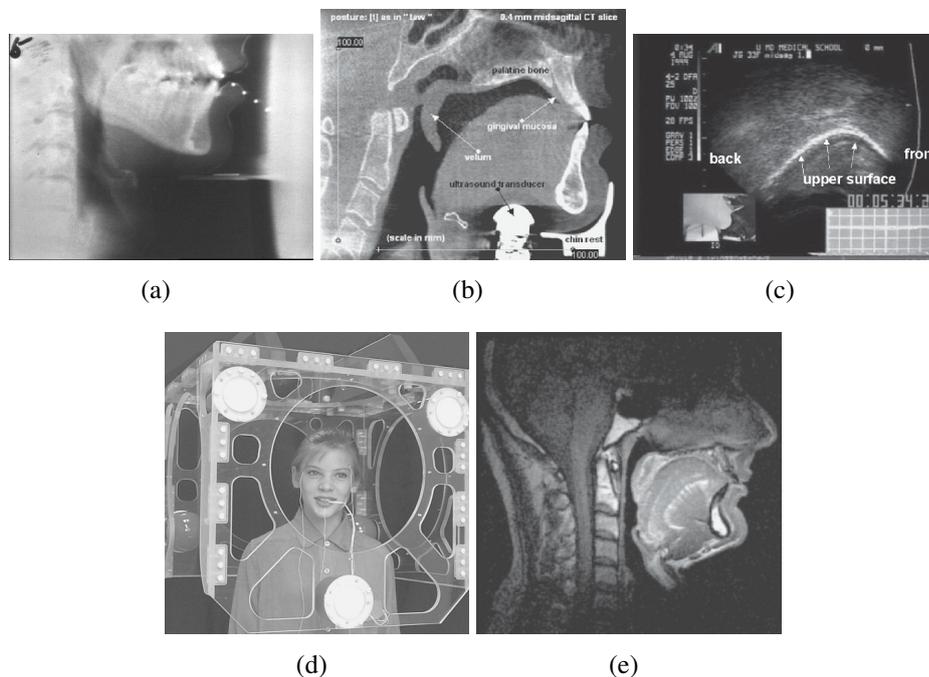


Figura 2.1: Imageamento por (a) raios-X, (b) tomografia computadorizada, (c) ultrassom, (d) artrografia eletromagnética e (e) ressonância magnética (imagens retiradas de Stone (1997)).

2.1.2 Estudos Dinâmicos baseados em MRI

O aumento de resolução temporal é desafiado pela queda na relação sinal/ruído e pelo aparecimento de artefatos. Como a MRI soma emissões de prótons ao longo do tempo, a reconstrução de uma única imagem normalmente é demorada e a coleta de dados durante a emissão da fala é um desafio. O cine-MRI consiste em várias aquisições do mesmo evento. Dados são somados para cada frame ao longo das repetições, possibilitando a formação de uma sequência de imagens de boa qualidade. Apesar de permitir a captura de informações importantes a respeito das mudanças na forma do trato vocal (Mathiak *et al.*, 2000; Takemoto *et al.*, 2006), essa técnica requer repetições com pouquíssimas variações entre elas a fim de prevenir borramento nas imagens, o que nem sempre é viável.

A coleta rápida de frames de MRI considerando apenas uma repetição de um evento foi alcançada utilizando a amostragem espiral do espaço-k (Narayanan *et al.*, 2004). A aquisição, reconstrução e análise seguindo essa metodologia são ilustradas na Figura 2.2(a). Entretanto,

como as amostras não caem em uma grade Cartesiana, é necessária uma estratégia de reconstrução modificada. Além disso, as taxas de amostragem alcançadas são suficientes para analisar apenas algumas articulações. Segundo Bresch *et al.* (2008), só é possível aumentar essas taxas de amostragem ao se empregar métodos acelerados como: imageamento paralelo, filtragem temporal ou *compressive sensing*.

A técnica *compressive sensing* (Candes *et al.*, 2006; Donoho, 2006) fornece um arcabouço teórico no qual um sinal pode ser recuperado ao se minimizar a norma-L1 de uma representação esparsa em um sistema linear sub-determinado. A maioria dos sinais naturais são compressíveis, o que equivale a dizer que eles podem ser representados em um domínio no qual o sinal é esparsa. De acordo com Schultz *et al.* (2009), técnicas de aquisição de imagens normalmente realizam amostragem seguida de compressão, ou seja, amostram a uma taxa alta e em seguida descartam a maior parte da informação adquirida seguindo um esquema de compressão que explora a representação esparsa do sinal. Nesse contexto, a *compressive sensing* surge como um novo paradigma de aquisição de dados. Ela apresenta um algoritmo estável e robusto que permite amostrar em taxas muito menores que o limite de Nyquist e mesmo assim recuperar o sinal livre de artefatos causados por *aliasing*. Dessa forma, o processo de aquisição é consideravelmente acelerado. A Figura 2.2(b) ilustra a aquisição, reconstrução e análise de uma imagem do trato vocal utilizando *compressive sensing*.

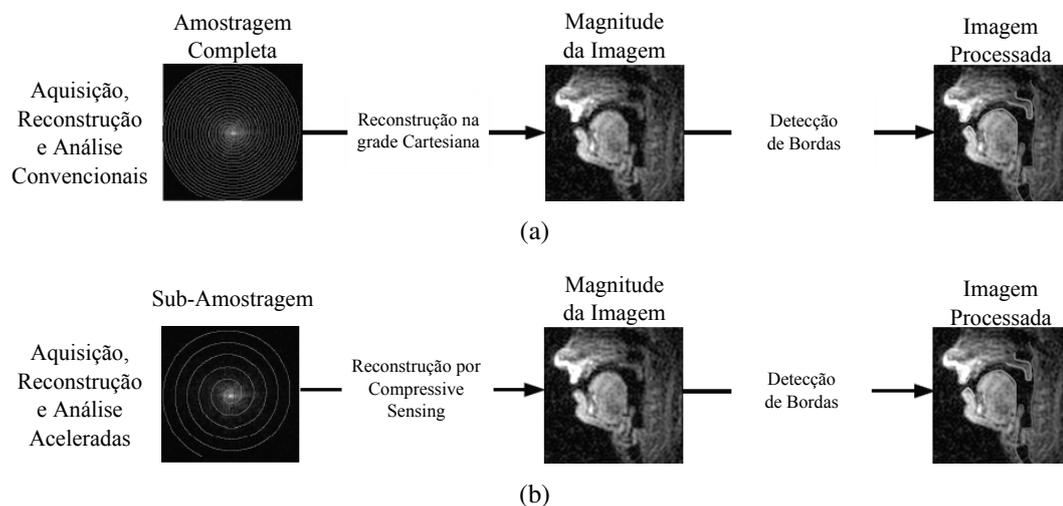


Figura 2.2: Aquisição, reconstrução e análise utilizando (a) uma amostragem completa do espaço-k seguindo uma trajetória espiral e (b) uma sub-amostragem do espaço-k e a reconstrução utilizando a esparsidade dos dados (adaptado de Bresch *et al.* (2008)).

De qualquer forma, apesar de ser uma metodologia muito promissora, a *compressive sensing* é dependente da base de representação esparsa e da escolha da estratégia de amostragem no

espaço-k, as quais ainda podem depender da aplicação. Além disso, ainda existe a limitação do hardware utilizado para executar a otimização.

Aparentemente não existe nenhum trabalho na literatura que proponha o aumento de resolução espaço-temporal das sequências já adquiridas com taxa de amostragem aquém da necessária para identificar o movimento articulatorio sendo estudado. A SRIR apresenta uma solução interessante para esse problema. Nessa metodologia são aplicadas técnicas de processamento de imagens digitais às imagens observadas a fim de reconstruir sequências de maior resolução (Park *et al.*, 2003). A seguir é apresentada a fundamentação teórica dessa metodologia e alguns dos principais métodos propostos na literatura.

2.2 Reconstrução de Imagens por Super-Resolução

A necessidade por sequências de imagens com alta resolução espaço-temporal é frequente em várias aplicações baseadas em imagens digitais. Detalhes a respeito de pequenas estruturas fornecem informação adicional na análise de imagens. O diagnóstico baseado em imagens médicas é mais preciso quando se possui uma alta resolução. Por outro lado, a identificação de objetos em imagens de sensoriamento remoto é muito mais precisa quando imagens de alta resolução estão disponíveis. Além disso, o sistema de televisão digital tem apresentado uma demanda crescente e significativa por imagens e vídeos de alta-resolução espaço-temporal. Esse tipo de imagens pode ser adquirido por dispositivos de aquisição de alta resolução. Entretanto, existem várias limitações financeiras e tecnológicas (Park *et al.*, 2003) (Levin e Hoffman, 1999). A maioria dos sensores de alta resolução são muito caros. Além disso, uma forma de aumentar a taxa de amostragem do sinal, e conseqüentemente a resolução espacial das imagens digitais, é aumentar o número de células foto-elétricas e diminuir sua área no sensor, aumentando a densidade de células. Porém, existe um limite para o tamanho da célula a partir do qual a imagem é degradada pela presença de *shot-noise* (Chaudhuri, 2001). Dessa forma, a reconstrução de imagens de alta resolução utilizando apenas técnicas de processamento de imagens digitais é um assunto de grande interesse.

A SRIR tenta reconstruir uma imagem, ou conjunto de imagens, de alta resolução a partir de um conjunto de observações de baixa resolução da mesma cena. Para que isso seja possível, as imagens observadas devem possuir deslocamentos de ordem sub-pixel entre si, ou seja, deslocamentos por uma fração da dimensão do pixel de baixa resolução nas direções vertical e/ou horizontal. Essa restrição permite que existam informações adicionais em cada observação, as quais podem ser utilizadas para aumentar a resolução espacial como ilustrado na Figura 1.2. A Figura 2.3 ilustra duas grades de baixa resolução que possuem deslocamentos da ordem de

um pixel entre elas. É possível notar que, com deslocamentos inteiros, existe apenas a mesma informação replicada em cada uma das observações. Imagens que possuem deslocamentos de ordem sub-pixel entre si podem ser adquiridas por meio de várias aquisições da mesma câmera, utilizando múltiplas câmeras localizadas em diferentes posições, por meio de movimento da cena ou dos objetos de interesse, por sistemas de imageamento vibratórios, utilizando frames de um vídeo, etc. (Park *et al.*, 2003).

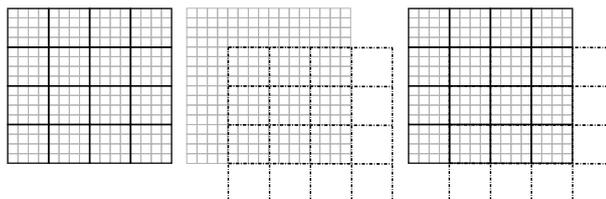


Figura 2.3: Ilustração de duas grades de baixa resolução que possuem deslocamentos inteiros entre si.

Tsai e Huang (1984) foram os primeiros a propor uma solução para o problema de se reconstruir uma imagem de alta resolução a partir de várias observações de baixa resolução da mesma cena. Eles utilizaram uma abordagem no domínio da frequência, baseada na propriedade de translação no espaço da transformada de Fourier, para modelar translação global da cena. Recentemente, várias abordagens para resolver o mesmo problema foram propostas, a maioria delas no domínio do espaço. Mesmo que as abordagens no domínio da frequência sejam mais simples, existem muitas desvantagens nessa formulação (Borman e Stevenson, 1998). Por exemplo, em geral essas abordagens não permitem modelos de deslocamentos mais genéricos. Abordagens no domínio do espaço possibilitam modelos de deslocamentos arbitrários, modelos de degradações complexos e, principalmente, a inserção de conhecimento *a priori* no processo de reconstrução. É importante notar que, de maneira similar aos problemas de restauração de imagens digitais, a reconstrução por super-resolução é um problema mal-condicionado. Dessa forma, soluções regularizadas que utilizam restrições *a priori* são necessárias e vários métodos que seguem essa formulação foram propostos. Abordagens baseadas em projeções em conjuntos convexos (POCS - *Projection onto convex sets*) impõem conhecimento *a priori* por meio de conjuntos convexos fechados (Stark e Oskoui, 1989; Wheeler *et al.*, 2005; Yeh e Stark, 1990). Essa é uma abordagem computacional baseada na teoria de que projeções iterativas em conjuntos convexos fechados convergem para a intersecção de todos os conjuntos. Entretanto, apesar de sua flexibilidade e simplicidade, se a intersecção dos conjuntos não for um único ponto, haverá mais de uma solução. Assim, o resultado dependerá da estimativa inicial. Além disso, essa abordagem demanda alto poder computacional. Por outro lado, métodos de regularização determinísticos utilizam informação a respeito da solução para estabilizar a inversão do

problema (Bose *et al.*, 2001; Hong *et al.*, 1997). Suavidade é a restrição mais comum entre os métodos existentes. Assume-se que, em geral, imagens apresentam atividade limitada nas altas frequências. Entretanto, em muitos casos, outros tipos de restrições preservariam as informações de alta frequência de forma mais adequada. Consequentemente, essas abordagens modelam a informação *a priori* de maneira desfavorável. Elas apenas incluem um termo de regularização na função de otimização. Técnicas de reconstrução probabilísticas incluem conhecimento *a priori* de forma mais natural, e a estimação Bayesiana de máxima probabilidade *a posteriori* (MAP - *Maximum a posteriori probability*) é o método mais promissor. Essa abordagem utiliza a função densidade de probabilidades *a priori* da imagem para impor restrições à solução. Nesse contexto, modelos *a priori* de MRFs são considerados os mais flexíveis e realísticos, já que eles permitem a inclusão de conhecimento utilizando apenas relações entre pixels vizinhos (Borman e Stevenson, 1998). Mesmo quando se possui informação *a priori* limitada, essas formulações permitem a imposição de características usuais de imagens digitais. No contexto da super-resolução, em geral existe apenas um conjunto de observações ruidosas e de baixa resolução da mesma cena. Assim, uma restrição usual a ser imposta é a suavidade. Em uma abordagem MAP-MRF, essa restrição é expressa por meio da probabilidade *a priori* da imagem de alta resolução a ser reconstruída, a qual é unicamente determinada por suas probabilidades condicionais locais (Besag, 1974). Além disso, em MRFs apenas pixels vizinhos possuem interação direta. Assim, a restrição de suavidade pode ser imposta simplesmente considerando que os valores de pixels vizinhos não devem mudar abruptamente. Entretanto, apesar dessa simplicidade, a maximização da probabilidade conjunta normalmente demanda alto poder computacional. Além disso, a estimação dos parâmetros do modelo de Markov é um problema computacionalmente inviável (Martins *et al.*, 2009b), a otimização global é difícil de ser implementada exatamente, e uma aproximação sempre se faz necessária (Li *et al.*, 1995). Nesse contexto, o algoritmo *Iterated Conditional Modes* (ICM) é uma alternativa interessante. Trata-se de um algoritmo iterativo proposto por Besag (1974), o qual maximiza as probabilidades condicionais locais sequencialmente. Um ponto interessante a respeito desse algoritmo é sua rápida convergência. Em geral são necessárias apenas seis iterações até o algoritmo convergir.

2.2.1 Modelo de Formação das Imagens

Para analisar de forma coerente o problema da reconstrução por super-resolução, primeiramente é preciso formular o modelo de formação das imagens, que relaciona a imagem de alta resolução desejada às imagens de baixa resolução observadas. Considerando $f[i, j]$, $0 \leq i, j \leq M$, a imagem ideal não degradada, amostrada da cena contínua de interesse na taxa de Nyquist, $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, em uma situação real, essa imagem sofre borramento pelo sistema ótico e é

corrompida por ruído durante sua aquisição. Seguindo a notação lexicográfica de f , o operador de borramento normalmente é considerado um operador linear invariante no espaço H_k , $M^2 \times M^2$, cujos elementos são dados por amostras da função de espalhamento pontual (PSF - *Point Spread Function*) do sistema. Assim, uma versão de baixa resolução borrada e ruidosa g_k , $N^2 \times 1$, $N \leq M$, da imagem de alta resolução f , $M^2 \times 1$, pode ser modelada como

$$g_k = D_k W_k H_k f + n_k, \quad (2.1)$$

onde n_k é o ruído na observação g_k , seguindo um modelo aditivo. W_k , $M^2 \times M^2$, modela uma transformação normalmente denominada *warping*, que provoca translações, rotações ou qualquer mapeamento mais sofisticado causando deslocamentos na imagem de alta resolução borrada $H_k f$. D_k , $N^2 \times M^2$, é o operador de sub-amostragem ou decimação e a composição dos operadores W_k e D_k é responsável pelos deslocamentos de ordem sub-pixel presentes na k -ésima imagem de baixa resolução g_k . A Figura 2.4 ilustra o modelo de formação mostrado na Equação (2.1).

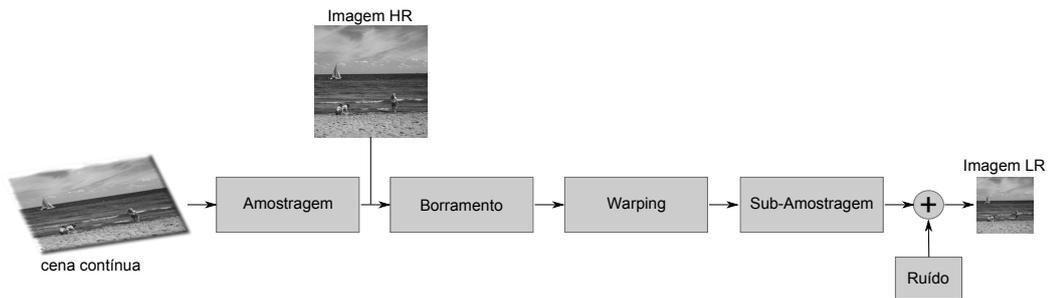


Figura 2.4: Modelo de formação das imagens de baixa resolução (adaptado de Katsaggelos *et al.* (2007)).

De acordo com Katsaggelos *et al.* (2007), existem dois tipos de modelo de formação das imagens de baixa resolução na literatura, os modelos *Warp-Blur* e *Blur-Warp*, os quais dependem da ordem em que o borramento e a transformação geométrica, H_k e W_k , são aplicados na geração da imagem de baixa resolução. No modelo *Warp-Blur*, primeiramente a imagem de alta resolução sofre a transformação geométrica e posteriormente é borrada:

$$g_k = D_k H_k W_k f + n_k. \quad (2.2)$$

Segundo os autores, esse modelo de formação foi primeiramente formulado em Irani e Peleg (1993) e posteriormente utilizado em vários trabalhos (Borman e Stevenson, 1998; Elad e Feuer, 1997; Eren *et al.*, 1997; Farsiu *et al.*, 2003, 2004; Patti *et al.*, 1997; Tekalp *et al.*, 1992). O modelo *Blur-Warp*, que considera que a imagem sofre primeiro o borramento, é mostrado na

Equação (2.1). Wang e Qi (2004) discutem qual desses modelos é mais adequado. Segundo os autores, as transformações geométricas responsáveis pelos deslocamentos sub-pixel normalmente são causadas por movimento relativo entre a câmera e a cena. Além disso, o borramento pode ser causado por esse movimento, pela própria câmera ou devido à atmosfera entre a câmera e a cena. Quando o borramento é predominantemente causado pela câmera, ele ocorre após os deslocamentos, e dessa forma, o modelo *Warp-Blur* estaria correto. Por outro lado, se o borramento predominante for devido à atmosfera entre a câmera e o objeto, ele ocorrerá antes dos deslocamentos e o modelo *Blur-Warp* estaria correto. Dessa forma, ambos os modelos podem estar de acordo com o processo físico de aquisição das imagens de baixa resolução, dependendo de que tipo de borramento é predominante. Na maioria dos trabalhos existentes na literatura considera-se que o borramento ocorre devido à PSF do sistema, e, conseqüentemente, o modelo *Warp-Blur* deveria ser adotado. Entretanto, como os deslocamentos causados pelo operador W_k normalmente não são conhecidos *a priori*, e, por isso, são aproximados pela estimação dos deslocamentos presentes nas observações de baixa resolução, o modelo *Warp-Blur* pode causar erros. Dessa forma, Wang e Qi (2004) concluem que o modelo *Blur-Warp* é mais adequado e apresenta melhores resultados, apesar de não refletir exatamente a realidade.

2.2.2 Reconstrução por Super-Resolução aplicada a Vídeos

Os modelos de formação das imagens de baixa resolução apresentados consideram a reconstrução de apenas uma imagem de alta resolução f , a partir da qual são formadas as observações de baixa resolução g_k , $k = 1, \dots, q$. Entretanto, em geral as técnicas de SRIR tradicionais que buscam reconstruir apenas uma imagem de alta resolução a partir de um conjunto de observações de baixa resolução, podem ser aplicadas na reconstrução de seqüências de alta resolução. Isso pode ser feito utilizando todas as (ou um sub-conjunto das) observações, na estimação de cada imagem de alta resolução, o que normalmente é implementado por meio de uma abordagem de janela deslizante como ilustrado na Figura 2.5 (Borman e Stevenson, 1998; Katsaggelos *et al.*, 2007). Nesse contexto, o método de estimação dos deslocamentos presentes entre as imagens observadas é a variação mais crítica. Dependendo da aplicação, esse método deve ser capaz de identificar deslocamentos globais ou locais, oclusão (pixels que existem em uma imagem, mas não em outras), deslocamentos não únicos, etc. No caso da reconstrução de uma imagem de alta resolução de uma cena estática, diferenças no registro de cada uma das imagens se devem principalmente a diferenças na posição da câmera. Em contrapartida, no caso de vídeos cada observação representa um momento diferente de uma cena dinâmica no tempo. Essa característica implica em várias complicações, já que objetos podem sofrer deformações; objetos diferentes podem se mover de forma diferente; movimento em profundidades diferentes pode

acarretar em sobreposições e, conseqüentemente, oclusão; características de um mesmo objeto podem se modificar ao longo do tempo; etc. (Prendergast e Nguyen, 2008).

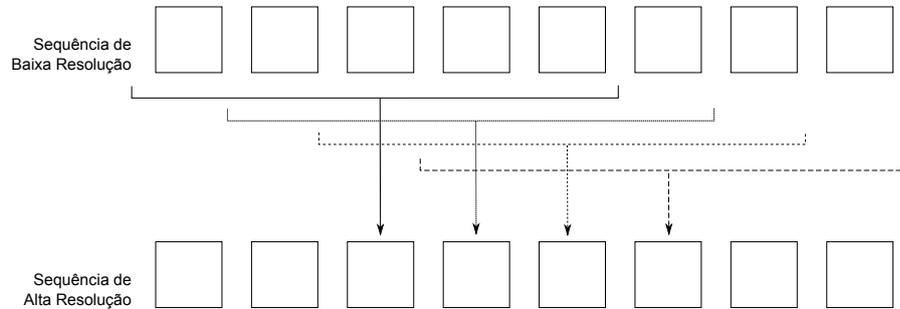


Figura 2.5: Obtenção de uma sequência de frames de alta resolução a partir de uma sequência de observações de baixa resolução (adaptado de Borman e Stevenson (1998)).

Seguindo Katsaggelos *et al.* (2007), o modelo de formação dos frames de baixa resolução é definido da seguinte forma. Considerando $f_l[i, j]$ uma cena dinâmica no espaço e contínua no tempo, devidamente amostrada na taxa de Nyquist em espaço e tempo, sendo $l = 1, \dots, L$ e $0 \leq i, j \leq M$ as coordenadas temporal e espaciais discretas, respectivamente, assume-se que f_1, \dots, f_L satisfazem

$$f_l[i, j] = f_k [i + d_{l,k}^x[i, j], j + d_{l,k}^y[i, j]], 1 \leq l, k \leq L, 0 \leq i, j \leq M, \quad (2.3)$$

onde $d_{l,k}^x[i, j]$ e $d_{l,k}^y[i, j]$ são os componentes horizontal e vertical do vetor de deslocamentos. É importante notar que esse modelo permite deslocamentos locais e globais na cena, já que os pixels se movem de forma independente. Além disso, a identificação dos vetores de deslocamentos $d_{l,k}^x[i, j]$ e $d_{l,k}^y[i, j]$ para $0 \leq i, j \leq M$, é equivalente à identificação do fluxo ótico da cena. Entretanto, é importante ressaltar que como os frames de alta resolução não estão disponíveis, esses vetores de deslocamentos são estimados com base nos frames de baixa resolução.

Seguindo a notação lexicográfica de f_l e f_k , ambas de dimensão $M^2 \times 1$, e do vetor de deslocamentos $d_{l,k}$, de dimensão $M^2 \times 2$, e considerando $C(d_{l,k})$, $M^2 \times M^2$, a matriz que mapeia o frame f_k para o frame f_l , a Equação (2.3) pode ser reescrita como

$$f_l = C(d_{l,k})f_k, 1 \leq l, k \leq L. \quad (2.4)$$

Considerando que para cada frame de baixa resolução será reconstruído um frame de alta resolução correspondente, de maneira semelhante ao caso da reconstrução de apenas uma imagem de alta resolução (apenas omitindo o operador de *warping*, já que os frames são corresponden-

tes), o modelo de formação dos frames de baixa resolução pode ser dado por

$$g_l = D_l H_l f_l + n_l, 1 \leq l \leq L, \quad (2.5)$$

onde, novamente, n_l é o ruído na observação g_l , H_l é formada por amostras da PSF do sistema e D_l é o operador de sub-amostragem.

Nesse contexto, de acordo com a relação mostrada na Equação (2.4), o frame de baixa resolução g_l está relacionado ao frame de alta resolução f_k da seguinte forma

$$g_l = D_l H_l C(d_{l,k}) f_k + n_l + \mu_{l,k}, 1 \leq l, k \leq L, \quad (2.6)$$

onde $\mu_{l,k}$ representa o ruído inerente ao registro das imagens.

O modelo apresentado na Equação (2.6) corresponde ao modelo *Warp-Blur* discutido anteriormente. O modelo *Blur-Warp* correspondente é dado por

$$g_l = D_l C(d_{l,k}) H_l f_k + n_l + \mu_{l,k}, 1 \leq l, k \leq L. \quad (2.7)$$

Além da aplicação das técnicas tradicionais à reconstrução de sequências de alta resolução, foram propostas as técnicas sequenciais e os métodos simultâneos. As técnicas sequenciais, apesar de também reconstruírem uma imagem de cada vez, sempre aproveitam as informações já estimadas em cada reconstrução (Elad e Feuer, 1999; Mateos *et al.*, 2000). Essas técnicas foram desenvolvidas com o objetivo de reduzir o custo computacional das abordagens existentes, entretanto a qualidade das estimativas foi inferior (Zibetti, 2007). Em contrapartida, os métodos simultâneos reconstróem toda a sequência de alta resolução em um único passo (Borman e Stevenson, 1999). Eles foram desenvolvidos a fim de se alcançar uma qualidade superior por meio de uma melhor utilização da informação sobre o movimento presente nas observações. Entretanto, o custo computacional em geral é maior do que o das técnicas tradicionais.

A seguir são descritas algumas das principais abordagens de SRIR propostas na literatura.

2.3 Métodos Propostos na Literatura

2.3.1 Abordagens no Domínio da Frequência

Tsai e Huang (1984) publicaram o primeiro trabalho utilizando várias imagens da mesma cena, a fim de gerar uma imagem de resolução espacial mais alta do que a presente em qualquer uma das imagens observadas. Foram utilizadas imagens de um satélite Landsat adquiridas a cada

18 dias, que invariavelmente possuíam deslocamentos espaciais uma em relação às demais. Os autores utilizaram uma abordagem no domínio da frequência com base na propriedade de deslocamento no espaço da Transformada Contínua de Fourier (CFT - *Continuous Fourier Transform*), segundo a qual deslocamentos no espaço são equivalentes ao deslocamento em fase no domínio de Fourier

$$f(x + \delta_k^x, y + \delta_k^y) \longleftrightarrow F(u, v)e^{j2\pi(\delta_k^x u + \delta_k^y v)}, \quad (2.8)$$

sendo $F(u, v)$ a CFT de $f(x, y)$.

Considerando $f(x, y)$ a cena original contínua, translações globais levam a q imagens deslocadas

$$f_k(x, y) = f(x + \delta_k^x, y + \delta_k^y), k = 1, \dots, q. \quad (2.9)$$

Nesse contexto, de acordo com a Equação (2.8), a CFT de $f_k(x, y)$, representada por $F_k(u, v)$, é dada por

$$F_k(u, v) = e^{j2\pi(\delta_{k,1}u + \delta_{k,2}v)} F(u, v). \quad (2.10)$$

A amostragem de $f_k(x, y)$, $k = 1, \dots, q$, idealmente por um trem de impulsos, dá origem às q imagens de baixa resolução observadas

$$g_k[i, j] = f(iT_x + \delta_k^x, jT_y + \delta_k^y), k = 1, \dots, q, \quad (2.11)$$

sendo T_x e T_y os períodos de amostragem nas direções horizontal e vertical, respectivamente. Denotando a Transformada Discreta de Fourier (DFT - *Discrete Fourier Transform*) de $g_k[i, j]$ por $G_k[u, v]$, a CFT da cena original e a DFT das imagens amostradas estão relacionadas pela propriedade de *aliasing* por

$$G_k[u, v] = \frac{1}{T_x T_y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} F_k\left(\frac{u}{NT_x} + \frac{m}{T_x}, \frac{v}{NT_y} + \frac{n}{T_y}\right). \quad (2.12)$$

Se $f(x, y)$ é limitada em banda, existem $\Omega_x, \Omega_y \in \mathbb{N}$ tais que $F(u, v) = 0$ para $|u| \geq \Omega_x/T_x$ e $|v| \geq \Omega_y/T_y$ e a soma infinita na Equação (2.12) se reduz a uma soma finita. Assim, com base na propriedade de deslocamento no espaço mostrada na Equação (2.8), a Equação (2.12) pode ser escrita na forma de matriz

$$\mathbf{G} = \Phi \mathbf{F}. \quad (2.13)$$

onde o k -ésimo elemento de \mathbf{G} , $q \times 1$, contém os coeficientes da DFT da imagem observadas $g_k[i, j]$, $G_k[u, v]$. Φ é uma matriz que relaciona as DFTs das observações às amostras da CFT de $f(x, y)$, $F(u, v)$, contidas em \mathbf{F} . Para finalizar, resolvendo a Equação (2.13) para \mathbf{F} e utilizando a DFT inversa, chega-se à imagem de alta resolução reconstruída.

Tsai e Huang (1984) não consideraram ruído e borramento em sua abordagem. Trabalhos posteriores, estenderam essa abordagem considerando a presença de borramento e ruído nas observações, e considerando diferentes tipos de borramento em cada uma das imagens de baixa resolução (Kim *et al.*, 1990; Kim e Su, 1993). Apesar da simplicidade oferecida pelo domínio da frequência, em geral não é possível utilizar modelos de deslocamentos mais genéricos. Por isso, atualmente a maior parte das abordagens são propostas no domínio do espaço.

2.3.2 Iterative Back Projection - IBP

Irani e Peleg (1991) propuseram uma abordagem iterativa simples no domínio do espaço, semelhante a técnicas de retroprojeção utilizadas na reconstrução de imagens via projeções em tomografia computadorizada. Considerando q imagens de baixa resolução observadas $g_k[i, j]$, $0 \leq i, j \leq N$, $k = 1, \dots, q$, e algum conhecimento a respeito do processo de aquisição dessas imagens, pretende-se reconstruir uma imagem de alta resolução da cena imageada $f[k, l]$, $0 \leq k, l \leq M$, $N < M$. Assumindo que existe uma estimativa inicial de f , é possível simular o processo de aquisição de g_k , para $k = 1, \dots, q$, e comparar as imagens simuladas com as observações encontrando um resíduo. Logicamente, quanto menor esse resíduo, mais próxima a estimativa de alta resolução está de f . Na abordagem de Irani e Peleg, esse resíduo é utilizado para melhorar a estimativa de f por meio de um processo denominado retroprojeção (*back projection*), que é repetido até que o resíduo seja mínimo. Dessa forma a abordagem se divide basicamente em duas etapas: simulação do processo de aquisição das imagens de baixa resolução e retroprojeção do erro entre as simulações e observações, atualizando a estimativa de alta resolução.

Nesse contexto, considerando g as imagens de baixa resolução empilhadas g_k , seguindo uma notação lexicográfica, e A o operador que relaciona a cena de alta resolução às observações, o processo de formação das imagens pode ser dado por

$$g = Af, \quad (2.14)$$

sendo que, como no modelo de formação mostrado nas equações (2.1) e (2.2), A geralmente modela borramento, warping e sub-amostragem. Dada uma estimativa da imagem de alta resolução $f^{(n)}$, as simulações de baixa resolução $g^{(n)}$ são encontradas por

$$g^{(n)} = Af^{(n)}, \quad (2.15)$$

e o erro entre as observações e as simulações é dado por

$$e^{(n)} = \sqrt{\sum_{k=1}^q \sum_{i,j} \left(g_k[i, j] - g_k^{(n)}[i, j] \right)^2}. \quad (2.16)$$

O algoritmo de retroprojeção iterativa (IBP - *Iterative Back Projection*) atualiza a estimativa de alta resolução por meio de um operador de retroprojeção A^{BP} , da seguinte forma

$$f^{(n+1)} = f^{(n)} + A^{\text{BP}} (g - g^{(n)}), \quad (2.17)$$

sendo que A^{BP} aproxima a inversa de A . É importante notar que, como discutido em Borman (2004), essa iteração é bastante semelhante à iteração de Landweber (Landweber, 1951) utilizada na resolução de problemas inversos mal-condicionados.

A atualização de cada pixel da estimativa de alta resolução $f^{(n)}[k, l]$ ocorre de acordo com todos os pixels de baixa resolução $g_k[i, j]$, $0 \leq i, j \leq N$, $k = 1, \dots, q$, que dependem de seu valor de acordo com o processo de aquisição das imagens. A dependência entre dois pixels de baixa resolução em imagens diferentes e um único pixel de alta resolução é ilustrada na Figura 2.6.

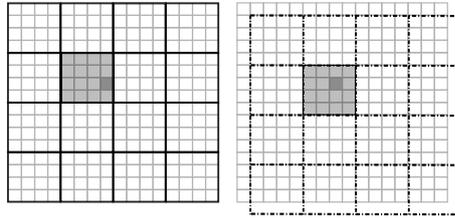


Figura 2.6: Pixels de baixa resolução que dependem de um pixel de alta resolução de acordo com o processo de aquisição das imagens de baixa resolução.

A contribuição de um pixel de baixa resolução é dada pelo erro $g_k[i, j] - g_k^{(n)}[i, j]$ multiplicado por um fator que mede a contribuição relativa do pixel de alta resolução $f[k, l]$ para esse pixel de baixa resolução $g_k[i, j]$. Dessa forma, $f^{(n+1)}[k, l]$ é dado por

$$f^{(n+1)}[k, l] = f^{(n)}[k, l] + \sum_{k=1}^q \sum_{i,j} (g_k[i, j] - g_k^{(n)}[i, j]) h^{\text{BP}}, \quad (2.18)$$

sendo h^{BP} o chamado kernel de retroprojeção, o qual modela a contribuição de $f[k, l]$ para cada pixel de baixa resolução $g_k[i, j]$. De acordo com Borman e Stevenson (1998), apesar dessa abordagem ser simples e eficiente, e condicionar a solução às observações, ela não garante uma solução única para o problema mal-condicionado da reconstrução por super-resolução. Além

disso, apesar dos autores argumentarem que a imposição de conhecimento *a priori* no processo de reconstrução pode ser inserida na definição do kernel de retroprojeção, isso não ocorre com facilidade.

2.3.3 Interpolação Não-Uniforme e Restauração

A interpolação não-uniforme seguida de restauração é uma das abordagens computacionalmente e intuitivamente mais simples dentre as metodologias de SRIR presentes na literatura (Alam *et al.*, 2000; Nguyen e Milanfar, 2000; Ur e Gross, 1992). Um método de registro é utilizado para identificar os deslocamentos de ordem sub-pixel presentes entre as observações de baixa resolução, sempre considerando uma delas como referência. Em seguida, os pixels de baixa resolução são distribuídos sobre uma grade comum na resolução desejada de acordo com os deslocamentos identificados. Como pode ser visto na Figura 2.7, nessa sobreposição as amostras observadas normalmente estarão espaçadas de forma não uniforme e será preciso um método de interpolação não linear para alcançar uma estimativa inicial de alta resolução. Como a interpolação das amostras não trata o borramento devido à PSF do sistema, ela geralmente é seguida pela aplicação de um método qualquer de restauração. Seja $g = [g_1^T, \dots, g_q^T]^T$ todos os pixels de baixa resolução observados, a estimativa de alta resolução proveniente da interpolação desses pixels de acordo com o deslocamento entre eles é dada por

$$\hat{h}_k = E_k g, \quad (2.19)$$

sendo E_k a matriz de interpolação. A estimativa inicial de alta resolução \hat{h}_k pode ser restaurada da seguinte forma

$$\hat{f}_k = \tilde{H}_k^{-1} \hat{h}_k, \quad (2.20)$$

onde \tilde{H}_k^{-1} é a inversa aproximada ou regularizada da PSF H_k .

A maioria dos métodos de interpolação-restauração tratam os dois processos de forma completamente separada. Entretanto, quando o movimento entre as imagens observadas é distribuído de maneira pobre, o resultado da interpolação geralmente apresentará *aliasing*. O processo de restauração independente só intensifica esse efeito. Nesses casos, a união dos dois processos em um único passo pode ser bastante vantajoso.

2.3.4 Projections Onto Convex Sets - POCS

Como discutido anteriormente, a reconstrução por super-resolução é um problema mal-condicionado. Dessa forma, soluções regularizadas são necessárias. Abordagens baseadas em pro-

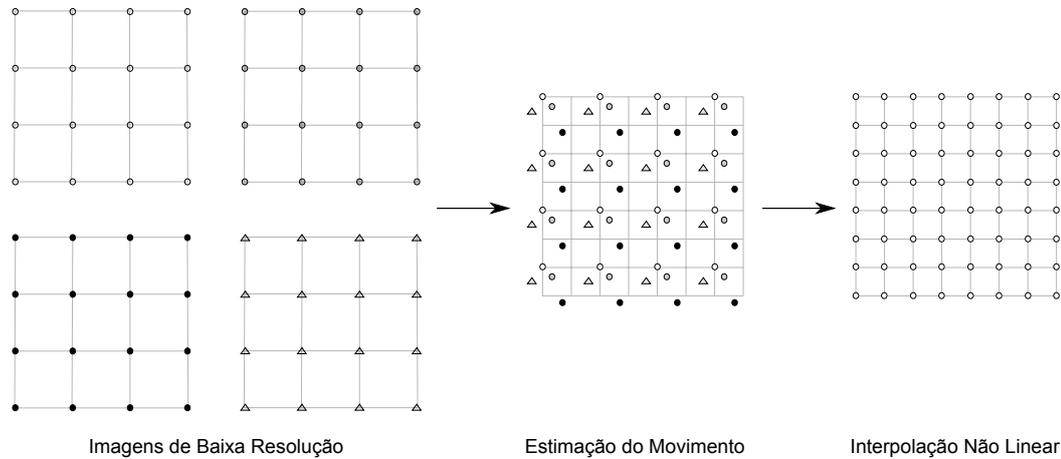


Figura 2.7: Interpolação dos pixels de baixa resolução espaçados não linearmente de acordo com o movimento estimado (adaptado de Park *et al.* (2003)).

jeções em conjuntos convexos (POCS) impõem conhecimento *a priori* por meio de conjuntos convexos fechados. Trata-se de uma técnica iterativa robusta, cuja característica principal é a facilidade na incorporação de restrições à solução no processo de reconstrução. Stark e Oskoui (1989) foram os primeiros a utilizarem essa abordagem para solucionar o problema da reconstrução por super-resolução e, posteriormente, Tekalp *et al.* (1992) estenderam o problema para incorporar a presença de ruído nas observações, além de incluírem a estimação do movimento presente entre elas.

Seguindo Stark e Yang (1998), a aplicação do método de POCS para resolver um problema prático segue a seguinte forma. Deseja-se determinar uma quantidade desconhecida sobre a qual se conhece alguma informação na forma de restrições. A quantidade desconhecida é tratada como um vetor no espaço de Hilbert, e as restrições são descritas na forma de conjuntos convexos fechados nesse espaço. Assim, assume-se que existem q conjuntos $C_k, k = 1, \dots, q$, cada um modelando uma restrição (entretanto é importante notar que é possível modelar mais de uma restrição com apenas um conjunto). Intuitivamente, a intersecção de todos esses conjuntos $C_0 = \cap_{i=1}^q C_i$, contém todas as possíveis soluções para o problema, já que todos os seus elementos satisfazem todas as restrições impostas pelos conjuntos $C_k, k = 1, \dots, q$.

A teoria fundamental de POCS afirma que, considerando T_k o projetor relacionado ao conjunto C_k , para $k = 1, \dots, q$, a iteração

$$x_{n+1} = T_q T_{q-1} \dots T_1 x_n, \tag{2.21}$$

com x_0 arbitrário, converge fracamente para um ponto em C_0 , como ilustrado na Figura 2.8. Nesse contexto, o sucesso dessa abordagem depende da definição apropriada dos conjuntos de restrição $C_k, k = 1, \dots, q$.

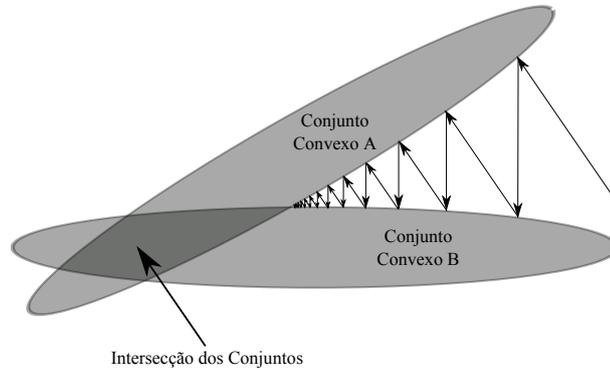


Figura 2.8: Ilustração da convergência das projeções sequenciais para a intersecção dos conjuntos convexos fechados (adaptado de Wheeler *et al.* (2005)).

Seguindo a solução proposta por Stark e Oskoui (1989) e Tekalp *et al.* (1992) para o problema da reconstrução por super-resolução, assume-se que existem q frames de baixa resolução $g_k, N \times N, k = 1, \dots, q$, a partir dos quais se pretende reconstruir uma imagem de alta resolução $f, M \times M, M > N$. Assim existem M^2 incógnitas e qN^2 equações. Na presença de degradações inerentes ao processo de aquisição das imagens observadas, as imagens de baixa resolução estão relacionadas à imagem de alta resolução por

$$g_k[i, j] = \sum_{m=1}^M \sum_{n=1}^M f[m, n] h_k[i, j; m, n] + n[i, j], k = 1, \dots, q, \quad (2.22)$$

onde $h[i, j; m, n]$ é a PSF (que pode ser variante ou invariante no espaço) da configuração do detector e do sistema ótico da câmera. Dessa equação é possível afirmar que, intuitivamente, o pixel de baixa resolução $g_k[i, j]$ é formado por uma superposição ponderada de pixels de alta resolução. Como ilustrado na Figura 2.9, assume-se que os pixels de alta resolução $f[m, n], m, n = 1, \dots, M$, são quadrados com lado igual a Δ , e que os pixels de baixa resolução $g_k[i, j], i, j = 1, \dots, N$ possuem área igual a $d^2 \Delta^2$. Pela figura é possível notar que apenas uma pequena parte dos pixels de alta resolução sobrepõe um pixel de baixa resolução, dentre os quais alguns terão toda sua área sobre o pixel de baixa resolução, e alguns terão apenas parte de sua área sobrepondo o pixel de baixa resolução. Nesse contexto, os autores utilizam um parâmetro $r_k[i, j; m, n], 0 \leq r_k[i, j; m, n] \leq 1$, o qual modela a contribuição do pixel de alta resolução $f[m, n]$ para o pixel de baixa resolução $g_k[i, j]$, de acordo com a porção de sua área sobrepondo $g_k[i, j]$.

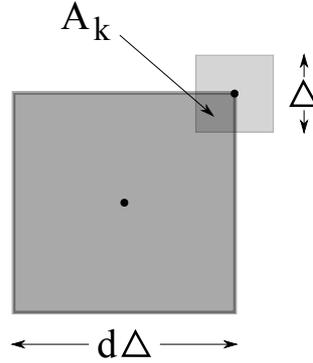


Figura 2.9: Ilustração da geometria que descreve a sobreposição dos pixels de alta resolução sobre um pixel de baixa resolução (adaptado de Stark e Yang (1998)).

Chamando de A_k a área de sobreposição, e de A_{LR} a área do pixel de baixa resolução, de acordo com o parâmetro $r_k[i, j; m, n]$, a PSF $h_k[i, j; m, n]$ pode ser modelada como

$$h_k[i, j; m, n] = \frac{r_k[i, j; m, n] \Delta^2}{d^2 \Delta^2} = \frac{A_k[i, j; m, n]}{A_{LR}}. \quad (2.23)$$

Dessa forma, conhecendo os deslocamentos entre as observações, é possível saber o valor de parâmetro de contribuição $r_k[i, j; m, n]$, e, conseqüentemente, é possível definir completamente a PSF $h_k[i, j; m, n]$. Ignorando a presença de ruído e reescrevendo a Equação (2.22) da seguinte forma

$$g_k[i, j] - \sum_{m=1}^M \sum_{n=1}^M f[m, n] h_k[i, j; m, n] = 0, \quad (2.24)$$

como não conhecemos apenas a imagem de alta resolução $f[m, n]$, $m, n = 1, \dots, M$, podemos escolher diferentes funções $\hat{f}[m, n]$, de forma que o resíduo

$$\epsilon'_k[i, j; \hat{f}] = g_k[i, j] - \sum_{m=1}^M \sum_{n=1}^M \hat{f}[m, n] h_k[i, j; m, n], \quad (2.25)$$

seja mínimo. Na presença de ruído esse resíduo é dado por

$$\epsilon_k[i, j; \hat{f}] = \epsilon'_k[i, j; \hat{f}] + n[i, j]. \quad (2.26)$$

De acordo com Stark e Yang (1998), uma possível restrição para a solução $\hat{f}[m, n]$ é que a magnitude do resíduo seja limitada, ou seja,

$$C_k[i, j] = \left\{ \hat{f} : \left| \epsilon_k[i, j; \hat{f}] \right| \leq \eta \right\}, \quad (2.27)$$

sendo η uma constante previamente definida. Trata-se de um conjunto convexo fechado e para encontrar a projeção de um vetor qualquer y em C_k , pode-se primeiramente considerar três casos:

1. $\epsilon_k[i, j; \hat{f}] > \eta$;
2. $\epsilon_k[i, j; \hat{f}] < -\eta$;
3. $-\eta \leq \epsilon_k[i, j; \hat{f}] \leq \eta$.

Para encontrar a projeção no primeiro caso, escreve-se o funcional de Lagrange

$$J(\hat{f}) = \sum_{m=1}^M \sum_{n=1}^M \left| y[m, n] - \hat{f}[m, n] \right|^2 + \lambda \left[\epsilon_k[i, j; \hat{f}] - \eta \right]. \quad (2.28)$$

Derivando com relação a $\hat{f}[m, n]$ e igualando a zero, a projeção é dada por

$$P_k \hat{f}[m, n] = y[m, n] + \frac{\epsilon_k[i, j; y] - \eta}{\|h_k\|_F^2} h_k[i, j; m, n]. \quad (2.29)$$

De maneira similar, a projeção no segundo caso é dada por

$$P_k \hat{f}[m, n] = y[m, n] + \frac{\epsilon_k[i, j; y] + \eta}{\|h_k\|_F^2} h_k[i, j; m, n], \quad (2.30)$$

e no terceiro caso por

$$P_k \hat{f}[m, n] = y[m, n], \quad (2.31)$$

sendo que $\|h_k\|_F^2$ é dado por

$$\|h_k\|_F^2 = \sum_{m=1}^M \sum_{n=1}^M h_k^2[i, j; m, n], \quad (2.32)$$

e $\epsilon_k[i, j; y]$ é o resíduo quando $\hat{f}[m, n]$ é substituído por $y[m, n]$, $m, n = 1, \dots, M$. Os operadores de projeção P_k , $k = 1, \dots, q$, são aplicados para todos os pixels de baixa resolução, e a seqüência

$$\hat{f}_{n+1} = P_q P_{q-1} \dots P_1 \hat{f}_n \quad (2.33)$$

converge para a imagem de alta resolução desejada \hat{f} .

É importante notar que em geral a intersecção dos conjuntos de restrições não é formada por um único ponto. Assim as abordagens baseadas em POCS possuem mais de uma solução e o resultado depende da estimativa inicial. Para amenizar essa limitação, restrições adicionais

podem ser impostas, de forma a favorecer uma dada solução. Patti *et al.* (1997) propuseram uma técnica mais elaborada de POCS para o problema da super-resolução, que modela borramento por movimento devido a um tempo de abertura diferente de zero durante a aquisição das imagens.

2.3.5 Modelagem Bayesiana do Problema

Normalmente, a formulação Bayesiana do problema da reconstrução por super-resolução provê uma forma flexível e realística de impor conhecimento *a priori* no processo de estimação. De acordo com Katsaggelos *et al.* (2007), um princípio fundamental da abordagem Bayesiana é considerar todos os parâmetros e variáveis observáveis como quantidades estocásticas desconhecidas, atribuindo a eles distribuições de probabilidade baseadas em critérios subjetivos. Dessa forma a imagem original f é tratada como uma amostra ou realização de um campo aleatório, cuja distribuição *a priori* $p(f)$, modela nosso conhecimento a respeito da natureza dessa imagem. As observações g_k , $k = 1, \dots, q$, as quais são função de f , também são tratadas como realizações de um campo aleatório com distribuição condicional que modela o processo de se obter g_k , $k = 1, \dots, q$, a partir de f . Considerando o modelo de formação das imagens dado pela Equação (2.1), na presença de ruído Gaussiano independente de média zero, essa distribuição condicional, denominada verossimilhança das observações, é dada por

$$p(g|f) = \frac{1}{(2\pi\sigma^2)^{q\frac{N^2}{2}}} \exp \left\{ - \sum_{k=1}^q \frac{\|g_k - D_k W_k H_k f\|^2}{2\sigma^2} \right\}, \quad (2.34)$$

onde g é o conjunto de todas as observações g_k , $k = 1, \dots, q$, e σ^2 é a variância do ruído.

Baseado nessa modelagem, a inferência Bayesiana é feita por meio da probabilidade *a posteriori* que, segundo a regra de Bayes, é dada por

$$P(f|g) = \frac{p(g|f)p(f)}{p(g)}, \quad (2.35)$$

A solução MAP é encontrada maximizando essa probabilidade

$$\hat{f} = \arg \max_f \{P(f|g)\}. \quad (2.36)$$

É comum modelar a imagem de alta resolução a ser reconstruída como um campo aleatório de Markov (MRF - *Markov Random Field*). Isso ocorre porque restrições contextuais são extremamente necessárias na interpretação de informações visuais. Objetos em uma cena são reconhecidos por suas características, as quais, em um nível mais baixo são descritas a partir

de interações entre pixels próximos. A teoria de MRFs fornece uma maneira conveniente de modelar informações contextuais, o que ocorre por meio da distribuição condicional de cada pixel com relação a seus vizinhos, dado um sistema de vizinhança. O uso prático dessa teoria foi possibilitado pelo teorema de Hammersley e Clifford (1971), posteriormente desenvolvido por Besag (1974), que institui a equivalência entre as distribuições de MRF e de Gibbs. A derivação da distribuição conjunta a partir das distribuições condicionais é bastante complicada. Por isso, a afirmação de que essa distribuição é dada pela distribuição de Gibbs, que toma uma forma bastante simples, fornece uma forma matematicamente tratável para a análise estatística de imagens utilizando a modelagem por MRFs (Geman e Geman, 1984). Além disso, as propriedades locais dos MRFs permitem implementações locais e extremamente paralelizadas. A modelagem por MRFs em conjunto com o critério MAP formam o framework MAP-MRF, adotado na maioria dos trabalhos de decisão estatística em processamento de imagens e visão computacional baseados em MRFs. É importante notar que as duas principais partes das abordagens MAP-MRF são a derivação da forma da distribuição *a posteriori*, e a determinação dos parâmetros dessa distribuição. Além disso, uma decisão muito importante é a escolha do algoritmo de otimização para encontrar o máximo da distribuição (Li, 2009).

Dado o espaço de configurações do campo aleatório \mathbb{F} , e um sistema de vizinhança η , a distribuição *a priori* conjunta de f , caracterizada como um MRF, é dada pela distribuição Gibbs que toma a seguinte forma

$$p(f) = \frac{1}{Z} \exp \left\{ -\frac{1}{T} U(f) \right\}, \quad (2.37)$$

onde

$$Z = \sum_{f \in \mathbb{F}} \exp \left\{ -\frac{1}{T} U(f) \right\}, \quad (2.38)$$

é uma constante de normalização denominada função de partição, T é uma constante chamada temperatura, e a função energia

$$U(f) = \sum_{c \in C} V_c(f), \quad (2.39)$$

é uma soma de funções potencial V_c no conjunto de todos os possíveis cliques C . Como é possível notar, para calcular a distribuição de Gibbs é preciso identificar a função de partição Z , a qual é formada pela soma de todas as possíveis configurações em \mathbb{F} . Essa identificação é proibitiva mesmo para problemas bastante simples. Caso $U(f)$ não contenha nenhum parâmetro desconhecido, a identificação de Z pode ser evitada já que a solução MAP é equivalente à solução de mínima energia. Entretanto, se isso não for verdadeiro e $U(f) = U(f|\theta)$, sendo θ um parâmetro que precisa ser estimado, a função de partição também será função de θ , $Z = Z(\theta)$, e necessariamente deverá ser identificada. Existem várias aproximações para solucionar esse

problema. Schultz e Stevenson (1996) utilizaram um modelo *a priori* baseado na função de Huber

$$p(f) = \frac{1}{Z} \exp \left\{ -\frac{1}{2T} \sum_{c \in C} \rho_{\alpha} (d_c^t f) \right\}, \quad (2.40)$$

sendo $d_c^t f$ uma medida de atividade espacial, a qual é pequena em regiões suaves e grande em regiões com maior atividade em frequência. As quatro medidas de atividade

$$\begin{aligned} d_{[i,j,1]}^t f &= f[i, j - 1] - 2f[i, j] + f[i, j + 1] \\ d_{[i,j,2]}^t f &= 0.5f[i + 1, j - 1] - f[i, j] + 0.5f[i - 1, j + 1] \\ d_{[i,j,3]}^t f &= f[i - 1, j] - 2f[i, j] + f[i + 1, j] \\ d_{[i,j,4]}^t f &= 0.5f[i - 1, j - 1] - f[i, j] + 0.5f[i + 1, j + 1] \end{aligned} \quad (2.41)$$

são calculadas para cada pixel de alta resolução. Elas aproximam as derivadas de segunda ordem direcionais em $f[i, j]$, com direções horizontal, vertical e diagonal. A verossimilhança das bordas é controlada pela função de penalização de Huber

$$\rho_{\alpha}(x) = \begin{cases} x^2, & |x| \leq \alpha \\ 2\alpha|x| - \alpha^2, & |x| > \alpha, \end{cases} \quad (2.42)$$

onde α é o limiar que separa regiões lineares de regiões quadráticas. Dessa forma, não é necessário identificar a função de partição e a solução é encontrada minimizando a energia

$$\sum_{i,j} \sum_{r=1}^4 \rho_{\alpha} (d_{[i,j,r]}^t f). \quad (2.43)$$

De acordo com Katartzis e Petrou (2008), do ponto de vista de estatísticas robustas, $\rho(x)$ é uma norma robusta do erro que considera descontinuidades predominantes na imagem como outliers. Restrições adaptativas a descontinuidades semelhantes podem ser introduzidas pela teoria de *line processes* (Geman e Geman, 1984), *plate models* (Blake e Zisserman, 1987), ou a função Lorentziana (Black e Anandan, 1996).

Algoritmo *Iterated Conditional Modes* (ICM)

Encontrar o mínimo global da função de energia não é uma tarefa trivial, ainda mais se essa função possui vários mínimos locais. Além disso, de acordo com Li (2009), não existe nenhum algoritmo eficiente que consiga encontrar o mínimo global da função com garantia de sucesso. Uma solução alternativa proposta por Besag (1986) é utilizar funções densidade de probabilidades condicionais locais (LCDF - *Local Conditional Density Functions*). Ele propôs o algoritmo

iterativo *Iterated Conditional Modes* (ICM) que utiliza uma estratégia de *greedy* para atualizar cada pixel f_i pelo valor de máxima probabilidade *a posteriori* local $P(f_i|g, f_{\eta_i})$, sendo f_{η_i} os valores dos pixels vizinhos a f_i , dado o sistema de vizinhança η .

No cálculo de $P(f_i|g, f_{\eta_i})$, primeiramente assume-se que as observações são independentes dado f , ou seja, considerando g_{ℓ_i} o conjunto de pixels de baixa resolução influenciados pelo pixel de alta resolução f_i , cada g_{ℓ_i} possui a mesma função densidade condicional $p(g_{\ell_i}|f_i)$ dependente apenas de f_i

$$p(g|f) = \prod_i p(g_{\ell_i}|f_i). \quad (2.44)$$

Em segundo lugar assume-se a propriedade chamada Markovianidade. De acordo com essa propriedade f depende apenas dos valores dos pixels em uma vizinhança local, e a partir dessas duas propriedades e do teorema de Bayes, é possível afirmar que

$$P(f_i|g, f_{\eta_i}) \propto p(g_{\ell_i}|f_i)P(f_i|f_{\eta_i}). \quad (2.45)$$

Obviamente é muito mais fácil maximizar $P(f_i|g, f_{\eta_i})$ do que $P(f|g)$, e esse é justamente o apelo do algoritmo ICM.

Uma iteração do ICM ocorre ao após maximizar $P(f_i|g, f_{\eta_i})$ para cada pixel de alta resolução f_i , $i = 1, \dots, M^2$, e as iterações continuam até a convergência. Segundo Besag (1986), a convergência é garantida e ocorre em poucas iterações. Entretanto o resultado é bastante dependente da estimativa inicial. Mesmo assim, sabe-se que uma boa estimativa inicial garante bons resultados.

Na maioria das abordagens de processamento de imagens baseadas no framework MAP-MRF, e que envolvem mais de duas classes ou tons de cinza, o modelo *Multi-Level Logistic* (MLL) isotrópico, também conhecido como processo de Strauss (Strauss, 1977) ou modelo de Ising generalizado (Geman e Geman, 1984), é adotado como modelo de MRF *a priori*. Esse modelo é definido da seguinte forma

$$P(f_i = I|f_{\eta_i}) = \frac{\exp\{-\beta n_i(I)\}}{\sum_{I=0}^L \exp\{-\beta n_i(I)\}}, \quad (2.46)$$

onde $n_i(I)$ é o número de pixels em η_i que possuem tom de cinza igual a I , L é o maior tom de cinza possível, e β é o chamado parâmetro de dependência espacial. No contexto da super-resolução, Martins *et al.* (2009a) adotaram o modelo denominado MLL isotrópico gene-

realizado (*Generalized Isotropic Multi-Level Logistic - GIMLL*), dado por

$$P(f_i = I | f_{\eta_i}) = \frac{\exp \left\{ -\beta \sum_{k \in \eta_i} [1 - 2 \exp \{ -(I - f_k)^2 \}] \right\}}{\sum_{I=0}^L \exp \left\{ -\beta \sum_{k \in \eta_i} [1 - 2 \exp \{ -(I - f_k)^2 \}] \right\}}, \quad (2.47)$$

para reconstruir uma imagem de alta resolução a partir de várias observações de baixa resolução. Os autores argumentam que esse modelo é capaz de modelar as interações entre pixels vizinhos de forma mais suave do que o modelo MLL isotrópico. Isso é justificado porque no modelo MLL isotrópico, apenas pixels vizinhos que possuem mesmo tom de cinza contribuem para a probabilidade *a priori*. No modelo GIMLL, pixels com tons de cinza próximos também contribuem para a probabilidade *a priori*.

2.3.6 Interpolação Estatística

Mascarenhas *et al.* (1996) apresentam um método de interpolação para fusão de dados de satélite utilizando técnicas estatísticas Bayesianas. A interpolação estatística espacial se baseia na simulação de 3 bandas com resolução 10 metros, a partir dos 3 canais multi-espectrais de 20 metros e do canal pancromático de 10 metros do satélite SPOT. Os valores dos pixels na grade original, assim como na grade interpolada, são considerados variáveis aleatórias. A estimação linear local dos pixels interpolados é feita considerando o critério de MMSE, com base no princípio da ortogonalidade (Papoulis, 1984). Os autores assumem, por simplicidade, a separabilidade da estrutura de correlação nos domínios espacial e espectral. A estrutura de correlação espacial nas direções horizontal e vertical também é assumida, sendo considerado um modelo Markoviano de primeira ordem em ambas as direções (Pratt, 2007).

Os autores utilizaram uma vizinhança 3×3 em cada uma das três bandas multi-espectrais de resolução de 20 metros do satélite SPOT para estimar quatro pixels com resolução de 10 metros cobrindo o pixel central dessa vizinhança em cada banda. Assim, são utilizados 27 pixels (nove em cada banda) para estimar 12 pixels (quatro para cada banda). A Figura 2.10 mostra a representação esquemática dessa re-amostragem.

Seja y o vetor 27×1 obtido pela notação lexicográfica das observações nas vizinhanças 3×3 de cada uma das bandas multi-espectrais, e seja x o vetor 12×1 obtido pela notação lexicográfica dos pixels estimados nas vizinhanças 2×2 de cada uma das bandas multi-espectrais. O estimador linear não homogêneo de x é dado por

$$\hat{x} = Ay + b. \quad (2.48)$$

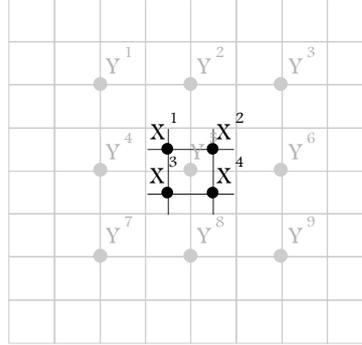


Figura 2.10: Representação esquemática da re-amostragem proposta por Mascarenhas *et al.* (1996).

Na estimação Bayesiana, a matriz A e o vetor b são obtidos por meio do princípio da ortogonalidade, chegando a

$$\hat{x} = E[x] + \Sigma_{xy} \Sigma_{yy}^{-1} (y - E[y]), \quad (2.49)$$

onde $E[.]$ indica a esperança matemática, Σ_{xy} é a matriz de covariância cruzada de x e y , e Σ_{yy} é a matriz de auto-correlação de y .

Considerando a separabilidade entre bandas e nas direções horizontal e vertical, e a notação lexicográfica de x e y , as matrizes de covariância são dadas por

$$\Sigma_{xy} = (R_h)_{xy} \otimes (R_v)_{xy} \otimes (\Sigma_s), \quad (2.50)$$

$$\Sigma_{yy} = (R_h)_{yy} \otimes (R_v)_{yy} \otimes (\Sigma_s). \quad (2.51)$$

O símbolo \otimes representa o produto de Kronecker entre duas matrizes, e h , v e s indicam as direções horizontal, vertical e espectral, respectivamente.

De acordo com a estrutura de correlação espacial Markoviana de primeira ordem, as matrizes $(R_h)_{xy}$ e $(R_h)_{yy}$ são dadas por

$$(R_h)_{xy} = \begin{bmatrix} \rho_h^{3/4} & \rho_h^{1/4} & \rho_h^{5/4} \\ \rho_h^{5/4} & \rho_h^{1/4} & \rho_h^{3/4} \end{bmatrix}, \quad (2.52)$$

$$(R_h)_{yy} = \begin{bmatrix} 1 & \rho_h & \rho_h^2 \\ \rho_h & 1 & \rho_h \\ \rho_h^2 & \rho_h & 1 \end{bmatrix}, \quad (2.53)$$

onde ρ_h é o coeficiente de correlação na direção horizontal. A mesma estrutura é válida para $(R_v)_{xy}$ e $(R_v)_{yy}$, substituindo ρ_h por ρ_v . As potências de ρ_h nas Equações (2.52) e (2.53)

dependem das distâncias entre os pixels na direção horizontal, adotando a estrutura Markoviana (Figura 2.10).

A matriz de covariância espectral (Σ_s), que é matriz de covariância entre as bandas multi-espectrais, é dada por

$$(\Sigma_s) = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{bmatrix}, \quad (2.54)$$

onde σ_{ij}^2 é a covariância entre as bandas i e j , e σ_{ii}^2 é a variância da banda i .

O método proposto por Mascarenhas *et al.* (1996) é rápido e facilmente implementável. Entretanto a imposição de conhecimento *a priori* ocorre apenas por meio do modelo adotado para caracterizar as estruturas de correlação das observações e cruzada. Além disso, devido ao contexto da imagens observadas, não existe movimento entre as bandas.

2.3.7 Filtro de Wiener Adaptativo

Hardie (2007) apresenta um algoritmo de reconstrução por super-resolução baseado em um filtro de Wiener adaptativo (Hardie, 2010). De maneira semelhante ao trabalho de Mascarenhas *et al.* (1996), os pixels de alta resolução são obtidos por meio da soma ponderada dos pixels observados de acordo com sua localização espacial. A Figura 2.11 ilustra a distribuição dos pixels de baixa resolução na grade de alta resolução de acordo com os deslocamentos de ordem sub-pixel presentes entre as imagens observadas. Como ilustrado na figura, é adotada uma janela de observação deslizante na grade de alta resolução que cobre W_x espaçamentos de um pixel de alta resolução na direção horizontal, e W_y na direção vertical. Todos os pixels de baixa resolução presentes nessa janela são dispostos em um vetor de observação $g_i = [g_{i,1}, g_{i,2}, \dots, g_{i,K_i}]$, onde i indica a posição da janela na grade de alta resolução, e K_i é o número de pixels de baixa resolução presentes na i -ésima janela de observação.

Para cada janela de observação são estimados pixels de alta resolução contidos em uma sub-janela de dimensão $D_x \times D_y$, $1 \leq D_x \leq W_x$, $1 \leq D_y \leq W_y$, como ilustrado na Figura 2.11. As estimações de alta resolução são obtidas utilizando uma soma ponderada dos pixels de baixa resolução na janela de observação

$$\hat{d}_i = W_i^T g_i, \quad (2.55)$$

onde $\hat{d}_i = [\hat{d}_{i,1}, \hat{d}_{i,2}, \dots, \hat{d}_{i,D_x D_y}]^T$ e W_i é uma matriz de pesos de dimensão $K_i \times D_x D_y$. Cada coluna de W_i contém os pesos utilizados para um pixel de alta resolução em particular.

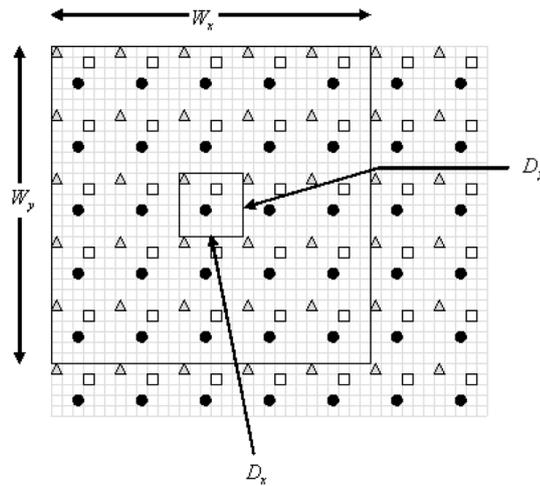


Figura 2.11: Distribuição dos pixels de baixa resolução na grade de alta resolução de acordo com os deslocamentos presentes entre eles (Hardie, 2007).

Os pesos na Equação (2.55) que minimizam o erro médio quadrático são dados por

$$W_i = R_i^{-1} P_i, \quad (2.56)$$

onde $R_i = E \{g_i g_i^T\}$ é a matriz de correlação do vetor de observações e $P_i = E \{g_i d_i^T\}$ é a correlação cruzada entre o vetor estimado d_i e o vetor de observações g_i . É importante notar que, até aqui, não existe nenhuma diferença entre as abordagens propostas por Hardie (2007) e por Mascarenhas *et al.* (1996). A modelagem das matrizes de covariância define as peculiaridades de cada uma das abordagens.

Considerando f_i a versão sem ruído do vetor de observação g_i , $g_i = f_i + n_i$, e assumindo que n_i modela ruído não correlacionado de média zero com elementos independentes e identicamente distribuídos com variância σ_n^2 , a matriz de auto-correlação de g_i é dada por

$$R_i = E \{g_i g_i^T\} = E \{f_i f_i^T\} + \sigma_n^2 I, \quad (2.57)$$

e a matriz de correlação cruzada é dada por

$$P_i = E \{g_i d_i^T\} = E \{f_i d_i^T\}. \quad (2.58)$$

Definindo uma função de auto-correlação $r_{dd}(x, y)$ para a imagem desejada d , a função de correlação cruzada entre $d(x, y)$ e $f(x, y)$ pode ser expressa em termos de $r_{dd}(x, y)$ como

$$r_{df}(x, y) = r_{dd}(x, y) * h(x, y), \quad (2.59)$$

e a auto-correlação de $f(x, y)$ expressa como

$$r_{ff}(x, y) = r_{dd}(x, y) * h(x, y) * h(-x, -y). \quad (2.60)$$

Hardie (2007) adotou o modelo de auto-correlação paramétrico circularmente simétrico

$$r_{dd}(x, y) = \sigma_d^2 \rho \sqrt{x^2 + y^2}, \quad (2.61)$$

onde σ_d^2 é a variância da imagem desejada e ρ é um parâmetro que controla o decaimento da auto-correlação com a distância.

O método é denominado adaptativo porque um conjunto de pesos único é utilizado para cada pixel de alta resolução, já que seus valores dependem da distribuição dos pixels de baixa resolução na grade de alta resolução. Além disso, Hardie (2007) propõe que a variância da imagem σ_d^2 também varie de acordo com a posição da janela de observação. Assim o modelo de auto-correlação é expresso como

$$r_{dd}(x, y) = \sigma_{d_i}^2 \rho \sqrt{x^2 + y^2}, \quad (2.62)$$

onde $\sigma_{d_i}^2$ é a variância da região local da imagem desejada que é responsável pela geração de g_i .

O método também é denominado rápido pois ao considerar apenas deslocamentos globais, dependendo da dimensão da janela de observação, o conjunto de pesos será único para todos os pixels de alta resolução. Entretanto, os autores não consideram o caso em que existem deslocamentos mais complexos.

2.4 Registro das Imagens

O registro de imagens é o processo de alinhar geometricamente duas ou mais imagens da mesma cena, adquiridas em momentos diferentes, de diferentes pontos de vista, ou por diferentes sensores. Trata-se de uma tarefa crucial quando o resultado da aplicação depende da combinação de dados de várias fontes. Atualmente o registro é utilizado rotineiramente em aplicações que envolvem o imageamento médico. Alguns exemplos de aplicações comuns são: a fusão de imagens de tomografia computadorizada com imagens de ressonância magnética para obter uma informação mais completa a respeito do paciente; monitoramento do crescimento de tumores; comparação dos dados de um paciente com um atlas anatômico; verificação de tratamento através de imagens anteriores e posteriores à intervenção médica; e a evolução de uma substância injetada no paciente sendo que esse pode se mover enquanto o acompanhamento é feito. O

registro rígido (baseado em rotações e translações globais da cena) é frequentemente utilizado, entretanto ele nem sempre é capaz de fornecer uma solução satisfatória. Isso ocorre porque os órgãos podem sofrer deformações, são corpos não rígidos. Dessa forma, são necessárias transformações não-lineares para corrigir as diferenças locais entre as imagens.

As imagens do trato vocal são adquiridas durante a emissão da fala de palavras ou fonemas. Dessa forma, o método de registro adotado deve ser capaz de identificar transformações que modelem deformações causadas pela movimentação dos articuladores da fala, como as evidenciadas na Figura 2.12(c). Essa figura mostra o erro entre as imagens mostradas nas Figuras 2.12(a) e 2.12(b) (dois momentos da emissão da fala de uma palavra). Como verificado empiricamente, essas deformações garantem a existência de deslocamentos de ordem sub-pixel entre as imagens, principalmente nas regiões referentes aos articuladores da fala.

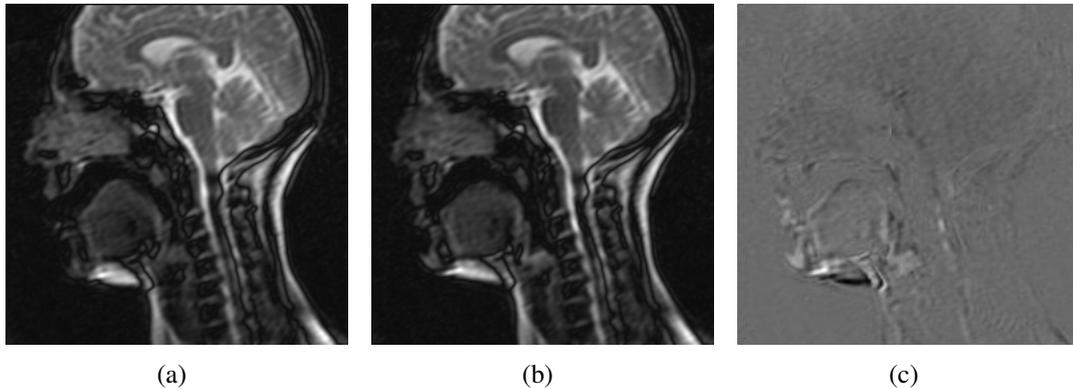


Figura 2.12: Exemplo de deformação presente nas imagens de MRI do trato vocal humano.

O processo de registro busca a identificação de uma transformação geométrica ótima que maximiza as correspondências entre duas imagens. Isso é feito com base nos seguintes componentes (Rueckert e Aljabar, 2010):

- Modelo de transformação que define o tipo de transformação geométrica que existe entre as imagens. Transformações não-rígidas podem ser definidas utilizando modelos paramétricos ou não-paramétricos. Alguns desses modelos são próprios para pequenas deformações e outros são capazes de representar grandes deformações.
- Medida de similaridade que mede o grau de alinhamento entre as imagens. A similaridade pode ser calculada utilizando diretamente as intensidades das imagens, ou utilizando a distância entre características identificadas, como pontos de referência, bordas ou superfícies.

- Método de otimização que maximiza a medida de similaridade. O registro não-rígido pode ser formulado como um problema de otimização cujo objetivo é maximizar uma função objetivo associada.

A seguir os modelos de transformação mais adequados para o contexto das imagens desse projeto, juntamente com as possíveis medidas de similaridade são discutidos.

2.4.1 Modelo de Transformação

O modelo de transformação define como estão relacionadas as coordenadas das duas imagens. Considerando duas imagens I_1 e I_2 , esse modelo é representado por uma transformação $T(\mathbf{x})$ que mapeia cada ponto \mathbf{x} de I_1 para a localização correspondente em I_2 . Quando a transformação modela apenas translações, rotações e mudanças de escala, as coordenadas de I_2 podem ser escritas como combinação linear das coordenadas de entrada. Porém, considerando o registro não-rígido, esse modelo linear e global não pode ser adotado. Normalmente um campo de deslocamentos espacialmente variável é utilizado na otimização. Além disso, duas restrições são impostas à transformação: suavidade e que ela seja inversível. Tais restrições garantem que as variações na anatomia dos corpos serão modeladas de forma coerente. Mudanças no tamanho e forma são comuns, entretanto mudanças na topologia são raras e, portanto, não devem ser permitidas.

Uma extensão do modelo rígido é o modelo de transformação afim

$$T(x, y, z) = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (2.63)$$

o qual possui doze graus de liberdade, e preserva colinearidade e a relação de distância entre pontos. Essa transformação é utilizada quando existe apenas uma quantidade limitada de variações na forma dos objetos. Ao se adicionar mais graus de liberdade, essa transformação linear pode ser estendida para modelos de transformações não-lineares. O modelo de transformação quadrático, por exemplo, possui trinta graus de liberdade e é definido por polinômios de segunda ordem

$$T(x, y, z) = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{00} & \dots & a_{08} & a_{09} \\ a_{10} & \dots & a_{18} & a_{19} \\ a_{20} & \dots & a_{28} & a_{29} \\ 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} x^2 \\ y^2 \\ \vdots \\ 1 \end{bmatrix}. \quad (2.64)$$

De maneira análoga, esse modelo pode ser estendido utilizando polinômios de terceira ordem (60 graus de liberdade), quarta ordem (105 graus de liberdade), ou mesmo quinta ordem (168 graus de liberdade). Porém essas transformações não permitem mudanças locais. Dessa forma, não são capazes de acomodar a variabilidade na anatomia de corpos não-rígidos. Além disso, polinômios de ordem mais alta geralmente provocam oscilações.

O modelo de transformação também pode ser definido por meio da combinação linear de funções base θ_i , $i = 1, \dots, n$,

$$T(x, y, z) = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{00} & \dots & a_{0n} \\ a_{10} & \dots & a_{1n} \\ a_{20} & \dots & a_{2n} \\ 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \theta_1(x, y, z) \\ \vdots \\ \theta_n(x, y, z) \end{bmatrix}. \quad (2.65)$$

Escolhas comuns são funções ortonormais como as funções base de Fourier ou wavelets.

Entretanto, os modelos de transformações mais utilizados são os baseados na família de splines. Normalmente é identificada a correspondência entre pontos de controle em ambas as imagens, e tal correspondência é utilizada, em conjunto com uma função spline, para identificar as coordenadas dos outros pontos de forma suave. Portanto existe a restrição

$$T(\phi_i) = \phi'_i, i = 1, \dots, n \quad (2.66)$$

onde ϕ_i denota a localização do i -ésimo ponto de controle na imagem de referencia, e ϕ'_i a localização do ponto de controle correspondente na outra imagem. As thin-plate splines (Bookstein, 1989; Grimson, 1982) tem sido extensamente utilizadas para o registro não-rígido de imagens médicas (Rohr, 2001). Elas são parte de uma família de splines que se baseiam em funções base radiais. Uma função base radial spline por ser definida como uma combinação linear de n funções base radiais $\theta(s)$ (Rueckert, 2001)

$$t(x, y, z) = a_1 + a_2x + a_3y + a_4z + \sum_{j=1}^n \theta(|\phi_j - (x, y, z)|). \quad (2.67)$$

No caso da thin-plate spline, a função base radial é definida como $\theta(s) = |s|^2 \log(|s|)$ no caso bidimensional, e $\theta(s) = |s|$ no caso tridimensional. Definindo a transformação como três funções thin-plate splines $T(t_1, t_2, t_3)$, um mapeamento entre as imagens é gerado de forma que os coeficientes a caracterizam uma transformação afim e os coeficientes b caracterizam a parte não afim da transformação. A restrição mostrada na Equação (2.66) forma $3n$ equações, sendo que é preciso determinar $3(n + 4)$ coeficientes. As outras 12 equações são definidas fazendo

com que a soma dos coeficientes b seja zero e seu produto com as coordenadas x , y e z também seja zero. Utilizando a notação matricial chega-se a

$$\begin{bmatrix} \Theta & \Phi \\ \Phi^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \Phi' \\ 0 \end{bmatrix}, \quad (2.68)$$

onde \mathbf{a} é o vetor de coeficientes a , \mathbf{b} é o vetor de coeficientes b , e Θ é uma matriz cujos elementos são dados por $\Theta_{ij} = \theta (|\phi_i - \phi_j|)$. Como mostra a Equação (2.67), nesse modelo cada ponto de controle possui uma influencia global sobre a transformação. Isso pode ser uma desvantagem já que a modelagem de deformações mais localizadas e complexas se torna difícil. Além disso, o custo computacional de se mover um único ponto de controle cresce consideravelmente à medida que o número de pontos de controle aumenta.

Por outro lado, B-splines são definidas considerando apenas pontos de controle espacialmente próximos. Dessa forma, ao se perturbar a posição de um ponto de controle, apenas sua vizinhança será afetada pela transformação. Por isso, diz-se que as B-splines possuem suporte finito. Atualmente, o registro não-rígido utilizando *free-form deformations* (FFD) baseadas em funções localmente controladas, como as B-splines, tem sido extensamente utilizado na literatura (Rueckert e Aljabar, 2010). Primeiramente apresentada por Sederberg e Parry (1986), a FFD é uma abordagem utilizada em aplicações de computação gráfica para modelar objetos tridimensionais deformáveis. O objeto é deformado de acordo com a manipulação de uma malha de pontos de controle. Considerando o caso bidimensional, a Figura 2.13 ilustra a deformação da imagem Lena ao se manipular a malha de pontos de controle uniformemente espaçados. Essa malha está sendo representada pela imagem mostrada na Figura 2.13(b), onde os pontos de controle estão localizados nas intersecções das linhas.

Rueckert *et al.* (1999) apresentam uma abordagem de registro não-rígido utilizando FFDs baseadas na interpolação por funções B-Spline cúbicas. A deformação entre as imagens é modelada por uma transformação $T(x, y, z)$ que combina movimento global e deformações locais

$$T(x, y, z) = T_{\text{global}}(x, y, z) + T_{\text{local}}(x, y, z), \quad (2.69)$$

onde T_{global} é modelada por uma transformação afim aplicada a todos os pixels da imagem, e T_{local} por FFDs utilizando funções B-spline cúbicas como funções base. Essa modelagem produz uma transformação suave e C^2 contínua. Seja $\Omega = \{(x, y, z) | 0 \leq x \leq X, 0 \leq y \leq Y, 0 \leq z \leq Z\}$ o suporte da imagem, e Φ a malha de pontos de controle $\phi_{i,j,k}$, $n_x \times n_y \times n_z$, com espaçamento uniforme δ , a FFD pode ser escrita como o produto tensorial da família

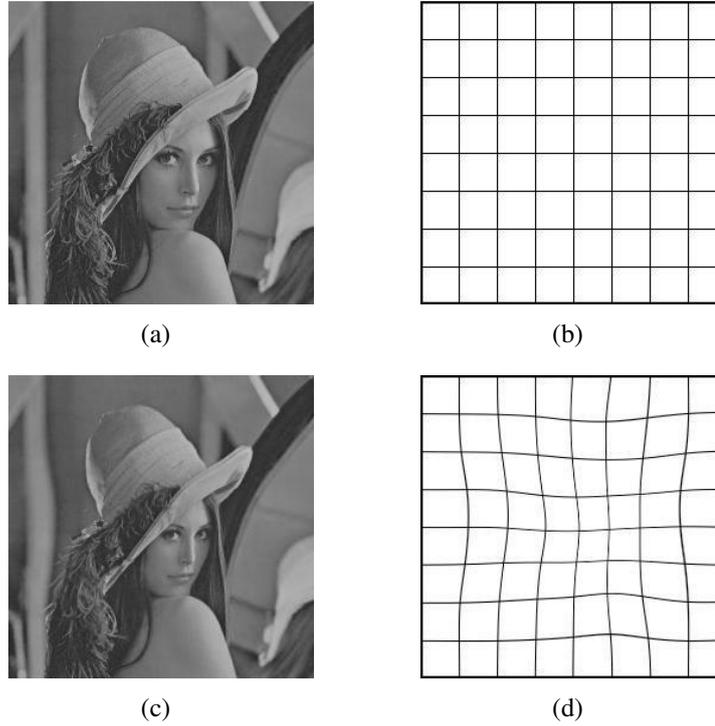


Figura 2.13: Deformação da imagem Lena, utilizando FFD baseada em funções base B-spline.

unidimensional de funções B-spline cúbicas

$$T_{\text{local}}(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(u) B_m(v) B_n(w) \phi_{i+l, j+m, k+n}, \quad (2.70)$$

onde $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $k = \lfloor z/n_z \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ e $w = z/n_z - \lfloor z/n_z \rfloor$ sendo que $\lfloor a \rfloor$ indica o menor inteiro maior que a . B_l é a l -ésima função base da família B-spline

$$\begin{aligned} B_0(u) &= (1 - u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6. \end{aligned} \quad (2.71)$$

Os pontos de controle $\phi_{i,j,k}$ agem como parâmetros da transformação. Tal característica torna essa abordagem bastante vantajosa em relação às demais, já que a identificação dos pontos de controle é independente do conhecimento de especialistas ou de características da imagem. Além disso, como as funções base utilizadas possuem suporte finito, a mudança na posição de

um ponto de controle afeta apenas sua vizinhança. Isso permite que deformações localizadas em pequenas partes da imagem possam ser modeladas facilmente.

A resolução da malha de pontos de controle define o grau da deformação e sua localidade. Espaçamentos menores entre os pontos permitem deformações abruptas e localizadas em pequenas partes da imagem, enquanto espaçamentos maiores permitem apenas deformações mais globais. Entretanto, quanto mais pontos de controle, maior o custo computacional do registro. Assim, para garantir um melhor compromisso entre custo computacional e número de graus de liberdade necessários para modelar as deformações, os autores utilizaram uma abordagem hierárquica multi-resolução desse método de registro (Lee *et al.*, 1996).

Seja Φ^1, \dots, Φ^L uma hierarquia de pontos de controle, na qual o espaçamento entre os pontos de controle diminui de Φ^l para Φ^{l+1} , a transformação local T_{local}^l é definida pela malha Φ^l juntamente com as funções B-spline baseadas na FFD. Dessa forma, a transformação local T_{local} é dada pela combinação das funções B-spline baseadas em FFD em diferentes resoluções da malha de pontos de controle

$$T_{\text{local}}(x, y, z) = \sum_{l=0}^L T_{\text{local}}^l(x, y, z). \quad (2.72)$$

Para que a transformação seja suave, um termo de penalização é introduzido. Esse termo, descrito por Wahba (1990), toma a seguinte forma

$$C_{\text{suavidade}}(T) = \frac{1}{V} \int_0^X \int_0^Y \int_0^Z \left[\left(\frac{\partial^2 T}{\partial x^2} \right)^2 + \left(\frac{\partial^2 T}{\partial y^2} \right)^2 + \left(\frac{\partial^2 T}{\partial z^2} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial xy} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial xz} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial yz} \right)^2 \right] dx dy dz, \quad (2.73)$$

onde V denota o volume da imagem. Esse termo de regularização é igual a zero para qualquer transformação afim, e, dessa forma, penaliza apenas as transformações não-rígidas modeladas pela transformação local T_{local} .

A medida de similaridade utilizada pelos autores foi a informação mútua normalizada (NMI - *Normalized Mutual Information*) que será descrita na próxima seção. Para encontrar a transformação ótima, minimiza-se a função custo associada aos parâmetros da transformação global, que serão denominados Θ , e aos parâmetros da transformação local Φ . Essa função é composta pelo custo relacionado à suavidade da transformação $C_{\text{suavidade}}$, e pelo custo relacionado à similaridade entre as imagens $C_{\text{similaridade}}$

$$C(\Theta, \Phi) = -C_{\text{similaridade}}(I_1(x, y, z), I_2(T(x, y, z))) + \lambda C_{\text{suavidade}}(T), \quad (2.74)$$

sendo λ um parâmetro de peso que define o compromisso entre esses dois custos.

A otimização é feita em duas etapas. Primeiramente os parâmetros da transformação afim são otimizados, o que equivale a maximizar a medida de similaridade, já que $C_{suavidade} = 0$ para transformações afim. Rueckert *et al.* (1999) utilizaram uma estratégia de busca iterativa multi-resolução (Studholme *et al.*, 1997). Entretanto, é possível adotar qualquer método de registro que considere apenas transformações afim (Brown, 1992; Zitová e Flusser, 2003). Posteriormente, os parâmetros da transformação local, Φ , são otimizados por uma técnica iterativa de gradiente descendente com passo de tamanho μ na direção do vetor gradiente. O algoritmo para quando um ótimo local da função custo é encontrado ($\|\nabla C\| \leq \epsilon$, para ϵ pequeno).

O registro baseado na transformação elástica modela a deformação como um processo físico semelhante ao comportamento de um corpo elástico feito de borracha. Esse processo físico é gerenciado por duas forças: uma força interna que combate qualquer força que deforma o corpo elástico de seu equilíbrio; e uma força externa que age no corpo elástico. A deformação para quando essas duas forças encontram o equilíbrio. De acordo com Rueckert (2001), o comportamento do corpo elástico pode ser descrito pela equação diferencial parcial de Navier

$$\mu \nabla^2 \mathbf{u}(x, y, z) + (\lambda + \mu) \nabla (\nabla \cdot \mathbf{u}(x, y, z)) + \mathbf{f}(x, y, z) = 0, \quad (2.75)$$

onde \mathbf{u} descreve o campo de deslocamentos, \mathbf{f} a força externa que atua no corpo elástico, ∇ é o operador gradiente, e ∇^2 o operador Laplaciano. Os parâmetros μ e λ são constantes de elasticidade de Lamé. O processo de registro é dirigido pela força externa que \mathbf{f} . Algumas alternativas para essa força são: o gradiente da medida de similaridade, características de intensidade como bordas ou curvaturas, distância entes curvas ou superfícies de estruturas anatômicas correspondentes, etc.

O registro elástico dificulta a modelagem de deformações bastante localizadas devido à ação das energias interna e externa que crescem proporcionalmente. Uma variação desse modelo de transformação é a transformação fluida, na qual essas forças são relaxadas com o tempo. Assim deformações mais localizadas, incluindo cantos, são permitidas. No registro baseado na transformação fluida, o frame de referencia Euleriano é utilizado. A transformação elástica utiliza o frame de referencia Lagrangiano, ou seja, a deformação é descrita a partir da configuração inicial da imagem. Dessa forma, é possível acompanhar o caminho que cada pixel percorre até que o registro seja encontrado. No registro fluido, a deformação é descrita por sua configuração corrente. A transformação é implementada em função de um campo de velocidade que é definido diferenciando o campo de deslocamentos \mathbf{u} com relação ao tempo

$$\mathbf{v}(x, y, z, t) = \left. \frac{\partial \mathbf{u}}{\partial t} \right|_{(x,y,z,t)} + \nabla \mathbf{u}(x, y, z, t) \mathbf{v}(x, y, z, t). \quad (2.76)$$

As deformações são caracterizadas pela equação diferencial parcial de Navier-Stokes

$$\mu \nabla^2 \mathbf{v}(x, y, z, t) + (\lambda + \mu) \nabla (\nabla \cdot \mathbf{v}(x, y, z, t)) + \mathbf{f}(x, y, z, t) = 0. \quad (2.77)$$

A transformação fluida é conveniente quando grandes deformações ou um alto grau de variabilidade são necessários. Entretanto, com essa liberdade a probabilidade de erro aumenta consideravelmente. Além disso, como o campo de velocidade pode variar com o tempo e com o espaço, a otimização pode ser um desafio.

2.4.2 Medida de Similaridade

A medida de similaridade indica o grau de alinhamento entre as imagens. Isso pode ser feito considerando diretamente as intensidades das imagens, ou características como pontos, linhas ou superfícies. No último caso se busca minimizar a distância entre tais características. Uma vantagem de se utilizar características das imagens é que o registro pode ser facilmente utilizado para o caso multi-modalidade. Entretanto, a necessidade da extração de tais características pode ser um processo oneroso. Além disso, qualquer erro que ocorre durante esse processo pode afetar de forma irreversível o resultado do registro. O uso das intensidades das imagens diretamente é muito mais simples e não exige nenhuma etapa de pré-processamento. Apesar da desvantagem de que o registro multi-modalidade é muito mais complexo, nos últimos anos as medidas de similaridade baseadas nas intensidades das imagens tem sido as mais utilizadas.

Considerando a imagem de referência I_1 , uma imagem I_2 a ser comparada com a imagem de referência, e a imagem transformada $I_2(T(\mathbf{x}))$, para $\mathbf{x} = (x, y, z)$, a soma das diferenças quadradas (SSD - *Sum of Squared Differences*) das intensidades das imagens é dada por

$$C_{\text{SSD}} [I_1(\mathbf{x}), I_2(T(\mathbf{x}))] = \frac{1}{n} \sum (I_2(T(\mathbf{x})) - I_1(\mathbf{x}))^2, \quad (2.78)$$

sendo n o número de voxels de cada imagem. Essa medida de similaridade assume que as imagens pertencem à mesma modalidade e possuem as mesmas características. Assim, se as imagens estão alinhadas, ela deve ser zero exceto pela presença de ruído. De acordo com Ruckert e Aljabar (2010), no caso de ruído Gaussiano, a SSD pode ser ótima.

Uma abordagem um pouco menos restritiva seria assumir a existência de uma relação linear entre as intensidades das imagens. Isso é feito pela correlação cruzada normalizada (*normalized cross-correlation*)

$$C_{\text{NCC}} [I_1(\mathbf{x}), I_2(T(\mathbf{x}))] = \frac{\sum (I_1(\mathbf{x}) - \mu_{I_1})(I_2(T(\mathbf{x})) - \mu_{I_2})}{\sqrt{\sum (I_1(\mathbf{x}) - \mu_{I_1})^2 \sum (I_2(T(\mathbf{x})) - \mu_{I_2})^2}} \quad (2.79)$$

onde μ_{I_1} e μ_{I_2} são as médias de I_1 e I_2 , respectivamente. Essa medida não é sensível a mudanças de amplitude em $I_1(\mathbf{x})$ e $I_2(T(\mathbf{x}))$. Uma propriedade interessante da correlação é que ela também pode ser realizada no domínio da frequência através da transformada rápida de Fourier (FFT - *Fast Fourier Transform*). Translações no domínio do espaço são equivalentes a mudanças de fase no domínio da frequência. Dessa forma, no domínio da frequência a correlação é chamada correlação por fase (*phase-correlation*), e pode ser mais eficiente se o tamanho das imagens for igual.

Em geral, as transformações que utilizam a correlação cruzada como critério de similaridade consideram apenas a presença de translações globais entre as imagens. Entretanto, segundo Hill e Batchelor (2001), o módulo do espectro de potência das imagens não contém nenhuma informação de fase e é, portanto, invariante a translação. Rotações no domínio do espaço são rotações pelo mesmo ângulo no domínio de Fourier. Utilizando uma representação polar no domínio da frequência, a rotação se torna um deslocamento da coordenada angular, que pode ser encontrado por correlação da representação polar da magnitude no domínio da frequência. Encontrada a rotação, a translação pode ser determinada pela diferença de fase no domínio de Fourier cartesiano.

Desde que foi independentemente proposta por Viola (1995) e Collignon *et al.* (1995), a informação mútua entre as imagens (MI - *Mutual Information*) tem sido muito utilizada como medida de similaridade. Trata-se de uma medida baseada na entropia das imagens. Seja $q \in \mathbb{N}$ e ρ uma densidade em \mathbb{R}^3 , $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$, $\rho(\mathbf{x}) \geq 0$ e $\int_{\mathbb{R}^3} \rho(\mathbf{x}) d\mathbf{x} = 1$. A entropia, também chamada entropia diferencial, é definida como

$$H(\rho) = -\mathbb{E}_{\rho} [\log \rho] = - \int_{\mathbb{R}^3} \rho \log \rho d\rho. \quad (2.80)$$

Dessa forma, C_{MI} é definida como

$$C_{MI} [I_1(\mathbf{x}), I_2(T(\mathbf{x}))] = H(\rho_{I_1(\mathbf{x})}) + H(\rho_{I_2(T(\mathbf{x}))}) - H(\rho_{I_1(\mathbf{x}), I_2(T(\mathbf{x}))}), \quad (2.81)$$

onde $\rho_{I_1(\mathbf{x})}$ e $\rho_{I_2(T(\mathbf{x}))}$ são as densidades de $I_1(\mathbf{x})$ e $I_2(T(\mathbf{x}))$, respectivamente, e $\rho_{I_1(\mathbf{x}), I_2(T(\mathbf{x}))}$ é a distribuição conjunta de $I_1(\mathbf{x})$ e $I_2(T(\mathbf{x}))$. Dessa forma, a MI mede a entropia da densidade conjunta, que é máxima quando as imagens estão maximamente relacionadas.

Modelagem por Campos Aleatórios de Markov

De acordo com Li (2009), restrições contextuais são extremamente necessárias na interpretação de informações visuais¹. A teoria de campos aleatórios de Markov fornece uma forma conveniente e consistente de modelar entidades dependentes de contexto, como é o caso dos pixels de uma imagem e das características relacionadas a sua distribuição espacial. Isso é feito por meio da caracterização de influências mútuas entre tais entidades utilizando distribuições condicionais. O uso prático dessa teoria se deve principalmente ao teorema proposto por Hammersley e Clifford (1971), futuramente desenvolvido por Besag (1974), que provou a equivalência entre os Campos Markovianos e a Distribuição de Gibbs. A maioria das aplicações requerem a distribuição conjunta e, para MRFs, derivá-la a partir das distribuições condicionais é muito difícil. A equivalência MRF-Gibbs afirma que a distribuição conjunta de um MRF é uma distribuição de Gibbs, a qual possui uma forma mais simples. Além disso, considerando o aspecto computacional, as propriedades locais dos MRFs possibilitam implementações locais e altamente paralelas, além de fornecer uma estrutura para aplicações multirresolução.

Nesse contexto, a teoria de MRFs apresenta uma forma de modelar a probabilidade *a priori* de padrões dependentes de contexto. Um modelo em particular irá favorecer uma dada classe de padrões associando a ela as maiores probabilidades. Essa teoria frequentemente é utilizada

¹Este capítulo é baseado no trabalho de Li (2009).

em conjunto com a teoria de estimação a fim de formular funções objetivo para atingir critérios ótimos. O critério de máxima probabilidade *a posteriori* (MAP) é o um dos mais populares. MRFs juntamente com o critério MAP formam o framework MAP-MRF. Geman e Geman (1984) discutem a aplicação desse framework para o problema da restauração de imagens. Nas abordagens MAP-MRF a função objetivo é a probabilidade *a posteriori* conjunta de todos labels do campo aleatório, a qual, de acordo com a fórmula de Bayes, é determinada pela distribuição *a priori* conjunta desses labels e pela probabilidade condicional das observações. Dessa forma, os principais desafios na modelagem MAP-MRF são: determinar a forma da distribuição *a posteriori*, determinar todos os parâmetros que a caracterizam e encontrar o melhor algoritmo de otimização para encontrar o máximo da distribuição *a posteriori*.

3.1 Sites e Labels

Seja S o conjunto de índices de um conjunto de m sites

$$S = \{1, \dots, m\}, \quad (3.1)$$

onde $1, \dots, m$ são índices. Nesse contexto, um site representa um ponto em uma região no espaço Euclidiano. Esses sites podem estar dispostos em uma grade regular, como os pixels de uma imagem, ou estar irregularmente dispostos no espaço, como no caso de características extraídas da imagem. Independente dessa regularidade, a relação entre sites é estabelecida por um sistema de vizinhança.

Um label é um evento que ocorre a um site. Seja L um conjunto de labels. Esse conjunto pode ser discreto ou contínuo. No caso contínuo, o conjunto de labels pode corresponder à linha real \mathbb{R} . No caso discreto, o label assume um valor discreto em um conjunto de M possibilidades

$$L = \{l_1, \dots, l_M\}. \quad (3.2)$$

É importante notar que, no caso em que os labels em L podem ser ordenados, uma medida numérica de similaridade entre dois labels quaisquer pode ser definida. Isso irá afetar a escolha do algoritmo para a atribuição de labels aos sites.

Considerando o conjunto

$$f = \{f_1, \dots, f_m\} \quad (3.3)$$

a atribuição de labels em L aos sites em S , quando apenas um label é atribuído a cada site, $f_i = f(i)$ pode ser considerada uma função de domínio S e imagem L . Como o suporte da

função é todo o domínio S , trata-se de um mapeamento

$$f : S \rightarrow L. \quad (3.4)$$

No contexto dos campos aleatórios, uma atribuição de labels a todos os sites é chamada configuração. Quando todos os sites possuem o mesmo conjunto de labels L , o conjunto de todas as possíveis configurações, ou seja, o espaço de configurações, é o produto Cartesiano

$$\mathbb{F} = L \times L \dots \times L = L^m. \quad (3.5)$$

A atribuição de labels pode ser feita considerando restrições contextuais. Em termos de probabilidades, essas restrições podem ser expressas localmente em termos de probabilidades condicionais $P(f_i | \{f_{i'}\})$, onde $\{f_{i'}\}$ representa um conjunto de labels nos outros sites $i' \neq i$. Caso os labels sejam independentes entre si, a probabilidade conjunta será o produto das probabilidades locais, o que facilita muito já que a atribuição global pode ser encontrada considerando cada label separadamente. Entretanto, considerando a informação contextual, os labels são mutuamente dependentes e a inferência global utilizando as informações locais se torna muito mais difícil.

3.2 Sistemas de Vizinhaça e Cliques

Os sites em S estão relacionados entre si por um sistema de vizinhaça

$$\eta = \{\eta_i | \forall i \in S\}, \quad (3.6)$$

onde η_i é conjunto de sites vizinhos a i . A relação de vizinhaça possui as seguintes propriedades:

1. Um site não é vizinho de si mesmo $i \notin \eta_i$.
2. A relação de vizinhaça é mútua $i \in \eta_{i'} \iff i' \in \eta_i$.

Como pode ser visto na Figura 3.1(a), adotando o sistema de vizinhaça de primeira ordem, todos os sites a , $a = 1, \dots, m$, possuem quatro vizinhos. De acordo com o sistema de vizinhaça de segunda ordem, como pode ser visto na Figura 3.1(b) todos os sites possuem oito vizinhos. A Figura 3.1(c) mostra sistemas de vizinhaça de graus mais altos. Em todos esses sistemas de vizinhaça, as duas propriedades são respeitadas.

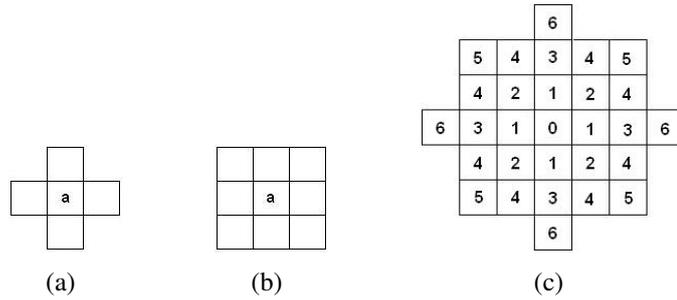


Figura 3.1: Sistemas de vizinhança (a) de primeira ordem, (b) de segunda ordem e (c) de ordens mais altas.

O par $(S, \eta) \triangleq G$ constitui um grafo onde S contém os nós e η determina as arestas que ligam os nós de acordo com o sistema de vizinhança. Um clique c em (S, η) é definido por um subconjunto de sites em S . Ele pode ser formado por um único site $c = \{i\}$, um par de sites vizinhos $c = \{i, i'\}$, três sites vizinhos $c = \{i, i', i''\}$, e assim por diante. As coleções de cliques formados por apenas um site, um par de sites, e três sites, denominadas C_1, C_2 e C_3 , respectivamente, são dadas por

$$C_1 = \{i | i \in S\} \tag{3.7}$$

$$C_2 = \{\{i, i'\} | i' \in \eta_i, i \in S\} \tag{3.8}$$

e

$$C_3 = \{\{i, i', i''\} | i, i', i'' \in S \text{ são vizinhos entre si}\}. \tag{3.9}$$

É importante notar que os sites em um clique são ordenados. Portanto, $\{i, i'\}$ não é o mesmo clique que $\{i', i\}$. As Figuras de 3.2(a) até 3.2(c) mostram os possíveis cliques para o sistema de vizinhança de primeira ordem, e as Figuras de 3.2(a) até 3.2(j) mostram os possíveis cliques para o sistema de vizinhança de segunda ordem.

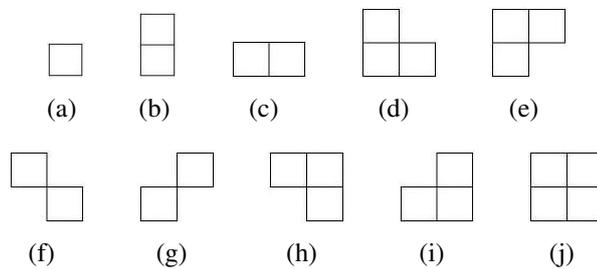


Figura 3.2: Sistemas de vizinhança (a) de primeira ordem, (b) de segunda ordem e (c) de ordens mais altas.

3.3 Campos Aleatórios de Markov

Um campo aleatório $F = \{F_1, \dots, F_m\}$ é uma família de variáveis aleatórias definidas em S , de forma que cada variável aleatória F_i assume um valor f_i de L . $F_i = f_i$ denota o evento em que F_i assume o valor f_i e $(F_1 = f_1, \dots, F_m = f_m)$ denota o evento conjunto, abreviado por $F = f$, onde $f = \{f_1, \dots, f_m\}$ é a configuração de F . Essa configuração corresponde a uma realização do campo. Considerando que L é um conjunto discreto de labels, $P(F_i = f_i)$, ou a abreviação $P(f_i)$ é a probabilidade de F_i assumir o valor f_i , e $P(F = f) = P(F_1 = f_1, \dots, F_m = f_m)$, ou a abreviação $P(f)$, é a probabilidade conjunta. Para o caso em que L é um conjunto contínuo, ao invés de probabilidades, existirão funções densidade de probabilidades $p(F_i = f_i)$ e $p(F = f)$.

O campo aleatório F pode ser denominado um MRF em S com relação a um sistema de vizinhança η se e somente se as condições a seguir são satisfeitas:

1. $P(f) > 0, \forall f \in \mathbb{F}$
2. $P(f_i | f_{S-\{i\}}) = P(f_i | f_{\eta_i})$

onde $S - \{i\}$ são todos os sites de S menos o site i , e $f_{S-\{i\}}$ é o conjunto de todos os labels em $S - \{i\}$. $f_{\eta_i} = \{f_{i'} | i' \in \eta_i\}$ é o conjunto de labels dos sites vizinhos a i . Quando a primeira condição é satisfeita, a probabilidade conjunta $P(f)$ é unicamente determinada por suas probabilidades condicionais locais (Besag, 1974). A segunda condição configura as características locais do campo F . Em MRFs apenas vizinhos possuem interação direta. No caso mais geral em que os vizinhos de cada site englobam toda a imagem, qualquer campos aleatório é um MRF com relação a esse sistema de vizinhança. Outra propriedade dos MRFs é que um MRF é dito homogêneo se $P(f_i | f_{\eta_i})$ é independente da localização de i em S . Assim, se $f_i = f_j$ e $f_{\eta_i} = f_{\eta_j}$, $P(f_i | f_{\eta_i}) = P(f_j | f_{\eta_j})$ mesmo se $i \neq j$.

De acordo com Li (2009), existem duas formas de se especificar um MRF: por meio das probabilidades condicionais $P(f_i | f_{\eta_i})$ ou por meio da probabilidade conjunta $P(f)$. Besag (1974) argumenta a favor da representação por meio da probabilidade conjunta já que não existe nenhum método para deduzir a probabilidade conjunta a partir das probabilidades condicionais associadas. Nesse contexto, o resultado teórico a respeito da equivalência entre MRFs e a distribuição de Gibbs apresenta uma forma matematicamente tratável de especificar essa probabilidade conjunta.

3.4 Equivalência entre MRFs e a Distribuição de Gibbs

Um campo aleatório F é dito um campo aleatório de Gibbs em S com relação a um sistema de vizinhança η se e somente se suas configurações obedecem a uma distribuição de Gibbs, a qual toma a forma

$$P(f) = Z^{-1} e^{-\frac{1}{T}U(f)} \quad (3.10)$$

onde

$$Z = \sum_{f \in \mathbb{F}} e^{-\frac{1}{T}U(f)} \quad (3.11)$$

é uma constante de normalização chamada função de partição, T é uma constante chamada temperatura e $U(f)$ é a função de energia definida como

$$U(f) = \sum_{c \in C} V_c(f) \quad (3.12)$$

é uma soma de funções potenciais $V_c(f)$ sobre todos os possíveis cliques C . O valor de $V_c(f)$ depende da configuração local na clique c . O campo aleatório de Gibbs é dito homogêneo se V_c independe da posição relativa de c em S e é dito isotrópico se V_c é independente da orientação de c .

O teorema de Hammersley e Clifford (1971) estabeleceu a equivalência entre MRFs e a distribuição de Gibbs. O teorema afirma que F é um MRF em S com relação a um sistema de vizinhança η se e somente se F é um campo aleatório de Gibbs em S com relação a um sistema de vizinhança η . Considerando a probabilidade condicional

$$P(f_i | f_{S-\{i\}}) = \frac{P(f_i, f_{S-\{i\}})}{P(f_{S-\{i\}})} = \frac{P(f)}{\sum_{f'_i \in L} P(f')} \quad (3.13)$$

onde f' é qualquer configuração diferente de f exceto em i . Assumindo que

$$P(f) = Z^{-1} e^{-\sum_{c \in C} V_c(f)}$$

chega-se a

$$P(f_i | f_{S-\{i\}}) = \frac{e^{-\sum_{c \in C} V_c(f)}}{\sum_{f'_i} e^{-\sum_{c \in C} V_c(f')}} \quad (3.14)$$

Dividindo C em dois conjuntos A e B , sendo A o conjuntos dos cliques que contém i e B o conjunto dos cliques que não contém i , $P(f_i|f_{S-\{i\}})$ pode ser escrita como

$$P(f_i|f_{S-\{i\}}) = \frac{[e^{-\sum_{c \in A} V_c(f)}][e^{-\sum_{c \in B} V_c(f)}]}{\sum_{f'_i} \{[e^{-\sum_{c \in A} V_c(f'_i)}][e^{-\sum_{c \in B} V_c(f'_i)}]\}}. \quad (3.15)$$

Como $V_c(f) = V_c(f')$ para qualquer clique que não contém i , é possível cancelar $e^{-\sum_{c \in B} V_c(f)}$ do numerador e do denominador. Assim, a probabilidade depende apenas das funções potencial dos cliques que contém i , ou seja, dos labels dos vizinhos de i . Isso prova que um campo aleatório de Gibbs é um MRF.

Esse teorema apresenta uma forma simples de especificar a probabilidade conjunta do MRF. Isso é feito especificando as funções potencial $V_c(f)$ de forma a caracterizar um comportamento desejado para o sistema. Dessa forma, é possível codificar com simplicidade a informação *a priori* a respeito da estimação.

Entretanto, apesar dessa simplicidade, a maximização da probabilidade conjunta normalmente demanda alto poder computacional. Além disso, a estimação dos parâmetros do modelo de Markov é um problema computacionalmente inviável, a otimização global é difícil de ser implementada exatamente, e uma aproximação sempre se faz necessária. Nesse contexto, o algoritmo ICM é uma alternativa interessante.

3.5 Algoritmo *Iterated Conditional Modes*

Como maximizar a probabilidade condicional de um MRF é difícil, Besag (1986) propôs um algoritmo determinístico que maximiza as probabilidades condicionais sequencialmente. Dadas as observações d e os labels dos outros sites $f_{S-\{i\}}^{(k)}$, o algoritmo atualiza sequencialmente $f_i^{(k)}$ para $f_i^{(k+1)}$ maximizando $P(f_i|d, f_{S-\{i\}})$, a probabilidade condicional *a posteriori* com relação a f_i . São feitas duas suposições no cálculo de $P(f_i|d, f_{S-\{i\}})$. A primeira é que as observações d_1, \dots, d_m são condicionalmente independentes da realização do campo f , e cada d_i possui a mesma função densidade condicional $p(d_i|f_i)$ dependente apenas de f_i . Assim

$$p(d|f) = \prod_i p(d_i|f_i). \quad (3.16)$$

A segunda suposição é que f depende dos labels na vizinhança local (segunda propriedade dos MRFs). A partir dessas duas suposições, segundo o teorema de Bayes, segue que

$$P(f_i|d, f_{S-\{i\}}) \propto p(d_i|f_i)P(f_i|f_{\eta_i}). \quad (3.17)$$

Claramente $P(f_i|d_i, f_{\eta_i}^{(k)})$ é bem mais fácil de maximizar que $P(f|d)$.

Maximizar a Equação (3.17) é equivalente a minimizar a função potencial *a posteriori* correspondente

$$f_i^{(k+1)} \leftarrow \arg \min_{f_i} V(f_i|d_i, f_{\eta_i}^{(k)}) \quad (3.18)$$

onde

$$V(f_i|d_i, f_{\eta_i}) = \sum_{i' \in \eta_i} V(f_i|f_{i'}^{(k)}) + V(d_i|f_i). \quad (3.19)$$

Quando a Equação (3.18) é aplicada a todos os sites i em S , é completado um ciclo do ICM. As iterações continuam até a convergência. De acordo com Besag (1986), a convergência é garantida e acontece rapidamente. É importante notar que o resultado do ICM é dependente da estimativa inicial $f^{(0)}$.

Estimador por Mínimo Erro Médio Quadrático

4.1 Estimador de Bayes

Suponha que se quer estimar f a partir das observações d , e na estimação comete-se um erro¹

$$\varepsilon = f - \hat{f}, \quad (4.1)$$

sendo \hat{f} o estimador de f . Associando uma função custo $C(f, \hat{f})$ a todos os pares (f, \hat{f}) , e supondo que essa função dependa apenas do erro de estimação, ela pode ser representada por $C(\varepsilon)$.

Na estimação de Bayes, um risco é minimizado a fim de se obter a estimação ótima. Por definição, o risco de Bayes da estimação \hat{f} é representado pela esperança do custo

$$R(\hat{f}) = E[C(\varepsilon)] = \int_f C(\varepsilon)P(f|d)df, \quad (4.2)$$

¹Este capítulo é baseado no trabalho de Candeias (1992).

onde $P(f|d)$ é a probabilidade *a posteriori* de f . De acordo com a regra de Bayes

$$P(f|d) = \frac{p(d|f)P(f)}{p(d)}, \quad (4.3)$$

onde $P(f)$ é a probabilidade *a priori* de f , $p(d|f)$ é a função densidade de probabilidade condicional de d também chamada função de verossimilhança, e $p(d)$ é a densidade de d que é constante quando d é dado.

Uma escolha bastante comum para $C(\epsilon)$ é o quadrado da norma do erro da estimação

$$C(\epsilon) = \|f - \hat{f}\|^2. \quad (4.4)$$

O risco de Bayes considerando a função custo quadrática, mede a variância da estimativa

$$R(\hat{f}) = \int_f \|f - \hat{f}\|^2 P(f|d) df. \quad (4.5)$$

A estimação de f será ótima se o risco de Bayes for mínimo. Para obter o risco mínimo, diferencia-se o risco $\frac{\partial R(\hat{f})}{\partial \hat{f}} = 0$, e tem-se que o estimador será dado por

$$\hat{f} = \int_f f P(f|d) df, \quad (4.6)$$

ou seja, obtém-se a estimativa de mínima variância que é equivalente à média da probabilidade *a posteriori*.

4.2 Princípio da Ortogonalidade

A estimativa ótima a partir do risco quadrático nem sempre é fácil de ser calculada, já que para isso é preciso conhecer a probabilidade *a posteriori* $P(f|d)$. Considerando o caso particular da estimação linear, para obter o estimador \hat{f} é necessário apenas o conhecimento dos dois primeiros momentos de f dado d (média e variância).

Pelo princípio da ortogonalidade, o risco será mínimo se o erro da estimação ϵ for perpendicular ao subespaço onde \hat{f} está contido (Figura 4.1). Assume-se que \hat{f} pertença a um espaço H^d , onde H^d é o conjunto de todas as variáveis aleatórias linearmente dependentes de d com probabilidade um. Assume-se também que

$$E[\|f - \hat{f}\|^2] \leq E[\|f - u\|^2] \forall u \in H^d, \quad (4.7)$$

ou seja, o risco é mínimo para o estimador \hat{f} . Nesse caso, assume-se que d é um vetor de dimensão $n \times 1$, f e \hat{f} são unidimensionais e \hat{f} é combinação linear das n variáveis aleatórias d_1, \dots, d_n

$$\hat{f} = a_1 d_1 + \dots + a_n d_n. \quad (4.8)$$

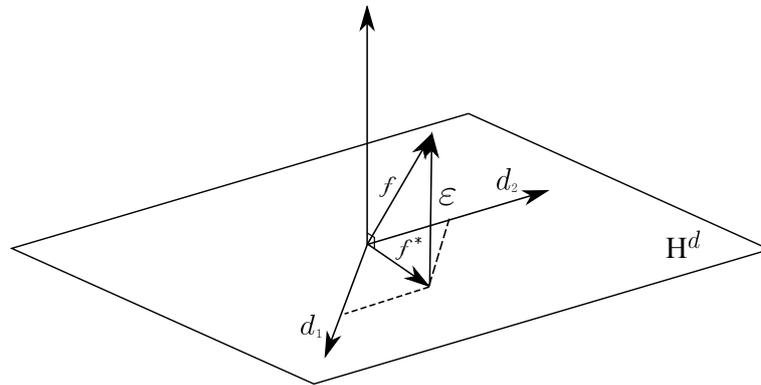


Figura 4.1: Interpretação geométrica da ortogonalidade (adaptado de Papoulis (1984)).

De acordo com o princípio da ortogonalidade, para que o risco seja mínimo o erro deve ser perpendicular ao vetor d . Assim as n constantes a_1, \dots, a_n devem ser determinadas de forma que (Papoulis, 1984)

$$E[(f - \hat{f})d^T] = 0. \quad (4.9)$$

A partir dessa Equação desenvolvem-se as equações de Yule-Walker

$$\sum_{j=1}^n R_{ij} a_j = R_{oi}, \quad (4.10)$$

onde $R_{ij} = E[d_i d_j]$ e $R_{oi} = E[f d_i]$, $i = 1, \dots, n$ e $j = 1, \dots, n$. Considerando a notação

$$d^T = [d_1, \dots, d_n],$$

$$A^T = [a_1, \dots, a_n],$$

$$R = [R_{ij}]_{i=1, \dots, n; j=1, \dots, n},$$

$$R_o = [R_{oi}]_{i=1, \dots, n},$$

chega-se ao seguinte sistema

$$\hat{f} = Ad \quad RA = R_o. \quad (4.11)$$

Resolvendo este sistema, o vetor de constantes A é dado por

$$A = R_o R^{-1} = E[fd^T]E[dd^T]^{-1}. \quad (4.12)$$

4.3 Estimador Linear Não Homogêneo

Assumindo que \hat{f} é um estimador linear não homogêneo do vetor aleatório f de dimensão $m \times 1$

$$\hat{f} = Ad + B, \quad (4.13)$$

onde A é uma matriz de constantes de dimensão $m \times n$, d é um vetor de dados observados de dimensão $n \times 1$, e B é um vetor de *offsets* de dimensão $m \times 1$. É importante notar que o caso linear homogêneo é um caso particular quando o vetor B é nulo.

O erro cometido na estimação de f é dado por

$$\varepsilon = f - \hat{f}. \quad (4.14)$$

Uma função custo é associada à estimação e supõe-se que essa função depende apenas do erro de estimação. Assume-se que ela seja dada pelo quadrado da norma de ε

$$C(\varepsilon) = \varepsilon^T \varepsilon. \quad (4.15)$$

A estimativa de f será ótima se o risco de Bayes for mínimo. Esse risco é dado pela esperança do custo

$$R = E[C(\varepsilon)]. \quad (4.16)$$

Deseja-se encontrar a matriz A e o vetor B de forma que o risco de Bayes seja mínimo. Supondo que o estimador é linear e diferenciando o risco em relação a cada elemento da matriz A , tem-se que a estimação ótima é dada pelo princípio da ortogonalidade

$$E[\varepsilon d^T] = 0. \quad (4.17)$$

Por este princípio é possível observar que o erro de estimação é não correlacionado com os dados e a projeção de ε em relação a d é zero.

Aplicando a esperança matemática na Equação (4.13) tem-se que

$$E[\hat{f}] = AE[d] + B. \quad (4.18)$$

Supondo que o estimador é não tendencioso tem-se que

$$B = E[f] - AEd. \quad (4.19)$$

As matrizes de covariância de d (Σ_{dd}) e de f em relação a d (Σ_{fd}) são dadas por

$$\begin{aligned} \Sigma_{dd} &= E[dd^t] - E[d]E[d^t], \\ \Sigma_{fd} &= E[fd^t] - E[f]E[d^t]. \end{aligned} \quad (4.20)$$

Substituindo a Equação (4.19) na Equação (4.17) e conhecendo as matrizes de covariância mostradas na Equação (4.20), tem-se que a matriz A será dada por

$$A = \Sigma_{fd}\Sigma_{dd}^{-1}, \quad (4.21)$$

onde Σ_{dd}^{-1} é a inversa de Σ_{dd} , e o vetor B será dado por

$$B = E[f] - (\Sigma_{fd}\Sigma_{dd}^{-1})E[d]. \quad (4.22)$$

Assim, o estimador \hat{f} é dado por

$$\hat{f} = E[f] + \Sigma_{fd}\Sigma_{dd}^{-1}(d - E[d]). \quad (4.23)$$

Método Proposto

5.1 Registro das Imagens

Como discutido na Seção 2.4, o método de registro adotado deve ser capaz de identificar transformações que modelem deformações causadas pela movimentação dos articuladores da fala, como as evidenciadas na Figura 2.12(c). Essas deformações se concentram na região da mandíbula, sendo que a maior parte da imagem não se move. Dentre os modelos de transformação discutidos na Seção 2.4.1, aqueles definidos por meio da combinação linear de funções base parecem os mais interessantes. No caso das thin-plate splines, além da exigência da identificação de pontos de controle, cada um desses pontos possui uma influencia global sobre a transformação. Assim, a modelagem de deformações mais localizadas se torna difícil. De acordo com Rueckert e Aljabar (2010), FFDs baseadas em funções localmente controladas, como as B-splines, tem sido extensamente utilizadas para o registro de imagens médicas. Como comentado anteriormente, FFD é uma abordagem utilizada em aplicações de computação gráfica para modelar objetos tridimensionais deformáveis. O objeto é deformado de acordo com a manipulação de uma malha de pontos de controle uniformemente espaçados. Os pontos de controle são os únicos parâmetros da transformação, o que torna essa abordagem bastante vantajosa já que a identificação desses pontos é independente do conhecimento de especialistas ou de características da imagem. Além disso, como as funções base utilizadas possuem suporte finito, a mudança

na posição de um ponto de controle afeta apenas sua vizinhança. Isso permite a modelagem de deformações localizadas em pequenas partes da imagem.

O registro não-rígido utilizando FFDs baseadas em funções B-spline como funções base (Rueckert *et al.*, 1999) foi capaz de modelar adequadamente as deformações existentes entre as imagens observadas. Considerando a imagem mostrada na Figura 5.1(a) como imagem de referência, esse método de registro foi aplicado na imagem mostrada na Figura 5.1(c). A transformação identificada é definida pela malha de pontos de controle mostrada na Figura 5.1(d), sendo que os pontos de controle se encontram nas intersecções entre as linhas. É importante ressaltar que a imagem dessa malha é apenas ilustrativa, não sendo necessária para a execução do método de registro. Além disso, a resolução dessa malha é determinada automaticamente pelo método de registro, de forma a obter a melhor relação entre a qualidade do registro e o custo computacional envolvido. No caso das malhas mostradas na Figura 5.1, a resolução alcançada foi 17×17 . As Figuras 5.1(e) e 5.1(f) mostram o erro entre as imagens 5.1(a) e 5.1(c) antes e após a aplicação do registro, respectivamente. Como se pode notar, o método de registro identificou a grande maioria das deformações existentes entre as imagens. Kroon (2010) disponibilizou uma implementação paralela do algoritmo proposto por Rueckert *et al.* (1999), que possibilita o uso de várias medidas de similaridade (diferença, erro médio quadrático, informação mútua normalizada, correlação cruzada normalizada, logaritmo da diferença absoluta, etc.) e dois métodos de otimização diferentes (algoritmo quasi-Newton Broyden-Fletcher-Goldfarb-Shanno e uma técnica iterativa de gradiente descendente). No contexto das imagens do trato vocal, os melhores resultados foram alcançados ao se utilizar o erro médio quadrático como medida de similaridade e o método de otimização Broyden-Fletcher-Goldfarb-Shanno (BFGS).

5.2 Aumento de Resolução Temporal

A fim de aumentar a resolução temporal de uma sequência de imagens de um dado evento, abordagens existentes utilizam múltiplas fontes de aquisição ou várias aquisições da mesma fonte com um pequeno atraso entre elas (Caspi e Irani, 2002; Singh *et al.*, 2007). Entretanto, quando se possui apenas uma única aquisição de um evento, o aumento de resolução temporal é alcançado por meio de interpolação entre as observações. Uma abordagem simples seria combinar as intensidades de mesma localização espacial de dois frames adjacentes para gerar um frame intermediário. Essa abordagem apresenta qualidade visual aceitável na ausência de movimento, mas ela causa borramento em partes que se movem de uma imagem para outra, como ilustrado na Figura 5.2. Técnicas de interpolação baseadas em movimento são capazes de

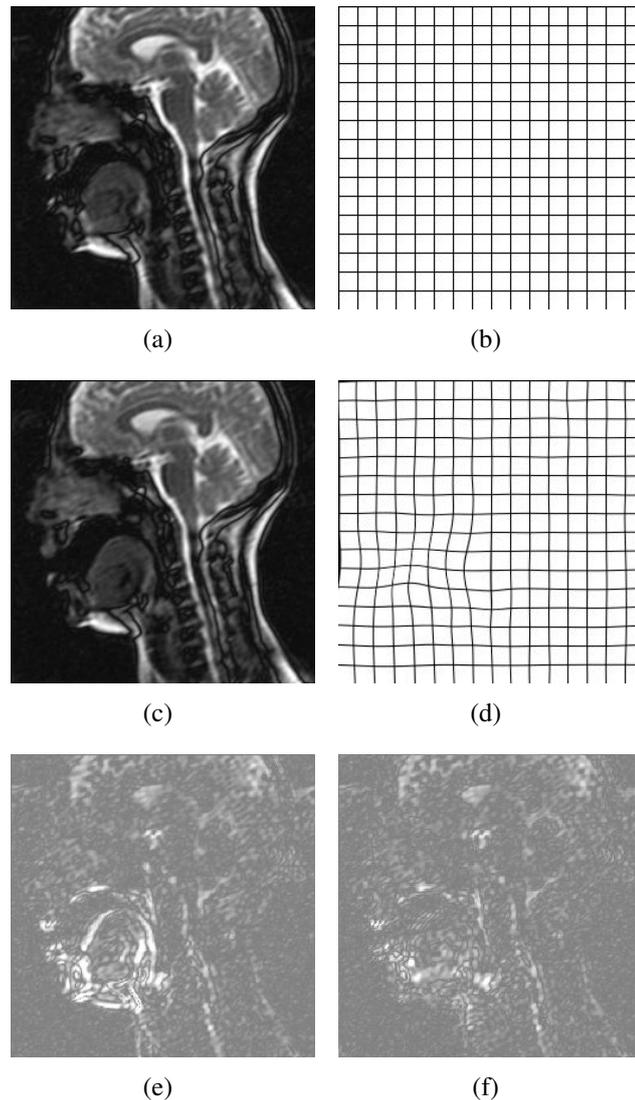


Figura 5.1: Erro entre duas imagens após o registro utilizando o método proposto por Rueckert *et al.* (1999): (a) imagem de referência; (b) malha de pontos de controle uniformemente espaçados; (c) imagem a ser registrada de acordo com a imagem de referência; (d) malha de pontos de controle identifica após a aplicação do registro; (e) erro entre as imagens antes da aplicação do registro; (f) erro entre as imagens após a aplicação do registro.

reduzir esses artefatos. Essas técnicas utilizam informação a respeito do movimento presente na sequência para aumentar a resolução temporal.

Penney *et al.* (2004) apresentam um método de interpolação entre fatias adjacentes de um conjunto de dados tomográficos tridimensionais. Fatias intermediárias às fatias existentes foram geradas a fim de se alcançar uma dimensão isotrópica dos voxels da imagem tridimensional. O método se baseia na correspondência espacial entre as fatias. Essa correspondência foi identi-

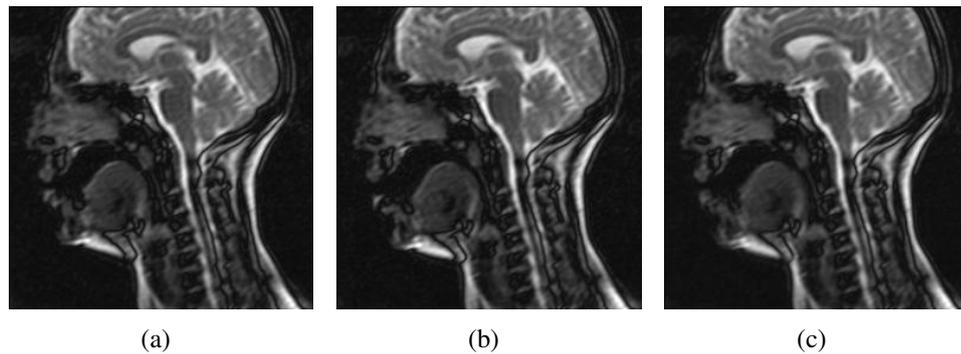


Figura 5.2: (a) e (b) Frames adjacentes adquiridos durante a emissão de uma palavra. (c) Combinação das intensidades de mesma localização espacial desses dois frames, a fim de gerar um frame intermediário.

ficada pelo método de registro não-rígido apresentado por Rueckert *et al.* (1999). De acordo com os autores, para que seja possível aplicar a interpolação baseada no registro assume-se duas hipóteses: imagens adjacentes possuem características semelhantes, e o algoritmo de registro é capaz de encontrar uma transformação que mapeia características de uma imagem nas características correspondentes na outra imagem. No caso das sequências de imagens do trato vocal, tais hipóteses são verdadeiras. Assim, é possível gerar imagens intermediárias de acordo com as transformações identificadas pelo método de registro. Como ilustrado na Figura 5.3, Penney *et al.* (2004) aplicaram interpolação linear na direção do movimento para identificar a posição de pontos na fatia intermediária.

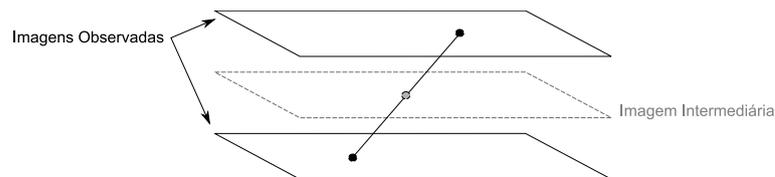


Figura 5.3: Ilustração da interpolação entre fatias adjacentes (adaptado de Penney *et al.* (2004)).

Nesse sentido, a representação por meio de malhas de pontos de controle se mostra uma ferramenta interessante para o aumento de resolução temporal. De acordo com a correspondência entre as malhas de pontos de controle, imagens intermediárias às imagens observadas podem ser geradas simplesmente posicionando pontos de controle em posições intermediárias. Como os únicos parâmetros da transformação são as coordenadas dos pontos de controle, a nova malha determina a geração da imagem intermediária. Assim, como descrito por Penney *et al.* (2004), a abordagem mais intuitiva é a interpolação linear das coordenadas de pontos de controle correspondentes das malhas das imagens adjacentes. Entretanto, no contexto das imagens do trato

vocal, uma sequência de imagens retrata o movimento contínuo e suave dos articuladores da fala durante a emissão de sons que constituem a fala. Assim, a interpolação linear na direção do vetor de movimento não parece ser a abordagem mais adequada. O movimento presente em imagens intermediárias deve ser coerente com o movimento presente em toda a sequência. Isso pode ser garantido se forem consideradas mais imagens na geração do frame intermediário. Segundo Thévenaz *et al.* (2000), existem várias motivações para o uso de splines ou funções baseadas em splines na interpolação. Essas funções são capazes de evitar a presença de artefatos, apresentando um ótimo compromisso entre qualidade e custo computacional. Nesse contexto, o aumento de resolução temporal de uma sequência de imagens de MR do trato vocal poderia ser feito por meio de interpolação por splines cúbicas como ilustrado na Figura 5.4.

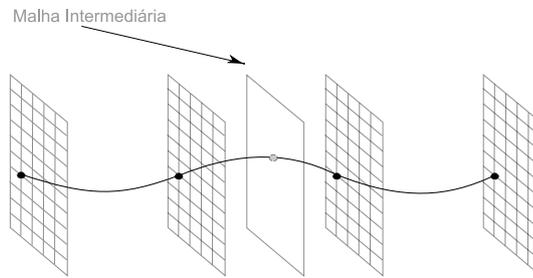


Figura 5.4: Ilustração da interpolação entre duas malhas de pontos de controle por meio de funções spline cúbicas.

Considerando uma sequência de imagens do trato vocal I_0, \dots, I_k , $k \in \mathbb{N}$, o algoritmo de registro foi aplicado nas imagens como ilustrado na Figura 5.5, sempre registrando a primeira imagem com relação às demais. É importante notar que essa abordagem foi adotada porque o movimento dos articuladores da fala é restrito a uma pequena área. Em outro contexto pode ser interessante registrar pares de imagens consecutivas. De qualquer forma, encontradas essas malhas de pontos de controle, é possível identificar a transformação entre qualquer par de imagens na sequência registrada.

A Figura 5.6 mostra o movimento de um ponto de controle correspondente em 20 imagens consecutivas de uma sequência observada. As Figuras 5.6(a) e 5.6(c) mostram o movimento considerando a interpolação por funções spline cúbicas nas direções horizontal e vertical, respectivamente. As Figuras 5.6(b) e 5.6(d) mostram o movimento considerando a interpolação linear nas direções horizontal e vertical, respectivamente. Nota-se que a interpolação por funções spline cúbicas parece ser mais coerente com o movimento presente entre as imagens.

Independente do método de interpolação, identificada a malha de pontos de controle da imagem intermediária, aplica-se a transformação referente a essa malha nas duas imagens adjacentes como ilustrado na Figura 5.7. As intensidades dos pixels da nova imagem são encontradas

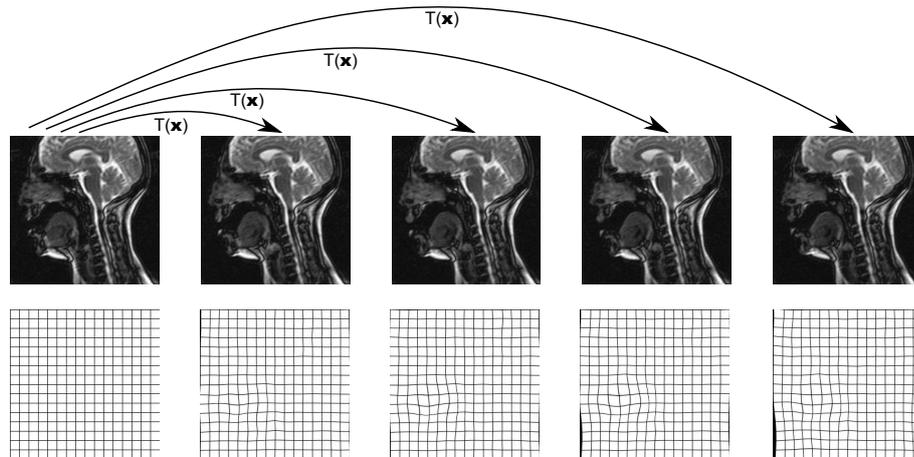


Figura 5.5: Ilustração da aplicação do método de registro, sempre registrando a primeira imagem com relação às demais.

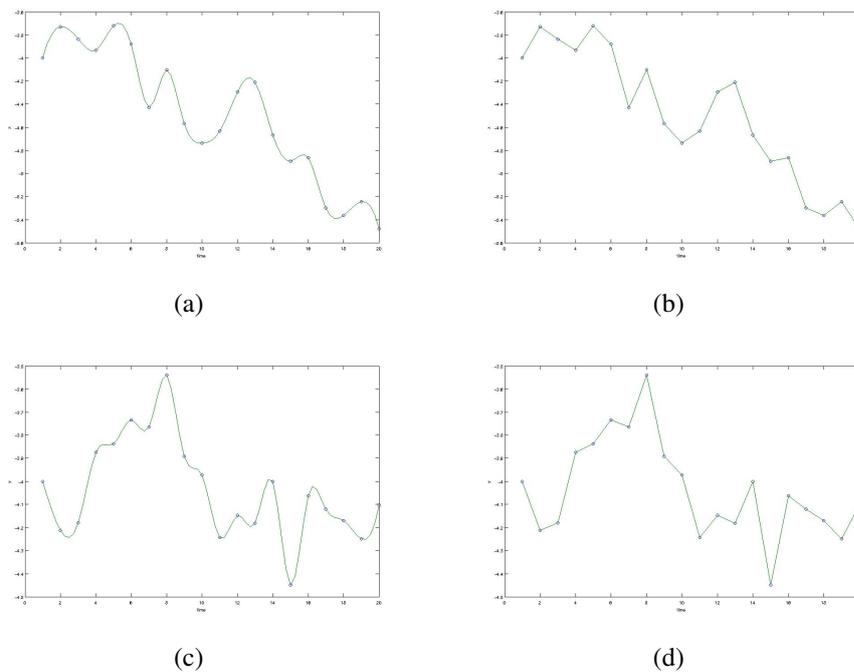


Figura 5.6: Movimento de um ponto de controle correspondente em 20 imagens consecutivas em uma sequência observada.

realizando a média ponderada das intensidades correspondentes nas imagens adjacentes transformadas. Os pesos w_1 e w_2 são definidos de acordo com a distância da nova imagem a cada uma de suas vizinhas. A imagem da qual a nova imagem estiver mais próxima recebe um peso mais alto. Por exemplo, considerando que duas imagens tenham sido adquiridas nos instantes 1 e 2, para gerar a imagem referente ao instante 1.75, os pesos seriam 0.25 para as intensidades da

imagem adquirida no instante 1 e 0.75 para as intensidades da imagem adquirida no instante 2 (a soma dos pesos sempre é igual a 1). Dessa forma, o aumento de resolução temporal preserva o movimento presente nas imagens observadas.

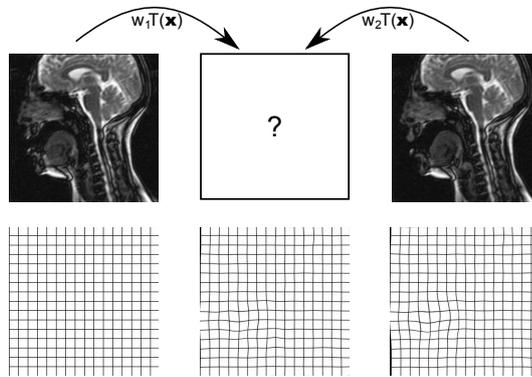


Figura 5.7: Aumento de resolução temporal através da soma ponderada das imagens vizinhas transformadas.

Como discutido no Capítulo 6, três experimentos foram conduzidos a fim de comparar objetivamente essas duas propostas de aumento de resolução temporal. Apesar da interpolação por splines cúbicas intuitivamente ser mais adequada, apenas em um dos experimentos ela apresentou melhores resultados e, considerando todos os experimentos, testes estatísticos evidenciaram que não há diferença significativa entre as médias das duas abordagens. Dessa forma, o aumento de resolução temporal utilizando interpolação linear na direção do movimento foi adotado por ser um procedimento mais simples e, conseqüentemente, apresentar menor custo computacional.

5.3 Aumento de Resolução Espacial

5.3.1 Abordagem Inicial

O principal objetivo estipulado inicialmente para o aumento de resolução espacial das imagens observadas foi a extensão da abordagem proposta pela aluna em seu projeto de mestrado. Nesse projeto, adotando um framework Bayesiano, as imagens de alta resolução foram modeladas utilizando campos aleatórios de Markov. No contexto da reconstrução por super-resolução, normalmente existem apenas observações de baixa resolução ruidosas. Dessa forma, é usual impor a restrição de suavidade à solução. Em uma abordagem baseada em uma solução de máxima probabilidade *a posteriori* (MAP), que caracteriza a imagem a ser estimada por um MRF (abordagem MAP-MRF), essa restrição de suavidade é expressa pela probabilidade *a priori* da

imagem de alta resolução, a qual é determinada unicamente pelas probabilidades condicionais locais do MRF (Besag, 1986).

Em MRFs, apenas pixels vizinhos possuem interação direta. Assim, a restrição de suavidade pode ser imposta apenas considerando que na vizinhança de um pixel os valores não podem mudar abruptamente. Entretanto, apesar dessa simplicidade, maximizar a probabilidade conjunta normalmente exige alto poder computacional. Além disso, a otimização global é difícil de ser calculada com exatidão e uma aproximação tem que ser utilizada (Li *et al.*, 1995). Nesse contexto, o algoritmo ICM é uma alternativa interessante. Trata-se de um algoritmo determinístico proposto por Besag (1974), que maximiza as probabilidades condicionais locais sequencialmente.

Seja $f[i, j]$, $0 \leq i, j \leq M$, uma imagem ideal não degradada e amostrada na taxa de Nyquist a partir de uma cena contínua $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Em uma situação real, a imagem digital sofre borramento pelo sistema ótico durante a aquisição, além de ser corrompida por ruído. Dessa forma, seguindo uma notação lexicográfica, uma versão de baixa resolução degradada $g_k[k, l]$, $0 \leq k, l \leq N$, $N \leq M$, da imagem de alta resolução f , pode ser modelada por

$$g_k = D_k f + n_k, \quad (5.1)$$

onde n_k é o ruído na k -ésima imagem de baixa resolução seguindo um modelo aditivo. O operador D_k , de dimensão $N^2 \times M^2$, modela a função do sensor de aquisição da imagem. Ele consiste na convolução com a função de espalhamento pontual (PSF) do sensor, seguida da aplicação de um operador de amostragem, o qual é dado pela multiplicação por uma soma de impulsos posicionados na grade de baixa resolução. De acordo com Park *et al.* (2003), a maioria dos métodos propostos na literatura modelam a PSF do sensor como um operador de média espacial, atribuindo a média de um bloco de alta resolução ao pixel de baixa resolução relacionado (Joshi e Chaudhuri, 2006; Joshi e Jalobeanu, 2010; Rajan e Chaudhuri, 2002; Wang e Qi, 2005). Além disso, alguns trabalhos aplicam esse operador de forma que ele já modele os deslocamentos de ordem sub-pixel entre as imagens observadas como ilustrado na Figura 5.8 (Schultz e Stevenson, 1996).

Dessa forma, na prática o operador D_k possui d^2 valores $1/d^2$ em cada linha, sendo d o fator de sub-amostragem:

$$D_k = \frac{1}{d^2} \begin{bmatrix} 11 \dots 1 & & & 0 \\ & 11 \dots 1 & & \\ & & \dots & \\ 0 & & & 11 \dots 1 \end{bmatrix}. \quad (5.2)$$

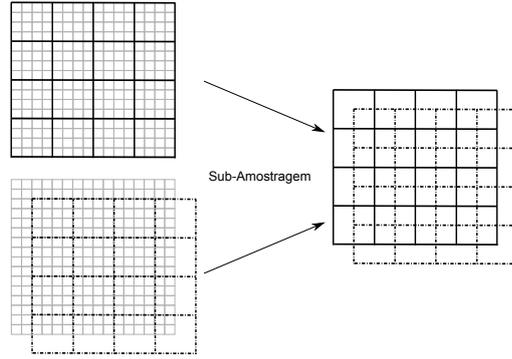


Figura 5.8: Ilustração de duas sub-amostragens de uma grade de alta resolução, provocando o deslocamento de ordem sub-pixel entre as grades de baixa resolução.

Como discutido na Seção 2.2.2, considerando que existem $q \in \mathbb{N}$ imagens de baixa resolução g_k , $k = 1, \dots, q$, as imagens de alta resolução correspondentes serão estimadas utilizando uma abordagem de janela deslizante (Figura 2.5). Assim, \hat{f}_k será estimada considerando um subconjunto das observações $g = [g_{k-n}, \dots, g_k, \dots, g_{k+n}]$, sendo g_k e imagem de referência. A solução MAP decide pela estimação que maximiza a densidade de probabilidades condicionais de f_k dadas todas as observações em g ,

$$\hat{f}_k = \arg \max_{f_k} \{p(f_k|g)\}. \quad (5.3)$$

Como apresentado anteriormente, maximizar a probabilidade *a posteriori* conjunta em geral exige alta poder computacional. Assim, o algoritmo ICM pode ser utilizado para alcançar uma aproximação da solução MAP. Utilizando um modelo *a priori* de MRF, esse algoritmo atualiza os labels f_k^i de cada pixel da imagem de alta resolução f_k , $i = 1, \dots, M^2$, maximizando sequencialmente as probabilidades *a posteriori* locais $P(f_k^i|g, f_k^{\eta_i})$, onde $f_k^{\eta_i}$ é o conjunto de vizinhos do pixel f_k^i dado o sistema de vizinhança η .

Como ilustrado na Figura 5.9, com base nos deslocamentos de ordem sub-pixel existentes entre as imagens observadas, no alinhamento dessas imagens com a grade de alta resolução, um pixel de alta resolução influencia um sub-conjunto de todos os pixels de baixa resolução. Considerando g^i o conjunto de pixels de baixa resolução influenciados pelo pixel de alta resolução f_k^i , pelo teorema de Bayes, $P(f_k^i|g, f_k^{\eta_i})$ pode ser aproximado por

$$P(f_k^i|g, f_k^{\eta_i}) \sim p(g^i|f_k^i)p(f_k^i|f_k^{\eta_i}). \quad (5.4)$$

Nesse contexto, o algoritmo ICM é dado por:

1. Defina um modelo de MRF para os valores de f_k^i ;

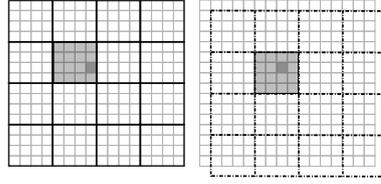


Figura 5.9: Ilustração de dois pixels observados que sobrepõem um pixel de alta resolução no alinhamento das imagens de baixa resolução com a grade de alta resolução.

2. Escolha uma estimativa inicial de alta resolução;
3. Para i de 1 a M^2 , atualize f_k^i pelo valor no intervalo de níveis de cinza que maximiza $p(g^i | f_k^i) p(f_k^i | f_k^{\eta_i})$;
4. Repita até que nenhuma modificação seja feita ou um número máximo de iterações.

Em seu projeto de mestrado a aluna utilizou o modelo de MRF de Potts (ou modelo *Multi-Level Logistic* (MLL) isotrópico, ou modelo de Ising generalizado)

$$p(f_k^i | f_k^{\eta_i}) \sim \exp \{ \beta \cdot \#\{t \in \eta_i | f_k^i = f_k^t\} \}, \quad (5.5)$$

para definir o conjunto de todas as distribuições condicionais *a priori*. Entretanto, nesse modelo, se dois vizinhos não possuem exatamente o mesmo label, eles não contribuem em nada para a distribuição, mesmo quando possuem valores próximos.

Li (2009) apresenta um modelo baseado na distribuição de Gibbs que incorpora a similaridade entre pixels de maneira mais suave. Esse modelo é definido como

$$p(f_k^i | f_k^{\eta_i}) = \frac{1}{Z} \exp \{ -U(f_k^i | f_k^{\eta_i}) \}, \quad (5.6)$$

onde a função potencial $U(f_k^i | f_k^{\eta_i})$ é dada por

$$U(f_k^i | f_k^{\eta_i}) = \sum_{i' \in \eta_i} \beta \left[1 - 2 \exp \left(-(f_k^i - f_k^{i'})^2 \right) \right]. \quad (5.7)$$

Z é chamada função de partição e β pode ser visto como um coeficiente de interação entre vizinhos. Esse modelo foi denominado MLL isotrópico generalizado (GIMLL). No contexto da reconstrução por super-resolução, em comparação com outros modelos, o GIMLL apresentou os melhores resultados (Martins *et al.*, 2009a,b). No Capítulo 6 essa comparação, utilizando as imagens do trato vocal e outros modelos além do modelo de Potts e do GIMLL, é discutida com mais detalhes e alguns resultados são apresentados. Esses resultados foram publicados no

periódico internacional *Integrated Computer-Aided Engineering* (ICAE), volume 18, número 2 de 2011 (Martins *et al.*, 2011).

Considerando o modelo de formação das imagens (Equação (5.1)), na presença de ruído Gaussiano independente de média zero, a distribuição de verossimilhança é dada por

$$p(g|f_k) = \frac{1}{(2\pi\sigma^2)^{q\frac{N^2}{q}}} \exp \left\{ - \sum_{n=1}^q \frac{\|g_n - D_n \hat{f}_k\|^2}{2\sigma^2} \right\}, \quad (5.8)$$

onde σ^2 é a variância do ruído. Com base nisso, $p(g^i|f_k^i)$ pode ser aproximada por

$$p(g^i|f_k^i) \sim \exp \left\{ - \sum_{n=1}^r \frac{\|(g^i)_n - (D^i)_n \hat{f}_k\|^2}{2\sigma^2} \right\}, \quad (5.9)$$

onde r é o número de pixels de baixa resolução influenciados pelo pixel f_k^i , $(g^i)_n$ é o n -ésimo pixel de baixa resolução influenciado por f_k^i , e $(D^i)_n \hat{f}_k$ é o pixel correspondente gerado pela estimativa \hat{f}_k .

Assim, com base nas Equações (5.6), (5.7) e (5.9), e negligenciando termos constantes, a maximização de $P(f^i|g, f^{n_i})$ é equivalente a

$$\arg \max_{f_k^i} p(g^i|f_k^i) \cdot p(f_k^i|f_k^{n_i}) \sim \arg \min_{f_k^i} \left[U(f_k^i|f_k^{n_i}) + \sum_{n=1}^r \frac{\|(g^i)_n - (D^i)_n \hat{f}_k\|^2}{2\sigma^2} \right]. \quad (5.10)$$

Como discutido em Martins *et al.* (2009b), a estimativa do parâmetro β pode ser feita seguindo um procedimento similar ao proposto por Levada e Tannús (2008). Entretanto, nos experimentos desenvolvidos no Capítulo 6, esse parâmetro foi decidido empiricamente.

No projeto de mestrado que embasou este projeto de doutorado foi adotado um sistema de vizinhança de segunda ordem onde apenas a relação do pixel com seus oito vizinhos na imagem era considerada, Figura 5.10(a). Entretanto, no contexto das imagens do trato vocal, como se trata de sequências semelhantes a frames de um vídeo, é possível explorar a relação que existe entre imagens consecutivas em uma sequência. Dessa forma, no projeto de doutorado, pretendia-se adotar um modelo de MRF tridimensional (semelhante ao discutido em Borman e Stevenson (1999)), no qual a relação do pixel com seus vizinhos nas imagens anteriores e posteriores à imagem corrente também seria considerada. Dois sistemas de vizinhança tridimensionais são ilustrados na Figura 5.10(b). Além disso, outro objetivo era comparar os resultados do modelo GIMLL com o modelo de MRF Gaussiano (GMRF). Entretanto, apesar de ter apresentado resultados promissores no contexto deste projeto, devido à dimensão do problema considerado, o algoritmo ICM apresentou alto custo computacional, mesmo conside-

rando apenas o sistema de vizinhança bidimensional. Assim, o uso de um sistema de vizinhança tridimensional apenas agravaria essa limitação.

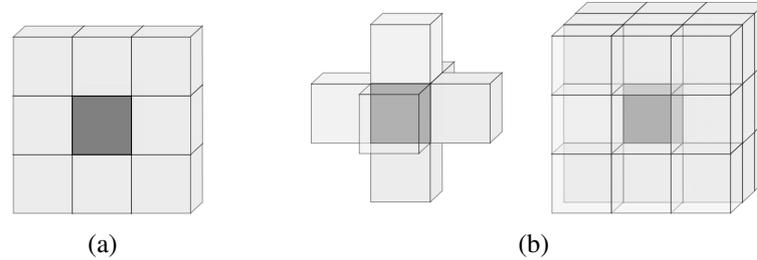


Figura 5.10: (a) Sistema de vizinhança bidimensional de segunda ordem; (b) Dois sistemas de vizinhança tridimensionais.

Considerando as limitações de desempenho do algoritmo ICM e com base no fato de que quando os dados e o modelo *a priori* são conjuntamente Gaussianos a abordagem MAP equivale à abordagem de mínimo erro médio quadrático (MMSE - *Minimum Mean Square Error*), a aluna e seu orientador decidiram verificar a viabilidade de se aplicar o filtro de Wiener discreto no contexto das imagens do trato vocal. A ideia seria adaptar o filtro discreto para a reconstrução por super-resolução. Para verificar se a abordagem seria de fato factível nesse contexto, o método foi implementado e comparado com os resultados anteriores do modelo GIMLL. Em comparação com a abordagem inicialmente investigada, o novo método apresentou resultados promissores com custo computacional bastante reduzido como mostrado no Capítulo 6. A nova abordagem se baseia no trabalho de Mascarenhas *et al.* (1996), com o qual o trabalho de Hardie (2007) possui várias similaridades. A seguir esses dois trabalhos, juntamente com as possibilidades exploradas, são discutidos.

5.3.2 Métodos Baseados no Filtro de Wiener

Como descrito na Seção 2.3.6, Mascarenhas *et al.* (1996) discutem um método de interpolação para fusão de dados de satélite utilizando técnicas estatísticas Bayesianas. Tanto os pixels observados como os pixels na grade interpolada são considerados variáveis aleatórias. A estimação linear dos pixels interpolados é feita de acordo com o critério de mínimo erro médio quadrático (MMSE - *Minimum Mean Square Error*), por meio da solução de um sistema que envolve as correlações espaciais e espectrais das observações. Assume-se a hipótese de separabilidade da estrutura de correlação nas direções horizontal, vertical e espectral e um modelo Markoviano de primeira ordem é utilizado para modelar as estruturas de correlação espaciais.

Na proposta inicial para o aumento de resolução espacial foi adotado o modelo de formação das imagens mostrado na Equação (5.1), no qual o operador D_k modela a função do sensor

de aquisição da imagem de acordo com os deslocamentos presentes entre as observações (na prática um operador de média espacial, atribuindo a média de um bloco de alta resolução ao pixel de baixa resolução relacionado). Mascarenhas *et al.* (1996) não consideram esse operador na formação da imagem multiespectral observada. Os pixels em cada banda apenas são dispostos em uma grade mais grossa, e os pixels interpolados são estimados entre os pixels observados em uma grade mais fina (Figura 2.10). Verificando como essa abordagem poderia ser aplicada ao contexto das imagens do trato vocal, chegou-se a duas possibilidades denominadas *Interpolação Estatística* e *Abordagem Multitemporal*, ambas desconsiderando o operador D_k :

1. *Interpolação Estatística*: Os pixels de um subconjunto das imagens observadas (de acordo com a abordagem de janela deslizante - Figura 2.5) são dispostos na grade de alta resolução de acordo com o fator de escala e os deslocamentos causados pelas deformações. A estimação proposta por Mascarenhas *et al.* (1996) é aplicada como se existisse apenas uma banda, ou seja, considerando apenas a correlação espacial entre as observações (estruturas de correlação horizontal e vertical), desconsiderando a correlação espectral.
2. *Abordagem Multitemporal*: Cada imagem de baixa resolução observada é considerada uma banda multiespectral, e os deslocamentos presentes entre elas são desconsiderados na estimação. Nesse caso, como as imagens são adquiridas ao longo do tempo, o termo *multitemporal* pode substituir o termo *multiespectral*. A estimação proposta por Mascarenhas *et al.* (1996) é aplicada sem grandes modificações.

Diferente da abordagem proposta por Mascarenhas *et al.* (1996), Hardie (2007) considera o modelo de formação das imagens mostrado na Equação (5.1), no qual o operador D_k modela a função do sensor de aquisição da k -ésima imagem de baixa resolução g_k de acordo com os deslocamentos presentes entre as observações (Equação (5.2)). Além disso, o modelo de autocorrelação paramétrico simétrico

$$R(x, y) = \sigma^2 \rho^{\sqrt{x^2 + y^2}} \quad (5.11)$$

é utilizado para modelar a estrutura de correlação dos pixels de alta resolução, como forma de inserir conhecimento *a priori* no processo de estimação. σ^2 é a variância da imagem desejada e ρ controla o decaimento da correlação com a distância. A terceira abordagem investigada, denominada *Filtro de Wiener Adaptativo*, se baseia na abordagem proposta por Hardie (2007).

A seguir, essas três propostas são detalhadas.

Interpolação Estatística

Nessa abordagem, a cada passo, considerando um subconjunto das imagens observadas $g = [g_{k-n}, \dots, g_k, \dots, g_{k+n}]$, sendo g_k a imagem de referência, é estimada a imagem de alta resolução f_k , correspondente à essa imagem de referência. Primeiramente, os pixels de baixa resolução de g são dispostos na grade de alta resolução de acordo com o fator de escala e os deslocamentos presentes entre eles. Esse processo é ilustrado na Figura 5.11. É importante notar que essa figura ilustra apenas a presença de translações globais entre as imagens observadas. No caso das imagens do trato vocal, devido às deformações presentes entre as imagens de uma sequência, os deslocamentos são bem mais complexos.

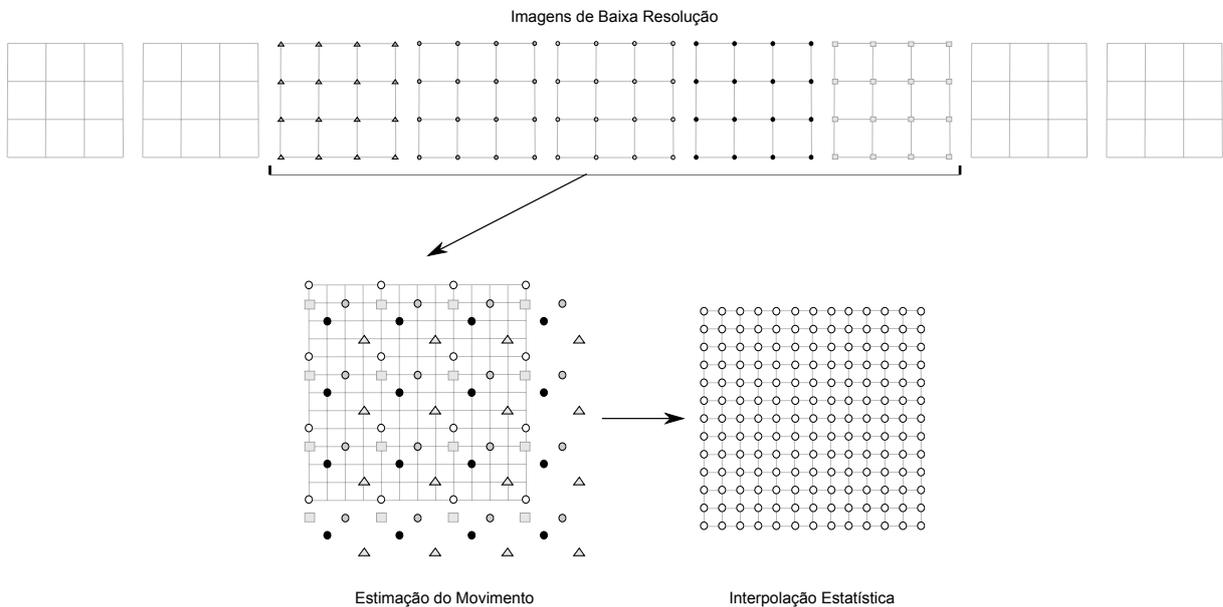


Figura 5.11: Pixels observados dispostos na grade de alta resolução de acordo com os deslocamentos presentes entre eles.

A estimação de subconjuntos de pixels de alta resolução é feita da seguinte forma. De forma semelhante à abordagem proposta por Hardie (2007), é empregada uma janela deslizante de observação. Como ilustrado na Figura 5.12, essa janela cobre W_x espaçamentos de um pixel de alta resolução na direção horizontal, e W_y espaçamentos de um pixel de alta resolução na direção vertical. Todos os pixels observados que se encontram nessa janela de observação são inseridos em um vetor de observação $g^i = [g^{i,1}, g^{i,2}, \dots, g^{i,k_i}]$, onde i indica a posição da janela na grade de alta resolução e k_i indica o número de pixels de baixa resolução que se encontram nessa janela. Como ilustrado na figura, para cada janela de observação são estimados os pixels de alta resolução presentes em uma subjanela de dimensão $D_x \times D_y$, com $1 \leq D_x \leq W_x$ e $1 \leq D_y \leq W_y$. Considerando que esses pixels de alta resolução estão dispostos em um vetor

$\hat{f}_k^i = [\hat{f}_k^{i,1}, \hat{f}_k^{i,2}, \dots, \hat{f}_k^{i,D_x D_y}]$, eles são estimados de forma semelhante ao estimador proposto por Mascarenhas *et al.* (1996):

$$\hat{f}_k^i = E[f_k^i] + \Sigma_{f_k^i g^i} \Sigma_{g^i g^i}^{-1} (g^i - E[g^i]). \quad (5.12)$$

Assume-se que $E[g^i]$ seja formado por um vetor de dimensão $k_i \times 1$, com valores iguais à média das imagens observadas. Além disso, considera-se que a média das imagens não deve ser alterada após a interpolação. Portanto, $E[f_k^i]$ também será formado por um vetor de dimensão $D_x D_y \times 1$, com valores iguais à média das imagens observadas.

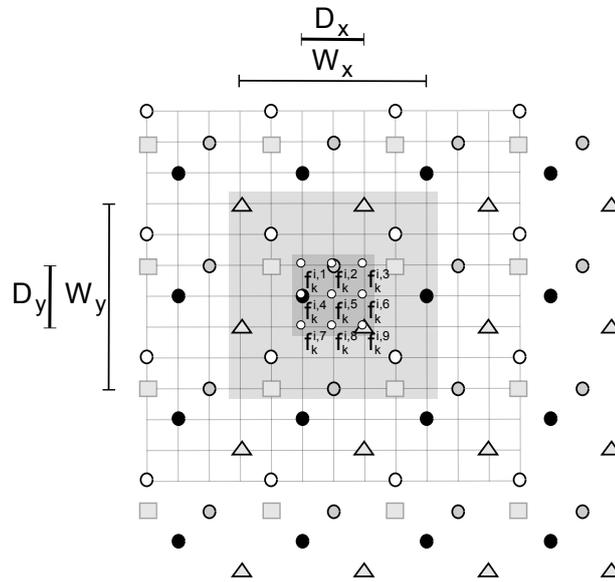


Figura 5.12: Estimação de blocos de pixels de alta resolução.

De acordo com os trabalhos de Mascarenhas *et al.* (1996) e Hardie (2007), para obter as matrizes de covariância $\Sigma_{f_k^i g^i}$ e $\Sigma_{g^i g^i}$ existem duas possibilidades. Uma delas é considerar a separabilidade nas direções horizontal e vertical. Assim, a matriz de covariância entre f_i e g_i é dada por

$$\Sigma_{f_k^i g^i} = (R_H)_{f_k^i g^i} \otimes (R_V)_{f_k^i g^i} \quad (5.13)$$

e a matriz de covariância de g^i é dada por

$$\Sigma_{g^i g^i} = (R_H)_{g^i g^i} \otimes (R_V)_{g^i g^i}, \quad (5.14)$$

sendo que R_H é a matriz de coeficientes de correlação horizontal, R_V a matriz de coeficientes de correlação vertical, e o símbolo \otimes representa o produto de Kronecker. Assumindo que pixels adjacentes ao longo de uma linha da imagem possuem correlação $0 \leq \rho_H \leq 1$, e auto-correlação

igual a 1, a matriz de covariância se reduz a

$$R_H = \sigma_H^2 \begin{bmatrix} 1 & \rho_H & \dots & \rho_H^{N-1} \\ \rho_H & 1 & \dots & \rho_H^{N-2} \\ \vdots & \vdots & & \vdots \\ \rho_H^{N-1} & \rho_H^{N-2} & \dots & 1 \end{bmatrix}, \quad (5.15)$$

onde σ_H^2 é a variância dos pixels ao longo das linhas e N o número de elementos em cada linha (R_V é descrita de forma semelhante). De acordo com Pratt (2007), esse é um exemplo de matriz de covariância de um processo de Markov e experimentos mostram que $\rho = 0.95$ é uma aproximação razoável. A segunda possibilidade para o cálculo das matrizes de covariância $\Sigma_{f_k^i g^i}$ e $\Sigma_{g^i g^i}$ é caracterizá-las pelo modelo isotrópico utilizado por Hardie (2007) (Equação (5.11)). Os resultados utilizando ambos os modelos são discutidos no Capítulo 6.

É importante ressaltar que, como pode ser visto na Figura 5.12, os pixels de baixa resolução não estão uniformemente distribuídos na janela de observação, e suas posições variam de acordo com a posição dessa janela. Dessa forma, $\Sigma_{g^i g^i}$ e $\Sigma_{f_k^i g^i}$ serão recalculadas a cada posição da janela de observação.

Abordagem Multi-Temporal

Nessa abordagem cada imagem de baixa resolução é considerada uma banda multiespectral e a interpolação estatística proposta por Mascarenhas *et al.* (1996) é aplicada sem grandes modificações. Assim, diferente da abordagem anterior, nessa abordagem a cada n imagens de baixa resolução $g = [g_k, \dots, g_{k+n}]$, são estimadas n imagens de alta resolução $f = [f_k, \dots, f_{k+n}]$. Esse processo é ilustrado na Figura 5.13. Como as imagens da sequência são adquiridas ao longo do tempo, essa abordagem é denominada *multitemporal*. É importante notar que, nessa abordagem, os deslocamentos presentes entre as imagens observadas são ignorados na estimação.

Seja g uma sequência de n imagens de baixa resolução observadas, $g(x, y, k)$ representa o nível de cinza nas coordenadas (x, y) da k -ésima imagem da sequência. Considerando novamente uma janela de observação deslizante que cobre o espaçamento de W_x pixels de alta resolução na direção horizontal e de W_y pixels de alta resolução na direção vertical, os pixels das n imagens de baixa resolução da sequência que se encontram nessa janela são inseridos em g^i , onde i indica a posição da janela. Essa inserção é feita seguindo uma ordenação lexicográfica (empilhando todas linhas de cada imagem da sequência). A Figura 5.14 ilustra a situação em que a janela de observação possui dimensão 5×5 . Dessa forma, de acordo com a figura, $g^i = [g(1, 1, 1), g(1, 2, 1), \dots, g(3, 3, 1), g(1, 1, 2), \dots, g(3, 3, n)]$. Para cada janela de

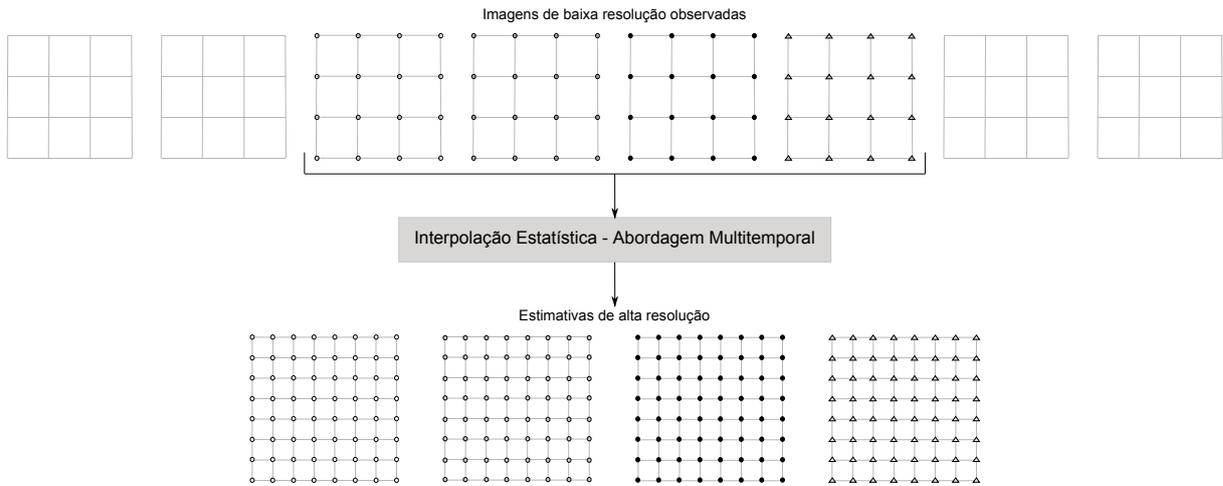


Figura 5.13: Estimação de quatro imagens de alta resolução a partir do mesmo número de imagens de baixa resolução observadas.

observação são estimados os pixels de alta resolução presentes em uma subjanela de dimensão $D_x \times D_y$, com $1 \leq D_x \leq W_x$ e $1 \leq D_y \leq W_y$. Esses pixels também são inseridos em um vetor f^i seguindo uma ordenação lexicográfica. Na Figura 5.14 a dimensão dessa subjanela da janela de observação é 2×2 , e $f^i = [f(1, 1, 1), f(1, 2, 1), \dots, f(2, 2, 1), f(1, 1, 2), \dots, f(2, 2, n)]$.

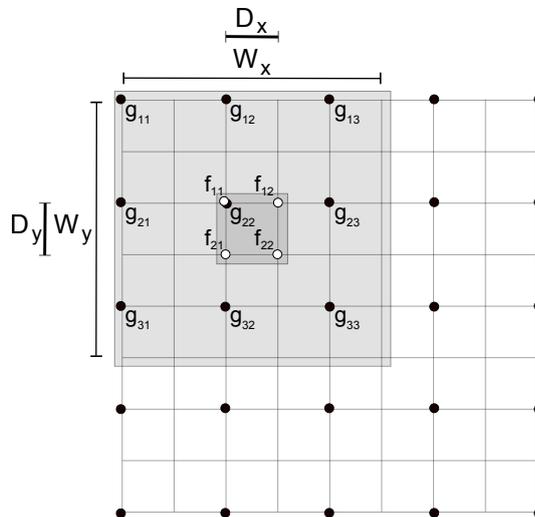


Figura 5.14: Estimação de cada banda multitemporal.

Seguindo Mascarenhas *et al.* (1996), o estimador de pixels multitemporais interpolados é dado por

$$\hat{f}^i = E[f^i] + \Sigma_{f^i g^i} \Sigma_{g^i g^i}^{-1} (g^i - E[g^i]). \quad (5.16)$$

Assim como na abordagem anterior, assume-se que $E[g^i]$ seja formado por um vetor com valores iguais à média das imagens observadas. Além disso, considera-se que a média das

imagens não deve ser alterada após a interpolação. Portanto, $E[f^i]$ também será formado por um vetor com valores iguais à média das imagens observadas.

Novamente existem duas possibilidades para obter as matrizes de covariância $\Sigma_{f^i g^i}$ e $\Sigma_{g^i g^i}$. A primeira possibilidade é admitir a hipótese de separabilidade da estrutura de correlação nas direções horizontal, vertical, e temporal. Assim, a matriz de covariância entre f^i e g^i é dada por

$$\Sigma_{f^i g^i} = (R_H)_{f^i g^i} \otimes (R_V)_{f^i g^i} \otimes \Sigma_T \quad (5.17)$$

e a matriz de covariância de g^i é dada por

$$\Sigma_{g^i g^i} = (R_H)_{g^i g^i} \otimes (R_V)_{g^i g^i} \otimes \Sigma_T, \quad (5.18)$$

sendo que R_H é a matriz de coeficientes de correlação horizontal, R_V a matriz de coeficientes de correlação vertical, Σ_T a matriz de covariância temporal, e o símbolo \otimes representa o produto de Kronecker. Seguindo um processo de Markov, R_H e R_V se reduzem a

$$R = \sigma^2 \begin{bmatrix} 1 & \rho_H & \dots & \rho_H^{N-1} \\ \rho_H & 1 & \dots & \rho_H^{N-2} \\ \vdots & \vdots & & \vdots \\ \rho_H^{N-1} & \rho_H^{N-2} & \dots & 1 \end{bmatrix}, \quad (5.19)$$

onde σ^2 é a variância dos pixels ao longo das linhas ou colunas e N o número de elementos na horizontal ou vertical, respectivamente. Novamente, a segunda possibilidade para o cálculo das matrizes de covariância $\Sigma_{f_k^i g^i}$ e $\Sigma_{g^i g^i}$ é caracterizá-las pelo modelo isotrópico utilizado por Hardie (2007) (Equação (5.11)). Os resultados utilizando ambos os modelos na Abordagem Multitemporal também são discutidos no Capítulo 6.

Nessa abordagem, como os pixels observados e a serem estimados estão uniformemente distribuídos na grade de alta resolução, é importante notar que, dependendo das dimensões da janela de observação e da subjunela das estimações, $\Sigma_{g^i g^i}$ e $\Sigma_{f^i g^i}$ não se modificam com a posição dessas janelas. Por exemplo, no caso do modelo Markoviano separável, de acordo com a Figura 5.14,

$$(R)_{g^i g^i} = \begin{bmatrix} 1 & \rho^2 & \rho^4 \\ \rho^2 & 1 & \rho^2 \\ \rho^4 & \rho^2 & 1 \end{bmatrix} \quad (5.20)$$

e

$$(R)_{f^i g^i} = \begin{bmatrix} \rho^2 & 1 & \rho^2 \\ \rho^3 & \rho & \rho \end{bmatrix}. \quad (5.21)$$

Como é possível notar na figura, a especificação das potências de ρ na matriz $R_{g^i g^i}$ depende da distâncias entre os pixels observados, e na matriz $R_{f^i g^i}$ da distância entre os pixels observados e interpolados. O fato de que essas matrizes não precisam ser recalculadas a cada posição da janela de observação diminui consideravelmente o custo computacional dessa abordagem com relação à anterior. A matriz de covariância temporal, que é igual à matriz de covariância entre as imagens da sequência, é dada por

$$\Sigma_T = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}, \quad (5.22)$$

sendo σ_{uv}^2 a covariância entre a u -ésima e a v -ésima imagens, e σ_{uu}^2 a variância da u -ésima imagem. Elas são calculadas diretamente da janela de observação.

Filtro de Wiener Adaptativo

Seja g um vetor formado por um subconjunto das imagens de baixa resolução observadas $g = [g_{k-n}, \dots, g_k, \dots, g_{k+n}]$, sendo g_k a imagem de referência, o modelo de formação dessas imagens pode ser definido como

$$g = Df_k + n, \quad (5.23)$$

sendo f_k a imagem de alta resolução correspondente à imagem de referência g_k . De acordo com esse modelo, o processo de estimação de f_k pode ser ilustrado pela Figura 5.15.

Nesse contexto, considerando o modelo de formação mostrado na Equação (5.23) e supondo que a matriz de autocorrelação da imagem a ser estimada $\Sigma_{f_k f_k}$ é definida por um dos modelos utilizados por Mascarenhas *et al.* (1996) e Hardie (2007), a matriz de autocorrelação das observações Σ_{gg} será dada por

$$\Sigma_{gg} = D\Sigma_{f_k f_k}D^T + \Sigma_{nn}, \quad (5.24)$$

onde Σ_{nn} é a autocorrelação do ruído. Além disso, a matriz de correlação cruzada entre a imagem a ser estimada e as observações $\Sigma_{f_k g}$ será dada por

$$\Sigma_{f_k g} = D\Sigma_{f_k f_k}. \quad (5.25)$$

Nesse contexto, a estimação pixel a pixel ocorre da seguinte forma:

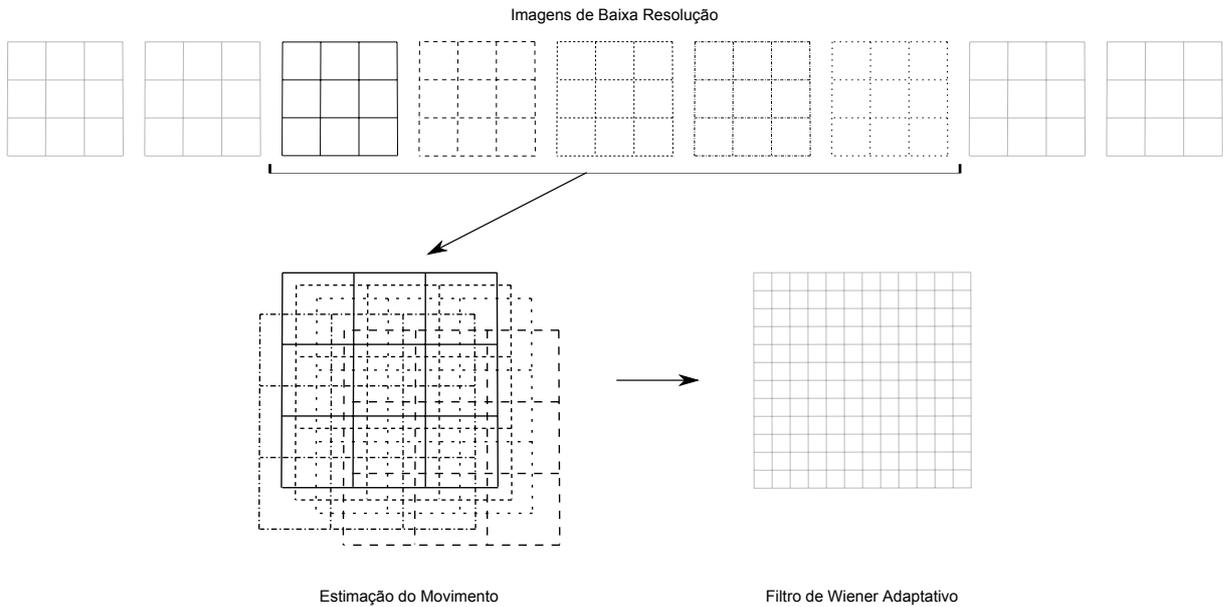


Figura 5.15: Estimação de uma imagem de alta resolução a partir de um subconjunto das imagens de baixa resolução observadas, considerando o modelo de formação das imagens mostrado na Equação (5.23).

1. Para cada pixel de alta resolução f_k^i , $i = 1, \dots, M^2$
2. Adicione ao vetor $g_{(\text{Índices})}^i$ todos os índices dos pixels de baixa resolução que são influenciados pelo pixel de alta resolução f_k^i . Isso é feito verificando no operador D , quais linhas possuem valores diferentes de zero na coluna correspondente ao pixel f_k^i ;
3. Adicione ao vetor g^i todos os pixels de baixa resolução que são influenciados pelo pixel de alta resolução f_k^i . Isso é feito verificando no operador D , quais linhas possuem valores diferentes de zero na coluna correspondente ao pixel f_k^i ;
4. Para cada pixel em g^i , adicione a esse vetor seus 8 vizinhos mais próximos como ilustrado na Figura 5.16, e adicione os respectivos índices ao vetor $g_{(\text{Índices})}^i$;
5. Para cada pixel em g^i , adicione a $f_{(\text{Índices})}^i$ os índices dos pixels de alta resolução que influenciam esse pixel de baixa resolução. Isso é feito verificando no operador D , quais colunas possuem valores diferentes de zero na linha correspondente ao pixel de baixa resolução em g^i ;
6. Defina D^i tomando apenas as linhas de índices $g_{(\text{Índices})}^i$ e colunas de índices $f_{(\text{Índices})}^i$ de D ;
7. Defina $\Sigma_{f^i f^i}$ referente aos índices em $f_{(\text{Índices})}^i$, de acordo com o modelo definido (por exemplo o modelo utilizado por Hardie (2007));

8. Defina $\Sigma_{n^i n^i}$ como $\sigma_n^2 I$, sendo σ_n^2 a variância do ruído e I a matriz identidade;
9. Defina $\Sigma_{g^i g^i} = D^i \Sigma_{f^i f^i} D^{iT} + \Sigma_{n^i n^i}$;
10. Defina $\Sigma_{f^i g^i} = D^i \Sigma_{f^i f^i}$;
11. Defina $W^i = \Sigma_{g^i g^i}^{-1} \Sigma_{f^i g^i}$;
12. Normalize as colunas de W_i ;
13. Defina $f^i = W^{iT} g^i$;
14. Atribua a f_k^i o valor de índice correspondente em f^i .

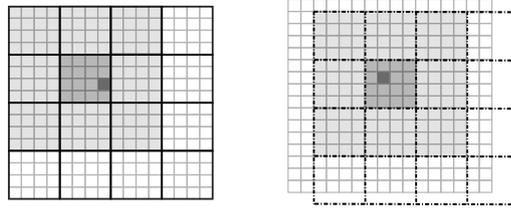


Figura 5.16: Ilustração das áreas dos pixels de baixa resolução utilizados na construção das matriz de covariância.

A estimação pode ser feita pixel a pixel ou considerando um conjunto de pixels utilizando uma janela de observação deslizante como descrito nas duas abordagens anteriores. Nesse caso, a execução do método ocorre de forma semelhante à explicada acima, porém f_k^i será uma sub-janela.

Filtro de Wiener Adaptativo - Considerando o Erro do Registro

Nas abordagens de SRIR, quando o registro não é perfeito, podem surgir artefatos nas regiões em que o erro do registro é grande. De forma semelhante ao trabalho de Borman e Stevenson (1999), para solucionar esse problema, a variância do ruído pode ser considerada proporcional ao erro absoluto do registro. Assim, na estimação discutida acima, $\Sigma_{n^i n^i}$ é definida como

$$\Sigma_{n^i n^i} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}, \quad (5.26)$$

sendo σ_k , $k = 1, \dots, n$, proporcional ao erro do registro para o k -ésimo pixel de baixa resolução. Como discutido no Capítulo 6, esse procedimento atenua consideravelmente os artefatos.

Experimentos Comparativos

O Projeto Heron, liderado pelo Instituto de Engenharia Eletrônica e Telemática de Aveiro (IE-ETA), Portugal, modelou um ambiente computacional para investigação em síntese articulatória do português. Isso foi feito por meio de sequências de imagens de ressonância magnética do trato vocal, adquiridas durante a emissão da fala de palavras ou fonemas. Através dessas sequências, foi possível modelar o movimento dos articuladores da fala, relacionados a palavras e fonemas específicos. As imagens foram adquiridas utilizando um equipamento de 1.5 Tesla (Magnetom Symphony, Maestro Class, Siemens, Erlanger, Alemanha) equipado com gradientes Quantum (máxima amplitude - 30mT/m; rise time - 0.24ms; Slew rate - 125 T/m/s; FOV - 50cm). Foram utilizadas antenas de crânio e de pescoço, simultaneamente, em todas as aquisições. Cada palavra foi repetida durante 20 segundos, o que dá em média 15 a 16 repetições da mesma palavra. A aquisição permite obter uma resolução temporal de 5 frames por segundo, o que dá cerca de 100 frames para cada palavra.

6.1 Procedimento estatístico

Em Estatística, uma simples observação das médias ou medianas de uma amostra de resultados não é suficiente para inferir algo sobre a população real. Isso acontece porque as diferenças observadas podem ser uma coincidência causada pela amostragem aleatória (no caso deste trabalho, as amostras de imagens utilizadas). Para verificar se as diferenças alcançadas são de fato

significantes, testes de hipótese podem ser aplicados. No caso dos resultados avaliados neste trabalho, como cada método é aplicado ao mesmo conjunto de imagens, testes pareados são os mais adequados. Quando comparados aos testes estatísticos não-pareados, os testes pareados oferecem maior precisão à inferência (Montgomery, 2006).

Antes da escolha do teste estatístico mais adequado a ser aplicado, deve-se verificar também se os resultados obtidos seguem uma distribuição normal. O teste de Shapiro-Wilk pode ser utilizado para verificar a normalidade das amostras. Nos resultados que serão discutidos a seguir, os testes indicaram que as observações obtidas não são normais. Dessa forma, os testes estatísticos não-paramétricos são mais indicados.

Para verificar a existência de diferença significativa entre um conjunto de abordagens de tamanho maior do que dois, utilizou-se a análise de variância não-paramétrica de Kruskal-Wallis. Nos casos em que se comparam apenas duas abordagens, adotou-se o teste pareado não-paramétrico de Wilcoxon. Para a significância estatística, adotou-se o nível de confiança clássico de 95%. Dessa forma, nas análises feitas considera-se que p-valores abaixo de 0,05 são significativos. Para todos os testes estatísticos realizados, utilizou-se a linguagem e ambiente R (<http://www.r-project.org/>).

6.2 Aumento de Resolução Temporal

Com o intuito de avaliar visualmente as abordagens propostas para o aumento de resolução temporal, considerando sequências de imagens de ressonância magnética do trato vocal, de dimensão 256×256 , foram geradas imagens intermediárias às imagens observadas. A Figura 6.1 mostra um recorte da região que concentra as deformações das quatro primeiras imagens de uma sequência observada. As Figura 6.2(b) e 6.2(d) apresentam imagens geradas entre a segunda e a terceira imagens dessa sequência por meio de interpolação linear e por interpolação por splines cúbicas, respectivamente. Considerando que as quatro imagens observadas foram adquiridas nos momentos $t = 1, \dots, 4$, as imagens intermediárias foram geradas no momento $t = 2.5$. É possível notar que não existem artefatos e, como indicado pelas malhas geradas (Figuras 6.2(f) e 6.2(h)), as imagens são coerentes com o movimento existente na sequência.

A fim de comparar objetivamente as duas abordagens, elas foram utilizadas para interpolar um conjunto de 25 imagens de ressonância magnética do corte sagital do trato vocal. Essas imagens foram fornecidas pelos professores Dr. António Joaquim Silva Teixeira e Dr. Augusto Marques Ferreira Silva do IEETA. No primeiro experimento, cada uma das 25 imagens (com exceção da primeira e da última) foi removida e os dois métodos foram utilizados para gerar

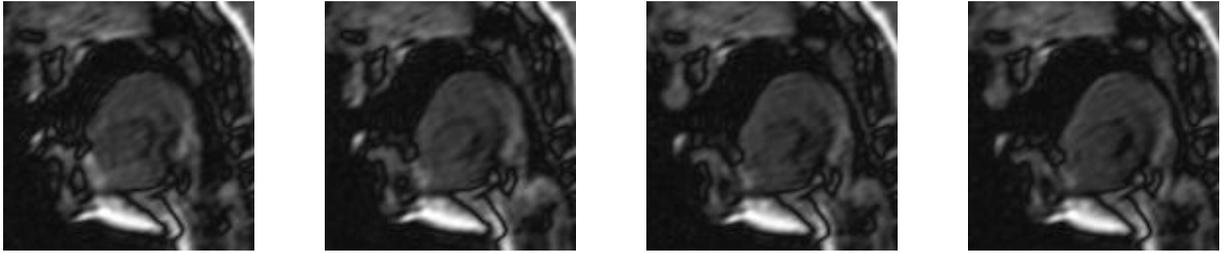


Figura 6.1: Recorte da região que concentra as deformações das primeiras quatro imagens de uma sequência observada.

uma versão interpolada da imagem removida. O erro médio quadrático normalizado (NMSE) entre as imagens interpoladas e reais foi utilizado como medida de erro nessa avaliação.

Como ilustrado na Figura 6.3, no segundo experimento as imagens foram removidas duas a duas consecutivamente. Novamente o NMSE entre as imagens interpoladas e reais foi utilizado na avaliação.

No terceiro experimento, a sequência observada foi sub-amostrada temporalmente como mostra a Figura 6.4, e, novamente, as imagens interpoladas foram comparadas às reais.

A Tabela 6.1 mostra as médias μ_1 , μ_2 e μ_3 dos experimentos 1, 2 e 3, respectivamente. É possível notar que apenas no primeiro experimento o NMSE da interpolação por splines foi mais alto do que o erro da interpolação linear na direção do movimento.

Tabela 6.1: Médias dos MSDs dos 3 experimentos.

	μ_1	μ_2	μ_3
Spline	40.06626	48.61732	48.87994
Linear	39.46613	54.75909	51.08056

Como comentado anteriormente, para analisar se a diferença entre as médias é estatisticamente significativa utilizou-se o teste de Wilcoxon com 95% de confiança. Todos os testes estatísticos evidenciaram que não há diferença significativa entre μ_1^{spline} e μ_1^{linear} , μ_2^{spline} e μ_2^{linear} , e μ_3^{spline} e μ_3^{linear} (p-valores avaliados em 0.754, 0.09173 e 0.1591, respectivamente). Para se ter uma ideia visual desses resultados, a Figura 6.5 apresenta o *boxplot* dos dados obtidos nos 3 experimentos. É importante notar que se o nível de confiança adotado fosse 90%, as médias poderiam ser consideradas significativamente diferentes, favorecendo o método baseado na interpolação por splines. O *boxplot* mostrado na Figura 6.5(b) é um indício visual dessa afirmação. Entretanto, considerando que com 95% de confiança, não existe diferença significativa entre os métodos, o aumento de resolução temporal utilizando interpolação linear na direção do movimento foi adotado por se tratar de um procedimento mais simples e, conseqüentemente, apresentar menor custo computacional.

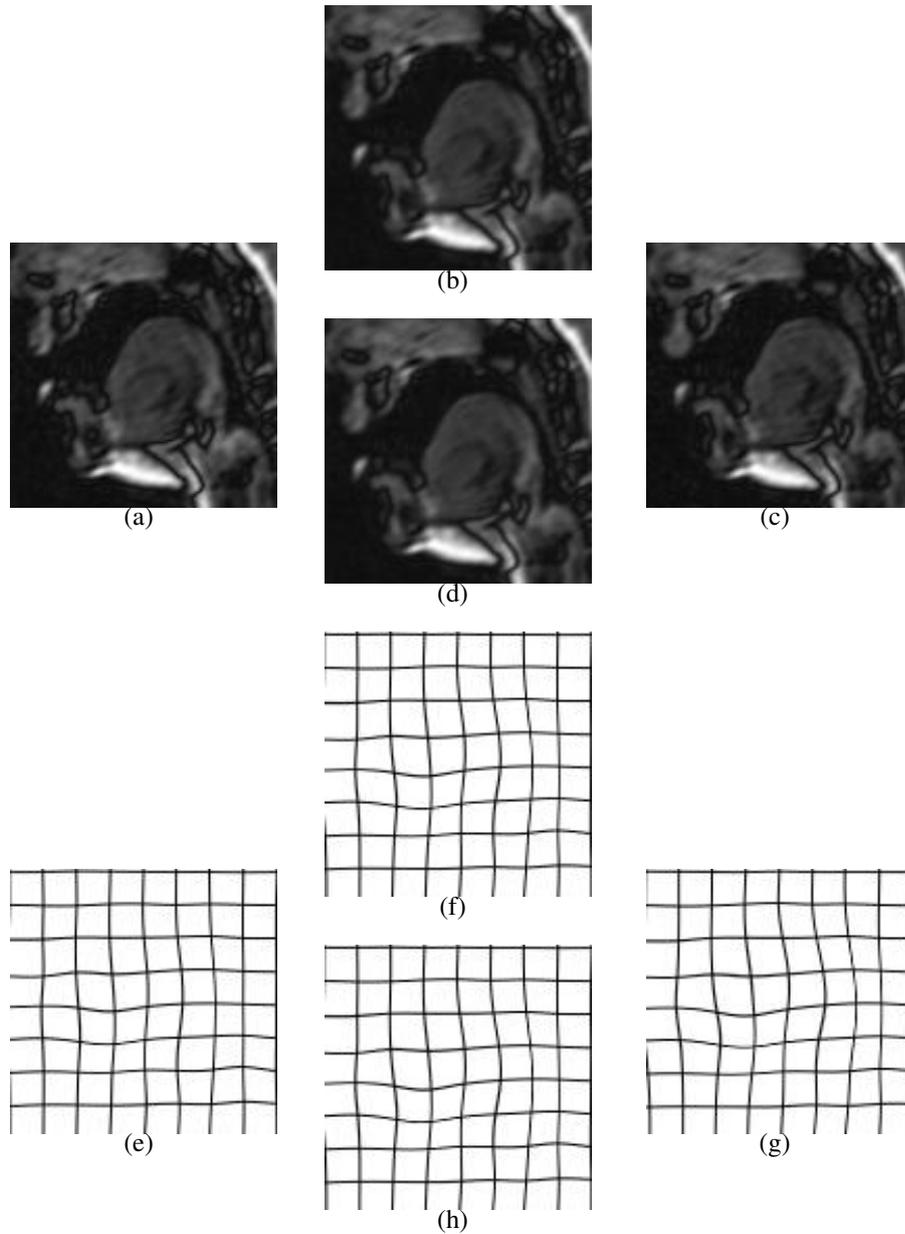


Figura 6.2: (a) e (c) Detalhe das segunda e terceira imagens de uma sequência observada. (e) e (g) Malhas de pontos de controle correspondentes a (a) e (c), respectivamente. (b) e (d) Imagens geradas por interpolação linear e utilizando splines cúbicas na direção do movimento, respectivamente. (f) e (h) Malhas de pontos de controle correspondentes a (b) e (d), respectivamente.

6.3 Aumento de Resolução Espacial

Uma avaliação numérica da abordagem MAP-MRF baseada no algoritmo ICM, utilizando o modelo *a priori* GIMLL, foi conduzida processando uma sequência de sete imagens de baixa

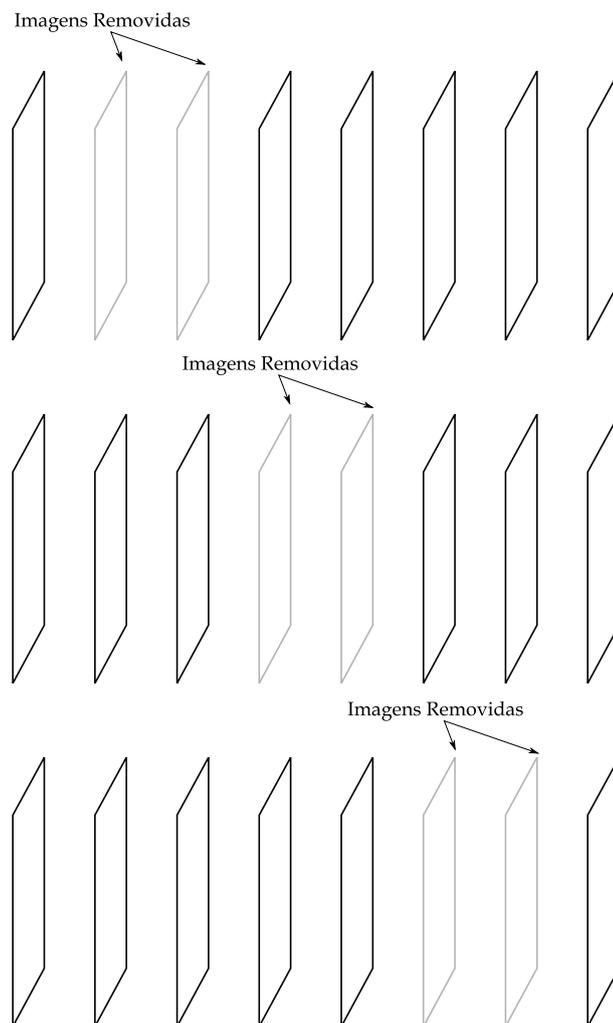


Figura 6.3: Aumento de Resolução Temporal - Experimento 2 - Remoção das imagens duas a duas consecutivamente.

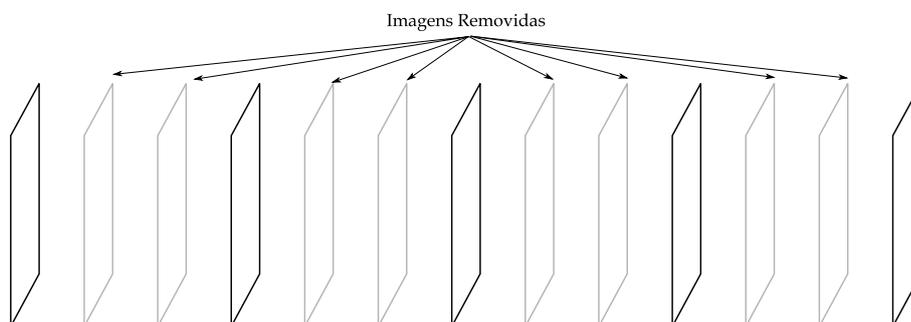


Figura 6.4: Experimento 3 - Sequência sub-amostrada temporalmente.

resolução simuladas. Transformações identificadas em uma sequência real foram utilizadas para simular o movimento dos articuladores da fala. Essas transformações foram aplicadas a uma imagem observada e a sequência simulada foi sub-amostrada considerado o fator de escala 2,

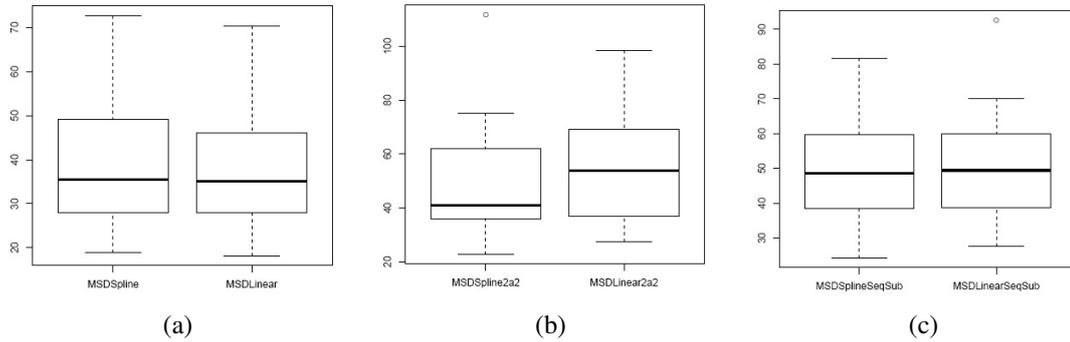


Figura 6.5: Boxplot dos dados obtidos nos experimentos (a) 1, (b) 2 e (c) 3.

gerando uma sequência de imagens de baixa resolução simulada. Esse processo é ilustrado na Figura 6.6. É importante notar que, nesse experimento, os deslocamentos de ordem sub-pixel são completamente conhecidos (o registro é perfeito). Além disso, como discutido anteriormente, Martins *et al.* (2009b) discute estimação do parâmetro β feita seguindo um procedimento similar ao proposto por Levada e Tannús (2008). Entretanto, nos experimentos desenvolvidos neste capítulo, esse parâmetro foi decidido empiricamente de forma a alcançar os melhores resultados (em todos os casos $\beta = 0.4$).

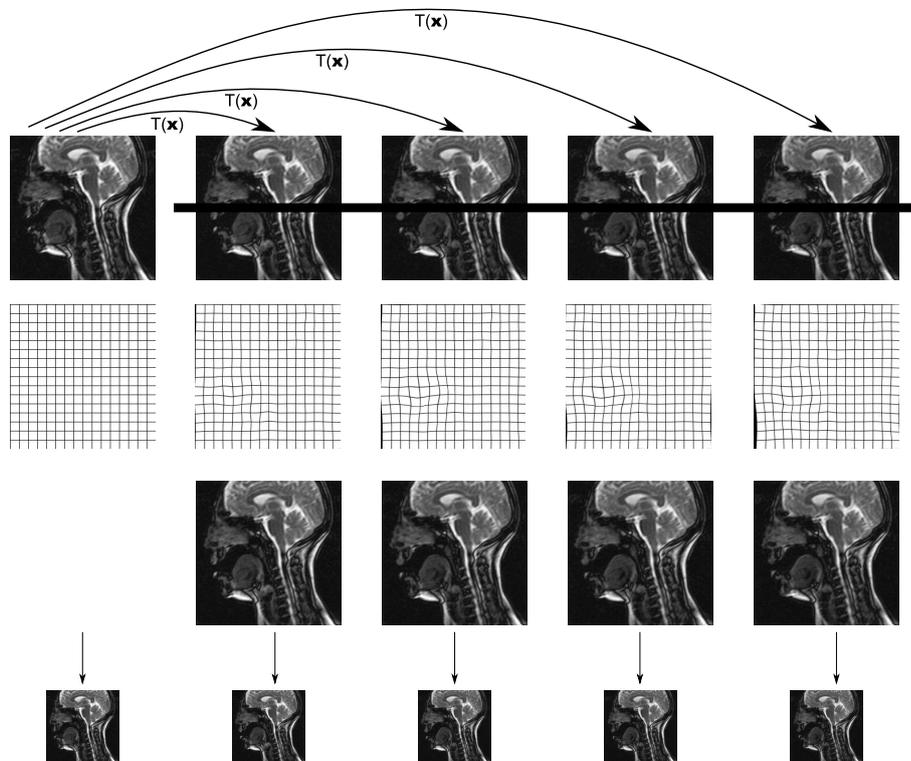


Figura 6.6: Ilustração do processo de geração das imagens de baixa resolução simuladas.

Aparentemente, existem apenas dois outros métodos de reconstrução por super-resolução baseados no algoritmo ICM na literatura (Martins *et al.*, 2007) (Suresh e Rajagopalan, 2007). Em Martins *et al.* (2007), o modelo de MRF MLL isotrópico (também conhecido como modelo de Potts) é utilizado para definir o conjunto de todas as distribuições condicionais locais $p(f_i | f_{\eta_i})$, $i = 1, \dots, M^2$. Entretanto, nesse modelo se dois vizinhos não possuem exatamente o mesmo nível de cinza, eles não contribuem nada para a distribuição, mesmo se forem níveis de cinza próximos. A fim de não suavizar descontinuidades, Suresh e Rajagopalan (2007) utilizam um modelo de MRF adaptativo a descontinuidades (DAMRF - *Discontinuity Adaptive MRF*) com o qual o grau de interação entre vizinhos é ajustado sempre que uma borda é encontrada. A interação entre vizinhos diminui nas descontinuidades.

A interpolação bilinear da imagem de referência e os métodos propostos em Martins *et al.* (2007) e Suresh e Rajagopalan (2007) foram utilizados para comparar os resultados do modelo GIMLL. Li (2009) discute vários outros modelos que podem ser adotados nesse contexto. O modelo *Total Variation*, que utiliza norma L1 como outra solução para a suavização adaptativa a descontinuidade, também foi utilizado na comparação. A Figura 6.7 ilustra a forma qualitativa das funções potenciais utilizadas na comparação. Os modelos GIMLL e DAMRF se baseiam na diferença quadrada entre vizinhos (norma L2). O modelo TV se baseia no erro absoluto entre vizinhos (módulo do gradiente). Finalmente, no modelo de Potts, independente da diferença entre vizinhos, se essa diferença não for zero, ela será penalizada (norma L0).

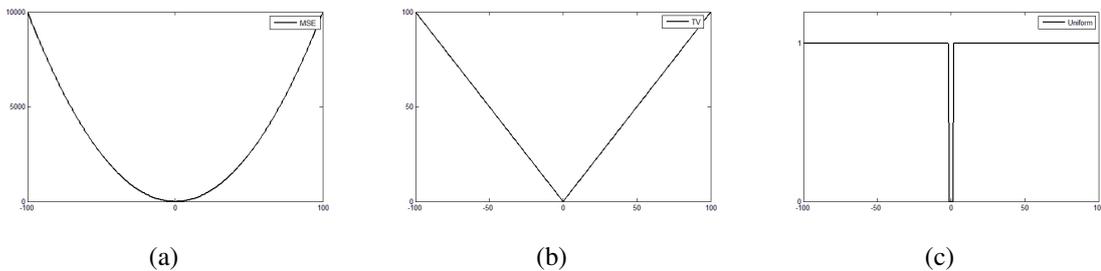


Figura 6.7: Forma qualitativa das funções potenciais utilizadas na comparação. (a) Funções baseadas na diferença quadrática (GIMLL e DAMRF). (b) Modelo TV. (c) Modelo de Potts.

O NMSE foi utilizado na avaliação dessas abordagens. A Tabela 6.2 mostra os resultados. É importante notar que o modelo GIMLL sempre apresenta os melhores resultados. A fim de verificar se as diferenças entre as médias foram estatisticamente significativas, o teste de análise de variância (ANOVA) foi utilizado. Como foi verificado (através do teste de Shapiro) que as distribuições não são Gaussianas, o método ANOVA de Kruskal-Wallis não paramétrico foi utilizado. Com 95% de confiança, o teste indicou que existe diferença estatística entre as médias ($\chi^2 = 31.9646$, $df = 4$, $p\text{-value} = 0.000001945$). Além disso, o teste de Wilcoxon pareado

mostrou diferença significativa entre as duas melhores abordagens (p -value = 0.007813). Esses resultados são uma evidência da efetividade do modelo GIMLL para a estimação das imagens de alta resolução nesse contexto. A Figura 6.8 mostra os resultados correspondentes a uma das imagens utilizadas nesse experimento. Essas imagens condizem com os resultados mostrados na Tabela 6.2.

Tabela 6.2: NMSE das 7 imagens simuladas, reconstruídas utilizando o modelo GIMLL em comparação com a interpolação bilinear das imagens de baixa resolução simuladas e as imagens reconstruídas utilizando os outros modelos.

	Img01	Img02	Img03	Img04	Img05	Img06	Img07	Média
GIMLL	0.000633	0.000644	0.000625	0.000594	0.000567	0.000680	0.000749	0.000642
DAMRF	0.000861	0.000729	0.000696	0.000662	0.000673	0.000812	0.000885	0.000760
TV	0.002263	0.002212	0.002156	0.002022	0.002068	0.002214	0.002249	0.002169
Potts	0.019171	0.018052	0.018914	0.019357	0.018613	0.018795	0.018296	0.018743
Bilinear	0.008334	0.007664	0.007550	0.007446	0.007434	0.007148	0.006877	0.007493

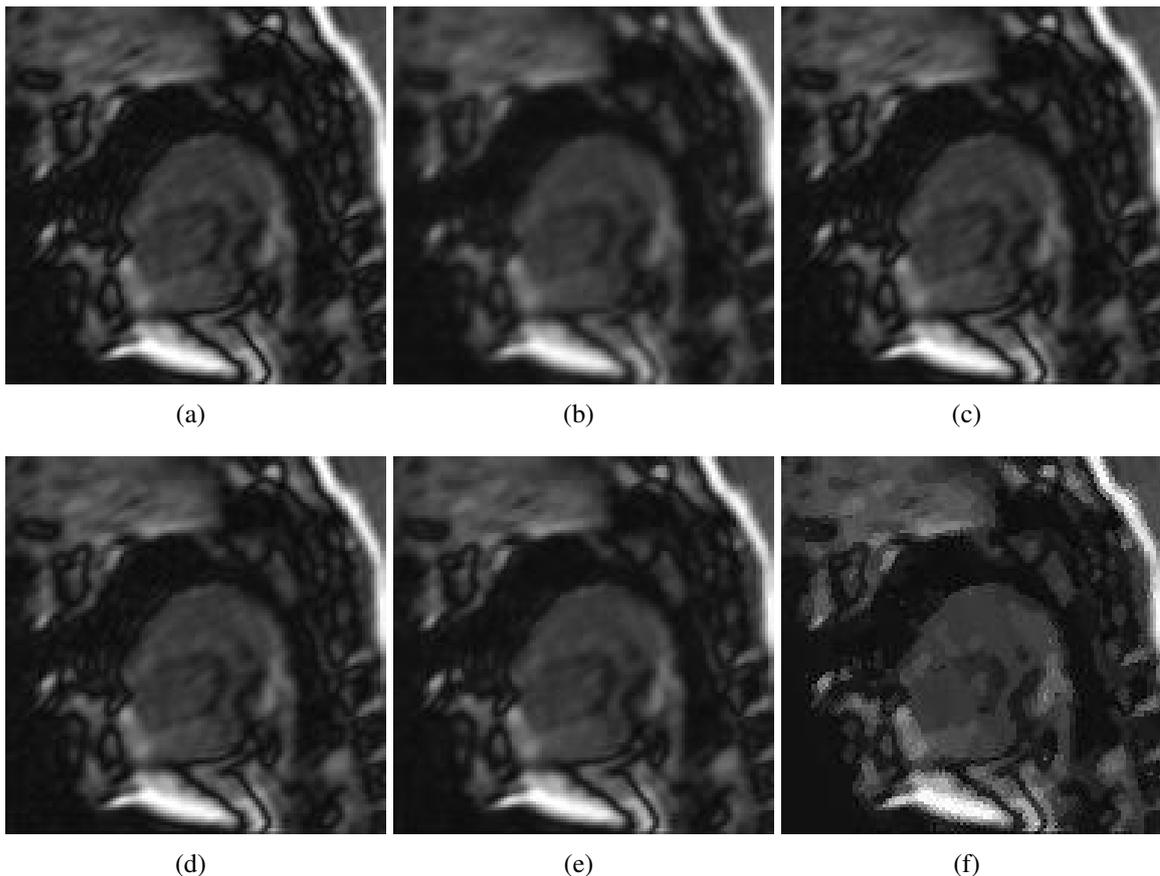


Figura 6.8: (a) Detalhe da imagem original utilizada nesse experimento. (b) Detalhe da interpolação bilinear da imagem de referência. Imagens reconstruídas pelos modelos (c) GIMLL; (d) DAMRF; (e) TV; e (f) Potts.

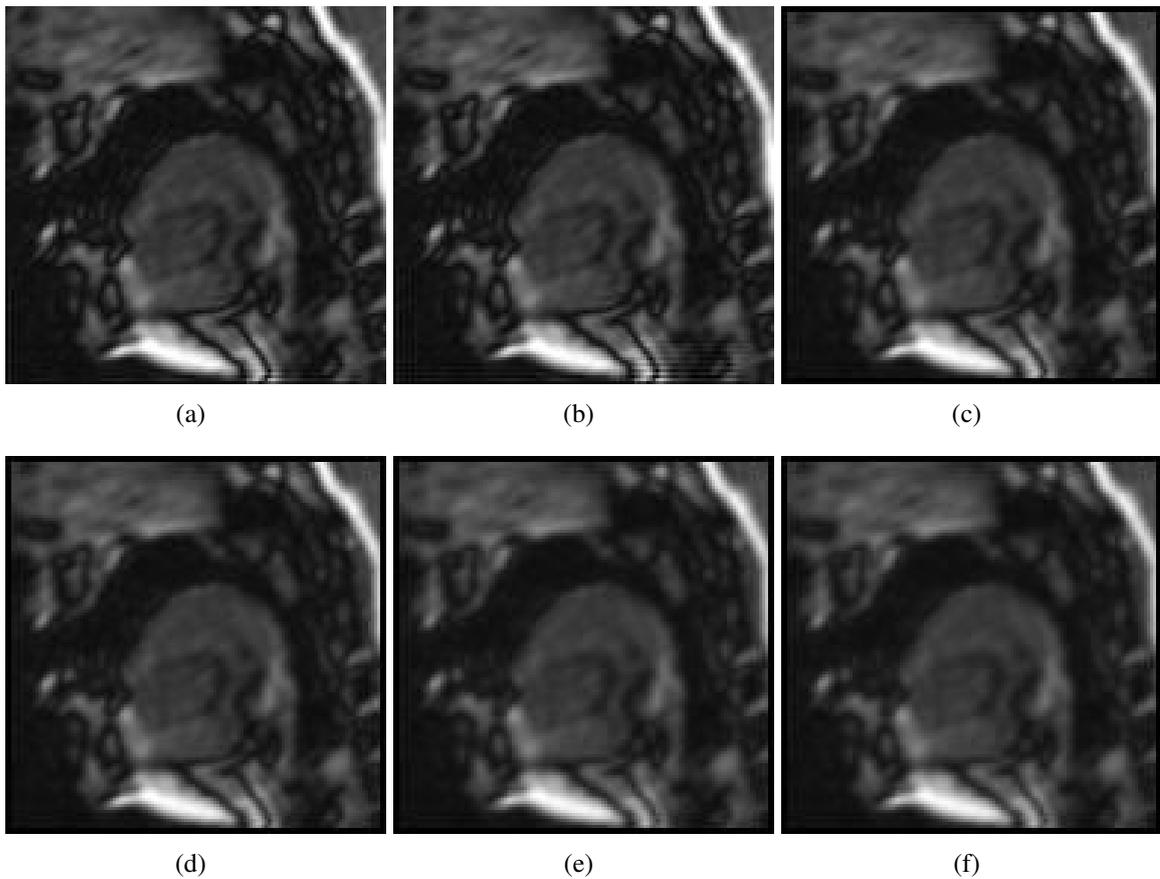
Como comentado anteriormente, apesar de ter apresentado resultados promissores no contexto das imagens do trato vocal, devido à dimensão do problema, o algoritmo ICM apresentou alto custo computacional. Os resultados dessa abordagem foram comparados com as propostas de aumento de resolução espacial baseadas no filtro de Wiener discutidas na Seção 5.3). Como comentado na Seção 5.3.2, as três propostas podem adotar o modelo isotrópico ou Markoviano separável para caracterizar as estruturas de correlação espacial. A Tabela 6.3 mostra o NMSE da abordagem inicial e das três novas propostas considerando os dois modelos, isotrópico e Markoviano, considerando o mesmo conjunto de 7 imagens simuladas utilizado na comparação acima. Na tabela, WI indica o filtro de Wiener adaptativo utilizando o modelo isotrópico, WM o filtro de Wiener adaptativo utilizando o modelo Markoviano separável, IEI a interpolação estatística utilizando o modelo isotrópico, IEM a interpolação estatística utilizando o modelo Markoviano separável, MTI a abordagem multitemporal utilizando o modelo isotrópico e MTM a abordagem multitemporal utilizando o modelo Markoviano separável. Em todos os casos o parâmetro ρ foi decidido empiricamente de forma a alcançar os melhores resultados (para o Filtro de Wiener Adaptativo $\rho = 0.75$, Interpolação Estatística $\rho = 0.95$ e Abordagem Multitemporal $\rho = 0.95$). A Figura 6.9 mostra os resultados correspondentes a uma das imagens utilizadas nesse experimento. Essas imagens condizem com os resultados mostrados na Tabela 6.3. Como é possível notar, em todos os casos o filtro de Wiener adaptativo apresentou os melhores resultados. A abordagem multitemporal, a qual não considera os deslocamentos presentes entre as imagens na estimação, apresentou os piores resultados, sendo pior do que a interpolação bilinear da imagem de referência em alguns casos. Esses resultados ficam evidentes no *boxplot* das três abordagens mostrado na Figura 6.10. A fim de verificar se as diferenças entre as médias foram estatisticamente significativas, o método ANOVA de Kruskal-Wallis não paramétrico foi utilizado. Com 95% de confiança, o teste indicou que existe diferença estatística entre as médias ($\chi^2 = 46.9057$, $df = 6$, $p\text{-value} = 0.00000001954$). Além disso, o teste de Wilcoxon pareado mostrou diferença significativa entre a abordagem baseada no modelo GIMLL e o filtro de Wiener adaptativo utilizando o modelo isotrópico ($p\text{-value} = 0.007813$).

Para cada uma das três propostas baseadas no filtro de Wiener, o modelo isotrópico apresentou desempenho superior quando comparado ao modelo Markoviano separável. Isso pode ser verificado visualmente através do *boxplot* de cada abordagem mostrados na Figura 6.11. Em todos os casos, com 95% de confiança, essa diferença foi significativa ($p\text{-value} = 0.007813$).

A Tabela 6.4 mostra os tempos de execução de cada método em segundos. Os algoritmos foram executados utilizando a versão 7.8.0.347 do software Matlab instalado no sistema operacional Windows 7, em um computador com processador Intel(R) Core(TM) i3, 2.13GHz e memória RAM de 4,00 GB. Como discutido anteriormente, a abordagem baseada no modelo GIMLL possui alto custo computacional, apresentando os piores tempos. Cada iteração do al-

Tabela 6.3: NMSE das 7 imagens simuladas, reconstruídas utilizando o modelo GIMLL em comparação com as propostas baseados no filtro de Wiener.

	Img01	Img02	Img03	Img04	Img05	Img06	Img07	Média
GIMLL	0.000633	0.000644	0.000625	0.000594	0.000567	0.000680	0.000749	0.000642
WI	0.000298	0.000226	0.000219	0.000200	0.000179	0.000231	0.000273	0.000233
WM	0.000503	0.000389	0.000370	0.000308	0.000290	0.000315	0.000372	0.000364
IEI	0.002427	0.002267	0.002213	0.002128	0.002058	0.002086	0.002106	0.002184
IEM	0.003350	0.003117	0.002997	0.002943	0.002927	0.002942	0.002867	0.003021
MTI	0.007008	0.006714	0.006618	0.006497	0.006472	0.006211	0.005949	0.006495
MTM	0.008057	0.007696	0.007581	0.007478	0.007466	0.007182	0.006912	0.007482

**Figura 6.9:** (a) Detalhe das imagens reconstruídas utilizando as abordagens (a) WI, (b) WM, (c) IEI, (d) IEM, (e) MTI, (f) MTM.

goritmo levou 300.389 segundos para ser completada. Dessa forma, apesar do filtro de Wiener adaptativo apresentar o segundo maior tempo de execução, ele é mais rápido do que apenas uma iteração da abordagem baseada no modelo GIMLL. A abordagem multitemporal possui os menores tempos, entretanto ela apresentou os piores resultados. Considerando a relação entre o NMSE das imagens geradas e o tempo de execução, a interpolação estatística se mostra a opção

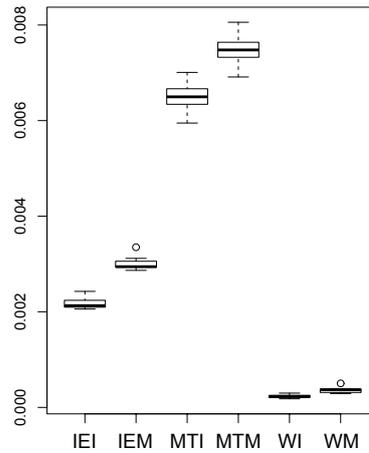


Figura 6.10: *Boxplot* das três propostas baseadas no filtro de Wiener.

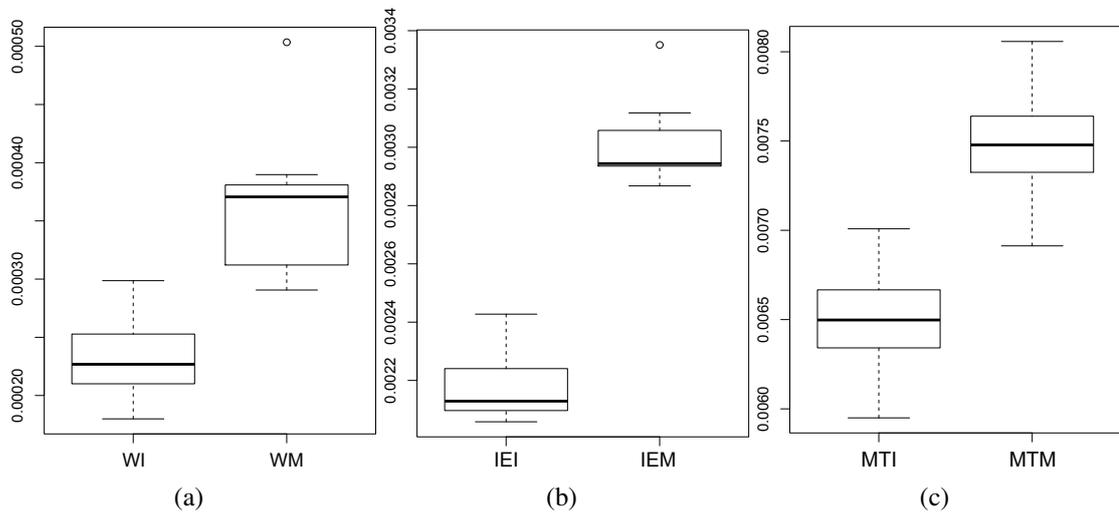


Figura 6.11: *Boxplot* das três propostas baseadas no filtro de Wiener considerando os dois modelos utilizados para caracterizar as estruturas de correlação espacial: (a) filtro de Wiener adaptativo, (b) interpolação estatística e (c) abordagem multitemporal.

mais interessante. Apesar de não ter apresentado os melhores resultados, o tempo de execução é extremamente baixo.

Tabela 6.4: Tempo de Execução da reconstrução das 7 imagens simuladas, utilizando o modelo GIMLL em comparação com as propostas baseados no filtro de Wiener.

	GIMLL	WI	WM	IEI	IEM	MTI	MTM
Tempo de Execução (segundos)	901.168	200.723	200.822	4.682	3.264	2.062	1.864

A Tabela 6.5 mostra o NMSE das imagens reconstruídas utilizando o modelo GIMLL e as abordagens baseadas no filtro de Wiener, considerando de sete a apenas uma imagem de baixa resolução. De acordo com a tabela é possível notar que todos os métodos, com exceção da

abordagem multitemporal, apresentam melhores resultados à medida em que cresce o número de observações. A abordagem multitemporal não apresenta grandes variações porque as informações adicionais que cada imagem oferece devido à presença de deslocamentos de ordem subpixel, não são consideradas na estimação. É importante notar que, em todos os casos, mesmo quando apenas uma imagem é utilizada na estimação, todas as abordagens apresentaram erro menor do que a interpolação bilinear da imagem de referência.

Tabela 6.5: NMSE de imagens reconstruídas utilizando o modelo GIMLL e as abordagens baseadas no filtro de Wiener, considerando de 7 a apenas uma observação de baixa resolução.

Número de observações	7	6	5	4	3	2	1
GIMLL	0.000633	0.000858	0.000939	0.001169	0.001374	0.002211	0.004229
WI	0.000298	0.000381	0.000443	0.000577	0.000681	0.001209	0.002251
WM	0.000503	0.000553	0.000558	0.000733	0.000846	0.001438	0.002624
IEI	0.002427	0.002566	0.002709	0.003036	0.003291	0.004279	0.006569
IEM	0.003350	0.003472	0.003594	0.003939	0.004209	0.005501	0.008061
MTI	0.007008	0.007010	0.007013	0.007014	0.007017	0.007023	0.007033
MTM	0.008057	0.008056	0.008055	0.008055	0.008054	0.008051	0.008048

Como discutido anteriormente, nas abordagens de SRIR, quando o registro não é perfeito, podem surgir artefatos nas regiões em que o erro do registro é grande. Apesar de identificar satisfatoriamente as deformações presentes nas sequências de imagens do trato vocal, o registro não rígido adotado não é perfeito. A Figura 6.12 mostra as imagens reconstruídas considerando que os deslocamentos de ordem subpixel não são previamente conhecidos. É possível identificar tais artefatos nas imagens, principalmente na região que engloba a epiglote e a base da língua (canto inferior direito das imagens). Como discutido no final da Seção 5.3.2, na definição da abordagem do filtro de Wiener adaptativa a variância do ruído pode ser considerada proporcional ao erro absoluto do registro. Como pode ser visto nas Figuras 6.12(h) e 6.12(i), esse procedimento atenua consideravelmente os artefatos. A Tabela 6.6 mostra o NMSE de uma das imagens reconstruídas por cada método considerando que o registro não é perfeito. Como é possível notar, a presença dos artefatos nas regiões em que o erro do registro é grande aumenta consideravelmente o NMSE. O procedimento adotado para atenuar esses artefatos melhora muito a qualidade das imagens.

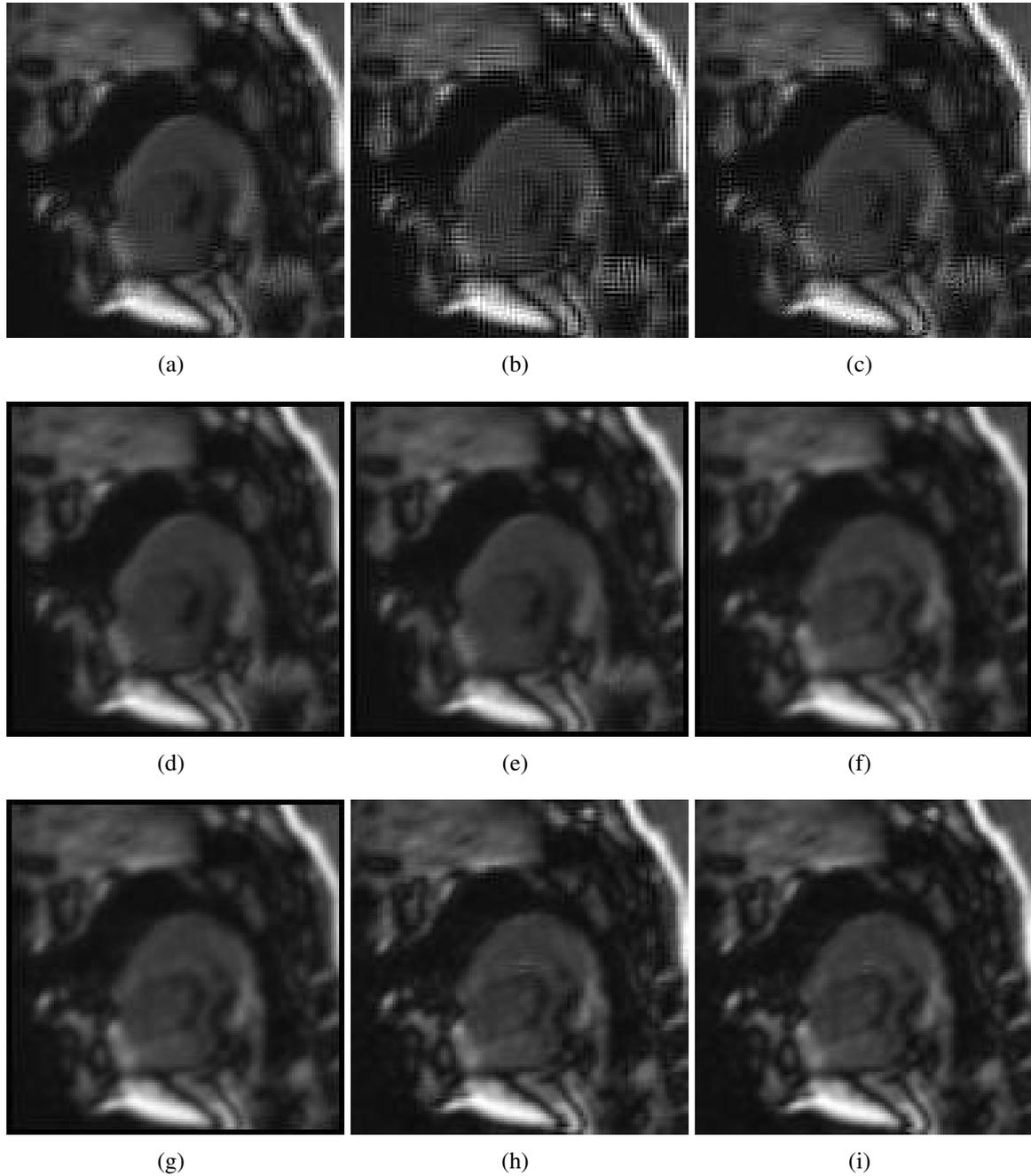


Figura 6.12: Detalhe das imagens reconstruídas considerando que o registro não é perfeito: (a) GIMLL, (b) WM, (c) WI, (d) IEM, (e) IEI, (f) MTM, (g) MTI, e o filtro de Wiener adaptativo considerando a variância do ruído proporcional ao erro do registro (h) WM e (i) WI.

Tabela 6.6: NMSE de imagens reconstruídas considerando que o registro não é perfeito.

	NMSE
GIMLL	0.043554
WI	0.058709
WM	0.084675
IEI	0.037692
IEM	0.039162
MTI	0.011797
MTM	0.013006
WI considerando o erro do registro	0.005259
WM considerando o erro do registro	0.006630

Conclusões e Trabalhos Futuros

7.1 Considerações Finais

Neste trabalho foi apresentada uma abordagem de aumento de resolução espaço-temporal de sequência de imagens de ressonância magnética do trato vocal, adquiridas durante a emissão da fala de palavras ou fonemas. A abordagem é formada por duas etapas: o aumento de resolução temporal por meio de uma técnica de interpolação por compensação de movimento; e o aumento de resolução espacial por meio de uma técnica de reconstrução de imagens por super-resolução. Primeiramente as deformações presentes entre as imagens observadas são identificadas por meio de um método de registro não rígido proposto na literatura (Rueckert *et al.*, 1999). Esse método se baseia em FFDs utilizando funções B-spline como funções base. As coordenadas dos pontos de controle que formam uma malha são os únicos parâmetros da transformação, e sua identificação é independente do conhecimento de especialistas ou de características da imagem, o que faz com que essa abordagem seja bastante vantajosa. Além disso, como as funções base utilizadas possuem suporte finito, a mudança na posição de um ponto de controle afeta apenas sua vizinhança. Isso permite a modelagem de deformações localizadas em pequenas partes da imagem. Como discutido anteriormente, esse método de registro foi capaz de modelar adequadamente as deformações existentes entre as imagens do trato vocal.

A fim de aumentar a resolução temporal de uma sequência de imagens de um dado evento, abordagens existentes utilizam várias aquisições do mesmo evento, com um pequeno atraso

entre elas. Entretanto, quando se possui apenas uma única aquisição desse evento, o aumento de resolução temporal é alcançado por meio de interpolação entre as observações. A fim de evitar a presença de artefatos ou borramento nas imagens interpoladas, o movimento presente na sequência de imagens observadas deve ser considerado nessa interpolação. Considerando as malhas de pontos de controle geradas pelo método de registro, a proposta para o aumento de resolução temporal gera malhas de pontos de controle intermediárias às malhas identificadas por meio de interpolação linear na direção do movimento de pontos correspondentes. A imagem intermediária é gerada pela média ponderada das imagens adjacentes transformadas de acordo com a malha interpolada. As imagens geradas são coerentes com o movimento existente na sequência. A interpolação por meio de splines cúbicas considerando as quatro malhas mais próximas foi considerada como forma de gerar malhas intermediárias mais coerentes com o movimento presente em toda a sequência. Porém, de acordo com os experimentos desenvolvidos, nesse contexto, apesar da interpolação por splines cúbicas ter apresentado melhores resultados, não existe diferença significativa entre esses dois métodos de interpolação. Acredita-se que isso pode ocorrer devido ao movimento que existe entre as imagens do trato vocal. A maior parte das deformações se deve ao movimento da mandíbula e trata-se de um movimento aproximadamente linear.

Cada imagem de alta resolução espacial é gerada considerando subconjuntos das imagens observadas por meio de uma abordagem de janela deslizante (Figura 2.5). O objetivo inicial para o aumento de resolução espacial das imagens observadas foi a extensão da abordagem proposta pela aluna em seu projeto de mestrado. Adotando um framework Bayesiano, as imagens de alta resolução foram modeladas utilizando campos aleatórios de Markov. Em uma abordagem MAP-MRF, a informação *a priori* é expressa pela probabilidade *a priori* da imagem de alta resolução, a qual é determinada unicamente pelas probabilidades condicionais locais do MRF. Como a otimização global é difícil de ser calculada com exatidão o algoritmo ICM é utilizado para minimizar as probabilidades condicionais locais sequencialmente. Como constatado nos experimentos desenvolvidos, o modelo de MRF GIMLL apresentou resultados superiores quando comparado a outros modelos, inclusive ao modelo de Potts utilizado pela aluna em seu projeto de mestrado. Entretanto, apesar de ter apresentado resultados promissores, devido à dimensão do problema tratado, o algoritmo ICM apresentou alto custo computacional.

Considerando as limitações de performance do algoritmo ICM, iniciou-se a investigação de uma abordagem não iterativa que fornecesse uma solução de mínimo erro médio quadrático. A matriz que caracteriza o filtro de Wiener é escolhida de forma que o erro médio quadrático seja minimizado e trata-se de uma abordagem não iterativa. Dessa forma, decidiu-se adaptar o filtro de Wiener para o problema da reconstrução por super-resolução. Nesse contexto, foram encontrados dois trabalhos na literatura que propõem abordagens para o aumento de resolução

espacial baseadas no filtro de Wiener (Hardie, 2007; Mascarenhas *et al.*, 1996). Mascarenhas *et al.* (1996) discutem um método de interpolação para fusão de dados de satélite baseado no critério de mínimo erro médio quadrático. Tanto para as estruturas de correlação das observações, como para as estruturas de correlação cruzada, o método implementa o aumento de resolução espacial das três bandas multiespectrais assumindo a hipótese de separabilidade nas direções horizontal, vertical e espectral e adotando um modelo Markoviano de primeira ordem para modelá-las. Diferente das imagens utilizadas neste projeto, não existem deslocamentos entre as bandas multiespectrais. Além disso, os autores não consideram a PSF do sistema de imageamento no modelo de formação das imagens.

Hardie (2007) apresenta um algoritmo de reconstrução por super-resolução baseado no filtro de Wiener. De maneira semelhante ao trabalho de Mascarenhas *et al.* (1996), os pixels de alta resolução são obtidos por meio da soma ponderada dos pixels observados de acordo com sua localização espacial. Porém o autor adotou um modelo de auto-correlação paramétrico circularmente simétrico para caracterizar a matriz de correlação *a priori*. Assim, a matriz de correlação das observações e matriz de correlação cruzada são definidas em função da matriz de correlação *a priori* de acordo com o modelo de formação das imagens observadas. Diferente da abordagem proposta por Mascarenhas *et al.* (1996), Hardie (2007) considera a PSF do sistema de imageamento no modelo de formação das imagens. Além disso, diferente das imagens utilizadas neste projeto, o autor considera apenas a presença de translações globais entre as observações de baixa resolução e não considera que o registro entre elas possa ser imperfeito.

Adaptando a solução proposta por (Mascarenhas *et al.*, 1996) para o contexto deste projeto, chegou-se a duas possibilidades. Na primeira abordagem, denominada Interpolação Estatística, os pixels de um subconjunto das imagens observadas (de acordo com a abordagem de janela deslizante ilustrada na Figura 2.5) são dispostos na grade de alta resolução de acordo com o fator de escala e os deslocamentos causados pelas deformações. Assim, não existe informação espectral. Adotando um dos modelos propostos (modelo Markoviano separável utilizado por Mascarenhas *et al.* (1996), ou modelo isotrópico utilizado por Hardie (2007)) para caracterizar as estruturas de correlação das observações e cruzada, os pixels de alta resolução de uma subjanela da janela de observação são identificados pelo estimador Bayesiano. Devido à presença de deformações entre as imagens observadas, a localização dos pixels observados varia de acordo com a posição da janela de observação. Dessa forma as matrizes de correlação das observações e cruzada devem ser recalculadas a cada posição dessa janela.

Na segunda abordagem, denominada multitemporal, semelhante a uma imagem multiespectral, cada imagem de baixa resolução é considerada uma banda multitemporal. Diferente da abordagem anterior, considerando um subconjunto das imagens de baixa resolução, é estimado o mesmo número de imagens de alta resolução. Além disso, os deslocamentos presentes

entre as observações são desconsiderados na estimação. Considerando uma janela de observação que engloba um subconjunto dos pixels de baixa resolução correspondentes de em todas as banda multitemporais, o estimador Bayesiano de Mascarenhas *et al.* (1996) é aplicado sem grandes modificações. Novamente é possível adotar o modelo Markoviano separável ou o modelo isotrópico para caracterizar as estruturas de correlação espacial das observações e cruzada. A matriz de covariância temporal é calculada considerando as observações.

Adaptando a solução proposta por Hardie (2007) para o contexto deste projeto, as estruturas de correlação espaciais *a priori* são caracterizadas por um dos dois modelos discutidos (separável Markoviano ou isotrópico). Essa abordagem foi denominada filtro de Wiener adaptativo. O modelo de formação das imagens mostrado na Equação (5.1), no qual o operador D modela a função do sensor de aquisição das imagem de baixa resolução de acordo com os deslocamentos presentes entre elas, é considerado na estimação. As matrizes de correlação das observações e cruzada são identificadas com relação à matriz de correlação *a priori*, de acordo com o modelo de formação das observações. Semelhante à abordagem denominada interpolação estatística, é adotada uma janela de observação e os pixels de uma subjanela dessa janela são estimados. Além disso, devido à presença de deformações entre as imagens observadas, a localização dos pixels observados varia de acordo com a posição da janela de observação. Dessa forma as matrizes de correlação das observações e cruzada devem ser recalculadas a cada posição dessa janela. Em aplicações de reconstrução por super-resolução, podem surgir artefatos nas regiões em que o erro do registro é grande. No contexto das imagens do trato vocal, geralmente o método de registro não é capaz de capturar todas as deformações presentes entre as imagens. Dessa forma, a fim de atenuar esses artefatos, a variância do ruído é considerada proporcional ao erro absoluto do registro. Dessa forma, as observações que não foram corretamente registradas terão menos peso na estimação. Esse procedimento atenua consideravelmente os artefatos.

De acordo com os experimentos desenvolvidos, para cada uma das três propostas baseadas no filtro de Wiener, o modelo isotrópico apresentou desempenho superior quando comparado ao modelo Markoviano separável. Em todos os casos, considerando todas as propostas baseadas no filtro de Wiener e a proposta inicial baseada no modelo de Markov GIMLL, o filtro de Wiener adaptativo apresentou os melhores resultados. Assim, o filtro de Wiener adaptativo, utilizando o modelo isotrópico para caracterizar a matriz de correlação *a priori*, foi avaliado como o melhor método com relação do NMSE das imagens estimadas. Em seguida estão o filtro de Wiener adaptativo utilizando o modelo Markoviano separável para caracterizar as estruturas de correlação espacial *a priori*, e a abordagem MAP-MRF baseada no modelo GIMLL. A abordagem multitemporal apresentou os piores resultados, sendo pior do que a interpolação bilinear da imagem de referência em alguns casos. Acredita-se que esse desempenho se deve ao fato de que essa abordagem desconsidera os deslocamentos presentes entre as imagens da sequência.

Considerando tais deslocamentos, cada imagem representa informações adicionais a respeito da imagem sendo reconstruída. Ao desconsiderá-los, a abordagem se aproxima de uma simples interpolação.

Com relação aos tempos de execução de cada método, a abordagem baseada no modelo GIMLL possui alto custo computacional, apresentando os piores tempos. Apesar do filtro de Wiener adaptativo apresentar o segundo maior tempo de execução, ele é mais rápido do que apenas uma iteração da abordagem baseada no modelo GIMLL. A abordagem multitemporal possui os menores tempos, entretanto ela apresentou os piores resultados. Considerando a relação entre o NMSE e o tempo de execução, a interpolação estatística se mostra a opção mais interessante. Apesar de não ter apresentado os melhores resultados, o tempo de execução é extremamente baixo.

Apesar de identificar satisfatoriamente as deformações presentes nas sequências de imagens do trato vocal, o registro não rígido adotado não é perfeito. A presença dos artefatos nas regiões em que o erro do registro é grande aumenta consideravelmente o NMSE de todas as abordagens. Para a abordagem denominada filtro de Wiener adaptativo, considerando que variância do ruído é proporcional ao erro absoluto do registro, grande parte dos artefatos são atenuados, o que melhora consideravelmente o NMSE das estimações.

7.2 Contribuições

A principal contribuição desta tese é uma abordagem efetiva para o aumento de resolução espaço-temporal de sequências de MRI do trato vocal baseado apenas em técnicas de processamento de imagens digitais. Um aspecto importante da abordagem é a utilização do método de registro não rígido baseado em FFDs. Esse método tem sido extensamente utilizado para o registro de imagens médicas (Rueckert e Aljabar, 2010), porém na literatura investigada não foi encontrado nenhum trabalho semelhante no contexto de sequências do trato vocal.

No que diz respeito ao aumento de resolução temporal por meio de uma técnica de interpolação por compensação de movimento, a principal contribuição foi a constatação de que, de acordo com os experimentos desenvolvidos, não existe diferença significativa entre os dois métodos investigados. Apesar da interpolação por splines cúbicas considerar mais informações a respeito do movimento presente em toda a sequência, acredita-se que, como o movimento da mandíbula é aproximadamente linear, esse acréscimo de informação não é estritamente necessário.

Com relação ao aumento de resolução espacial das sequências observadas por meio de uma abordagem de SRIR baseada no algoritmo ICM, a contribuição mais importante foi a constata-

tação de que o modelo de Markov GIMLL é o mais adequado dentre os modelos utilizados na comparação. Existem evidências de que isso não ocorre somente no contexto das imagens do trato vocal (Martins *et al.*, 2009a). Ainda com relação ao aumento de resolução espacial, outras contribuições relevantes foram a exploração e comparação de possíveis variações de uma abordagem de SRIR baseada no filtro de Wiener, além da comparação entre os modelos Markoviano separável e isotrópico utilizados para caracterizar as estruturas de correlação espacial. De acordo com os experimentos apresentados, no contexto das imagens utilizadas neste projeto, existem evidências de que o modelo isotrópico é o mais adequado. Além disso, uma característica importante que melhorou consideravelmente a estimação foi a consideração da PSF do sensor de aquisição das imagens juntamente com a imposição de informação *a priori* por meio da matriz de correlação dos pixels de alta resolução a serem estimados. Por fim, considerando que o método de registro adotado pode não ser perfeito, o que ocorre com certa frequência, o procedimento adotado para atenuar os artefatos que são amplificados nas regiões em que o erro do registro é grande também é uma contribuição significativa desta tese.

7.3 Trabalhos Futuros

A abordagem apresentada para o aumento de resolução das sequências do trato vocal é formada por duas etapas completamente disjuntas: o aumento de resolução temporal por meio de uma técnica de interpolação por compensação de movimento, e o aumento de resolução espacial por meio de uma técnica de SRIR. Como trabalho futuro propõe-se que o aumento de resolução espaço-temporal ocorra em passo único, possivelmente utilizando as imagens das várias repetições do mesmo evento.

O valor do parâmetro ρ , que caracteriza as estruturas de correlação espacial de todas as propostas baseadas no filtro de Wiener foi decidido empiricamente e é único para toda a imagem. Entretanto, de acordo com alguns experimentos desenvolvidos, seu valor influencia na qualidade das estimções de alta resolução. Assim, na continuação deste trabalho pretende-se investigar a estimação automática desse parâmetro e a possibilidade de seu valor variar de acordo com as observações presentes na janela de observação.

Recentemente, o imageamento por ressonância magnética tem utilizado técnicas de *compressive sensing* a fim de reduzir o tempo de aquisição das imagens (Lustig *et al.*, 2007, 2008). A esparsividade inerente a essas imagens é explorada para subamostrar significativamente o espaço k . No contexto das imagens do trato vocal, um dos principais desafios ainda é a aquisição rápida e de alta qualidade das sequências de imagens. Iniciativas recentes no uso de *compressive sensing* para a aquisição das imagens do trato vocal apresentam evidência de que se trata de

um nicho de pesquisa interessante e importante a ser explorado (Bresch *et al.*, 2008; Kim *et al.*, 2009a,b). Dessa forma, futuramente pretende-se investigar a respeito de *compressive sensing* e sua aplicação na aquisição das imagens de ressonância magnética do trato vocal.

Referências Bibliográficas

- AKGUL, Y. S.; KAMBHAMETTU, C.; STONE, M. Automatic extraction and tracking of the tongue contours. *IEEE Transactions on Medical Imaging*, v. 18, n. 10, p. 1035–1045, 1999.
- ALAM, M. S.; BOGNAR, J. G.; HARDIE, R. C.; YASUDA, B. J. Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames. *IEEE Transactions on Instrumentation and Measurement*, v. 49, p. 915–923, 2000.
- BADIN, P.; BAILLY, G.; RAYBAUDI, M.; SEGEBARTH, C. A three-dimensional linear articulatory model based on MRI data. In: *5th Internat. Conf. on Spoken Language Processing (ICSLP)*, 1998, p. 417–420.
- BADIN, P.; SERRURIER, A. *Three-dimensional Modeling of Speech Organs: Articulatory Data and Models*. Technical Report 177, Institute of Electronics, Information, and Communication Engineers (IEICE), 2006.
- BAER, T.; GORE, J.; GRACCO, L. C.; NYE, P. W. Analysis of vocal tract shape and dimensions using Magnetic Resonance Imaging: Vowels. *Journal of the Acoustic Society of America (JASA)*, v. 90, n. 2, p. 799–828, 1991.
- BEHRENDTS, J.; WISMULLER, A. A segmentation and analysis method for MRI data of the human vocal tract. In: *Proceedings of the Symposium on Human and Machine Perception in Acoustic and Visual Communication*, 2001.
- BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, v. 36, n. 2, p. 192–236, 1974.
- BESAG, J. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, v. 48, p. 259–302, 1986.
- BLACK, M.; ANANDAN, P. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, v. 63, n. 1, p. 75–104, 1996.
- BLAKE, M.; ZISSERMAN, A. *Visual reconstruction*. Cambridge, MA: The MIT Press, 1987.

- BOOKSTEIN, F. L. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 11, n. 6, p. 567–585, 1989.
- BORMAN, S. *Topics in multiframe superresolution restoration*. Tese de Doutorado, University of Notre Dame, Notre Dame, IN, 2004.
- BORMAN, S.; STEVENSON, R. L. *Spatial resolution enhancement of low-resolution image sequences - A comprehensive review with directions for future research*. Relatório Técnico, University of Notre Dame, 1998.
- 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, p. 469–473.
- BOSE, N. K.; LERTRATTANAPANICH, S.; KOO, J. Advances in superresolution using L-curve. In: *The 2001 IEEE International Symposium on Circuits and Systems (ISCAS 2001)*, IEEE, 2001, p. 433–436.
- BRESCH, E.; KIM, Y.-C.; NAYAK, K.; BYRD, D.; NARAYANAN, S. Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP]. *Signal Processing Magazine, IEEE*, v. 25, n. 3, p. 123–132, 2008.
- BRESCH, E.; NARAYANAN, S. Region Segmentation in the Frequency Domain Applied to Upper Airway Real-Time Magnetic Resonance Images. *IEEE Transactions on Medical Imaging*, v. 28, n. 3, p. 323–338, 2009.
- BROWN, L. A survey of image registration techniques. *ACM Computing Surveys*, v. 24 (4), p. 325–376, 1992.
- CANDEIAS, A. L. B. *Uso da Teoria de Estimaco Bayesiana na Fuso de Dados de Satlites*. Tese de doutorado, Instituto Nacional de Pesquisas Espaciais (INPE), 1992.
- CANDES, E.; ROMBERG, J.; TAO, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, v. 52, n. 2, p. 489–509, 2006.
- CASPI, Y.; IRANI, M. Spatio-Temporal Alignment of Sequences. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, v. 24, n. 11, p. 1409–1424, 2002.
- CHAUDHURI, S. *Super-Resolution Imaging*. Norwell, MA, USA: Kluwer Academic Publishers, 2001.
- CHIBA, T.; KAJIYAMA, M. *The Vowel - Its Nature and Structure*. Tokyo: Kaiseikan, 1941.
- COLLIGNON, A.; MAES, F.; DELAERE, D.; VANDERMEULEN, D.; SUETENS, P.; MARCHAL, G. Automated multi-modality image registration based on information theory. *Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI)*, p. 263–274, 1995.

- DANG, J.; HONDA, K.; SUZUKI, H. MRI measurements and acoustic investigation of the nasal and paranasal cavities. *Journal of the Acoustical Society of America*, v. 94, n. 3, p. 1765, 1993.
- DEMOLIN, D.; METENS, T.; SOQUET, A. Three-dimensional measurements of the vocal tract by MRI. In: *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA, USA: University of Delaware and Alfred I. du Pont Institute, 1996, p. 272–275.
- DEMOLIN, D.; METENS, T.; SOQUET, A. Real-time MRI and articulatory coordinations in vowels. In: *Proceedings of the 5th Seminar on Speech Production*, Kloster Seeon, Germany, 2000, p. 86–93.
- DONOHO, D. L. Compressed Sensing. *IEEE Transactions on Information Theory*, v. 52, n. 4, p. 1289–1306, 2006.
- ELAD, M.; FEUER, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured image. *IEEE Trans. Image Process.*, v. 6, p. 1646–1658, 1997.
- 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, 1999.
- 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, p. 1446–1451, 1997.
- FANT, G. *Acoustic theory of speech production*. Mouton, 1960.
- FARSIU, S.; ROBINSON, D.; ELAD, M.; MILANFAR, P. Robust shift and add approach to super-resolution. In *Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing*, p. 121–130, 2003.
- FARSIU, S.; ROBINSON, D.; ELAD, M.; MILANFAR, P. Fast and robust multiframe super-resolution. *IEEE Trans. Image Process.*, p. 1327–1344, 2004.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, n. 6, p. 721–741, 1984.
- GRIMSON, W. E. L. A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, v. 298, n. 1092, p. 395–427, 1982.
- HAMMERSLEY, J. M.; CLIFFORD, P. Markov field on finite graphs and lattices, unpublished, 1971.
- HARDIE, R. A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter. *IEEE Transactions on Image Processing*, v. 16, n. 12, p. 2953–2964, 2007.

- HARDIE, R. C. Super-Resolution using adaptive Wiener filters. In: MILANFAR, P., ed. *Super-Resolution Imaging*, cap. 2, CRC Press, p. 35–61, 2010.
- HEINZ, J. M.; STEVENS, K. N. On the derivation of area functions and acoustic spectra from cineradiographic films of speech. *The Journal of the Acoustical Society of America*, v. 36, n. 5, p. 1037–1038, 1964.
- HILL, D. L. G.; BATCHELOR, P. Registration Methodology: Concepts and Algorithms. In: HAJNAL, J. V.; HILL, D. L.; HAWKES, D. J., eds. *Medical Image Registration*, CRC Press, 2001.
- HONG, M. C.; KANG, M. G.; KATSAGGELOS, A. K. A regularized multichannel restoration approach for globally optimal high resolution video sequence. In: *Visual Communications and Image Processing (VCIP'97)*, San Jose, Calif, USA: Proceedings of SPIE, 1997, p. 1306–1316.
- IRANI, M.; PELEG, S. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, v. 53, n. 3, p. 231–239, 1991.
- IRANI, M.; PELEG, S. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation (JVCIR)*, v. 4, p. 324–335, 1993.
- ISKAROUS, K. Patterns of tongue movement. *Journal of Phonetics*, v. 33, n. 4, p. 363–381, 2005.
- JALLON, J. F.; BERTHOMMIER, F. A semi-automatic method for extracting vocal tract movements from X-ray films. *Speech Communication*, v. 51, n. 2, p. 97–115, 2009.
- JOSHI, M.; CHAUDHURI, S. Simultaneous estimation of super-resolved depth map and intensity field using photometric cue. *Computer Vision and Image Understanding*, v. 101, n. 1, p. 31–44, 2006.
- JOSHI, M.; JALOBEANU, A. MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model. *IEEE Transactions on Geoscience and Remote Sensing*, v. 48, n. 3, p. 1245–1255, 2010.
- KATARTZIS, A.; PETROU, M. Current trends in super-resolution image reconstruction. In: STATHAKI, T., ed. *Image Fusion: Algorithms and Applications*, Academic Press, 2008.
- KATSAGGELOS, A. K.; MOLINA, R.; MATEOS, J. *Super Resolution of Images and Video*. Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan and Claypool Publishers, 2007.
- KIM, S. P.; BOSE, N. K.; VALENZUELA, H. M. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 18, p. 1013–1027, 1990.

- KIM, S. P.; SU, W. Y. Recursive high resolution reconstruction of blurred multiframe images. *IEEE Transactions on Image Processing*, v. 2, p. 534–539, 1993.
- KIM, Y. C.; NARAYANAN, S.; NAYAK, K. S. Accelerated 3D upper airway MRI using compressed sensing. *Magnetic Resonance in Medicine*, v. 61, 2009a.
- KIM, Y. C.; NARAYANAN, S.; NAYAK, K. S. Rapid three-dimensional magnetic resonance imaging of vocal tract shaping using compressed sensing. *The Journal of Acoustical Society of America*, v. 125, n. 4, p. 2498–2498, 2009b.
- KITAMURA, T.; TAKEMOTO, H.; HONDA, K.; SHIMADA, Y.; FUJIMOTO, I.; SYAKUDO, Y.; MASAKI, S.; KURODA, K.; OKU-UCHI, N.; SENDA, M. Difference in vocal tract shape between upright and supine postures: observations by an open-type MRI. *Acoustical Science and Technology*, v. 26, n. 5, p. 465–468, 2005.
- KROON, D. Non-rigid b-spline grid image registration. Disponível em: <http://www.mathworks.com/matlabcentral/fileexchange/20057-non-rigid-b-spline-grid-image-registration>, (acessado em 07/06/2010), 2010.
- LANDWEBER, L. An iterative formula for Fredholm integral equations of the first kind. *American Journal of Mathematics*, v. 73, p. 615–624, 1951.
- LEE, S.; WOLBERG, G.; CHWA, K.-Y.; SHIN, S. Y. Image Metamorphosis with Scattered Feature Constraints. *IEEE Transactions on Visualization and Computer Graphics*, v. 2, p. 337–354, 1996.
- LEVADA, A. L. M. MASCARENHAS, N. D. A.; TANNÚS, A. Pseudolikelihood Equations for Potts MRF model Parameter Estimation on Higher-Order Neighborhood Systems. *IEEE Geoscience and Remote Sensing Letters*, v. 5, n. 3, p. 522–526, 2008.
- LEVIN, C. L.; HOFFMAN, E. J. Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution. *Physics in Medicine and Biology*, v. 44, p. 781–799, 1999.
- LI, S. Z. *Markov random field modeling in image analysis*. Springer Publishing Company, Incorporated, 2009.
- LI, S. Z.; HUANG, Y. H.; FU, J. S. Convex MRF potential functiona. In: *Proceedings of the 1995 International Conference on Image Processing (ICIP 1995)*, Washington, DC, USA: IEEE Computer Society, 1995, p. 269–299.
- LUSTIG, M.; DONOHO, D.; PAULY, J. M. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, v. 58, n. 6, p. 1182–1195, 2007.
- LUSTIG, M.; DONOHO, D. L.; SANTOS, J. M.; PAULY, J. M. Compressed sensing mri. *IEEE Signal Processing Magazine*, v. 25, n. 2, p. 72–82, 2008.

- MARTINS, A. L. D.; HOMEM, M. R. P.; MASCARENHAS, N. D. A. Super-Resolution Image Reconstruction using the Generalized Isotropic Multi-Level Logistic Model. In: *Proceedings of the ACM Symposium on Applied Computing*, Honolulu, Hawaii, USA: ACM, 2009a, p. 934–938.
- MARTINS, A. L. D.; LEVADA, A. L. M.; HOMEM, M. R. P.; MASCARENHAS, N. D. A. Super-resolution image reconstruction using the ICM algorithm. In: *Proceeding of the IEEE International Conference on Image Processing*, San Antonio, Texas, USA: IEEE Computer Society, 2007, p. IV 205–IV 208.
- MARTINS, A. L. D.; LEVADA, A. L. M.; HOMEM, M. R. P.; MASCARENHAS, N. D. A. MAP-MRF Super-Resolution Image Reconstruction using maximum pseudo-likelihood parameter estimation. In: *Proceedings of the IEEE International Conference on Image Processing*, Cairo, Egypt: IEEE Computer Society, 2009b, p. 1165–1168.
- MARTINS, A. L. D.; MASCARENHAS, N. D. A.; SUAZO, C. A. T. Spatio-Temporal Resolution Enhancement of Vocal Tract MRI Sequences Based on Image Registration. *Integrated Computer-Aided Engineering*, v. 18, n. 3, p. 143–155, 2011.
- MARTINS, P.; CARBONE, I.; PINTO, A.; SILVA, A.; TEIXEIRA, A. European Portuguese MRI based speech production studies. *Speech Communication*, v. 50, n. 11-12, p. 925–952, 2008.
- MASCARENHAS, N. D. A.; BANON, G. J. F.; CANDEIAS, A. L. B. Multispectral image data fusion under a bayesian approach. *International Journal of Remote Sensing*, v. 17, n. 8, p. 1457–1471, 1996.
- 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, p. 1919–1922.
- MATHIAK, K.; KLOSE, U.; ACKERMANN, H.; HERTRICH, I.; KINCSES, W.; GRODD, W. Stroboscopic articulography using fast magnetic resonance imaging. *International Journal of Language & Communication Disorders*, v. 35, n. 3, p. 419–425, 2000.
- MÁDY, K.; SADER, R.; ZIMMERMANN, A.; HOOLE, P.; BEER, A.; ZEILHOFE, H.; HANNIG, C. Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. In: *Proceedings of the 4th International Speech Motor Conference*, Nijmegen, The Netherlands, 2001, p. 142–145.
- MONTGOMERY, D. C. *Design and analysis of experiments*. John Wiley & Sons, 2006.
- MUNHALL, K. G.; BATESON, E. V.; TOKHURA, Y. X-Ray Film Database for Speech Research. *Journal of the Acoustical Society of America*, v. 98, n. 2, p. 1222–1224, 1995.
- NARAYANAN, S.; BYRD, D.; KAUN, A. Geometry, kinematics, and acoustics of Tamil liquid consonants. *Journal of the Acoustical Society of America*, v. 106, n. 4, p. 1993–2007, 1999.

- NARAYANAN, S.; NAYAK, K.; LEE, S.; SETHY, A.; BYRD, D. An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America*, v. 115, n. 4, p. 1771–1776, 2004.
- NARAYANAN, S. S.; ALWAN, A. A.; HAKER, K. An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, v. 98, n. 3, p. 1325–1347, 1995.
- NGUYEN, N.; MILANFAR, P. An efficient wavelet-based algorithm for image superresolution. In: *Proceedings of the International Conference on Image Processing (ICIP 2000)*, 2000.
- ONG, D.; STONE, M. Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope*, v. 1, p. 1–13, 1998.
- PAPOULIS, A. *Probability, random variables and stochastic processes*. 2nd edition ed. McGraw-Hill, 1984.
- 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, 2003.
- PATTI, A. J.; SEZAN, M. I.; TEKALP, A. M. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Image Processing Magazine*, v. 6, p. 1064–1076, 1997.
- PENNEY, G. P.; SCHNABEL, J. A.; RUECKERT, D.; VIERGEVER, M. A.; NIESSEN, W. J. Registration-based interpolation. *IEEE Transactions on Medical Imaging*, v. 23, n. 7, p. 922–926, 2004.
- PRATT, W. K. *Digital image processing: The scientific inside*. Wiley-Interscience, 2007.
- PRENDERGAST, R. S.; NGUYEN, T. Q. A block-based super-resolution for video sequences. In: *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP 2008)*, San Diego, CA, 2008, p. 1240–1243.
- RAJAN, D.; CHAUDHURI, S. Data fusion techniques for super-resolution imaging. *Information Fusion*, v. 3, n. 1, p. 25–38, 2002.
- ROHR, K. *Landmark-based image analysis: using geometric and intensity models*. Computational imaging and vision. Kluwer Academic Publishers, 2001.
- RUECKERT, D. Nonrigid Registration: Concepts, Algorithms, and Applications. In: HAJNAL, J.; HAWKES, D.; HILL, D., eds. *Medical Image Registration, The Biomedical Engineering Series*, 1 ed, New York: CRC Press, (Chapter 13), 2001.
- RUECKERT, D.; ALJABAR, P. Nonrigid Registration of Medical Images: Theory, Methods, and Applications. *IEEE Signal Processing Magazine*, v. 27, n. 4, p. 113–119, 2010.
- RUECKERT, D.; SODONA, L. I.; HAYES, C.; HILL, D. L. G.; LEACH, M. O.; HAWKES, D. J. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, v. 18, n. 8, p. 712–721, 1999.

- SANTA-MARTA, C.; LAFUENTE, J.; VAQUERO, J. J.; GARCIA-BARRENO, P.; DESCO, M. Resolution recovery in Turbo Spin Echo using segmented Half Fourier acquisition. *Magnetic Resonance Imaging*, v. 22, p. 369–378, 2004.
- SCHULTZ, A.; VELHO, L.; SILVA, E. Uma investigação empírica do desempenho da amostragem compressiva em codificação de imagens. In: *Brazilian Telecommunications Symposium (SBrT)*, 2009.
- 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, 1996.
- SEDERBERG, T. W.; PARRY, S. R. Free-form deformation of solid geometric models. *SIGGRAPH Comput. Graph.*, v. 20, n. 4, p. 151–160, 1986.
- SHAWKER, T. H.; SONIES, B.; STONE, M.; BAUM, B. J. Real-time ultrasound visualization of tongue movement during swallowing. *Journal of Clinical Ultrasound*, v. 11, n. 9, p. 485–490, 1983.
- SINGH, M.; BASU, A.; MANDAL, M. Event Dynamics Based Temporal Registration. *Multimedia, IEEE Transactions on*, v. 9, n. 5, p. 1004–1015, 2007.
- STARK, H.; OSKOUI, P. High-resolution image recovery from image-plane arrays, using convex projections. *Journal of the Optical Society of America A*, v. 6, n. 11, p. 1715–1726, 1989.
- STARK, H.; YANG, Y. *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics*. New York, NY, USA: John Wiley & Sons, Inc., 1998.
- STONE, M. *The handbook of phonetic sciences*, cap. Laboratory techniques for investigating speech articulation Blackwell, Oxford, p. 11–32, 1997.
- STONE, M.; DAVIS, E.; DOUGLAS, A.; NESSAIVER, M.; GULLAPALLI, R.; LEVINE, W.; LUNDBERG, A. Modeling Tongue Surface Contours From Cine-MRI Images. *Journal of Speech, Language, and Hearing Research*, v. 44, n. 5, p. 1026–1040, 2001.
- STONE, M.; LUNDBERG, A. Three-dimensional tongue surface shapes of english consonants and vowels. *The Journal of the Acoustical Society of America*, v. 99, n. 6, p. 3728–3737, 1996.
- STRAUSS, D. J. Clustering on colored lattice. *Journal of Applied Probability*, v. 14, p. 135–143, 1977.
- STUDHOLME, C.; HILL, D. L. G.; HAWKES, D. J. Automated 3D registration of MR and PET brain images by multi-resolution optimization of voxel similarity measures. *Med. Phys.*, v. 24, n. 1, p. 25–35, 1997.
- SURESH, K. V.; RAJAGOPALAN, A. N. Robust and computationally efficient superresolution algorithm. *Journal Optical Society of America A*, v. 24, n. 4, p. 984–992, 2007.

- TAKEMOTO, H.; HONDA, K.; MASAKI, S.; SHIMADA, Y.; FUJIMOTO, I. Measurement of temporal changes in vocal tract area function from 3D CINE-MRI data. *Journal of the Acoustical Society of America*, v. 119, n. 2, p. 1037–1049, 2006.
- TEKALP, A. M.; OZKAN, M. K.; SEZAN, M. I. High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration. In: *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, San Francisco, USA, 1992, p. 169–172.
- THÉVENAZ, P.; BLU, T.; UNSER, M. Interpolation Revisited. *IEEE Transactions on Medical Imaging*, v. 19, n. 7, p. 739–758, 2000.
- TSAI, R. Y.; HUANG, T. S. Multi-frame image restoration and registration. *Advances in Computer Vision and Image Processing*, v. 1, n. 2, p. 317–339, 1984.
- UR, H.; GROSS, D. Improved resolution from sub-pixel shifted pictures. *CVGIP: Graphical Models and Image Processing*, v. 54, p. 181–186, 1992.
- VASCONCELOS, M. J. M.; RUA VENTURA, S. M.; FREITAS, D. R. S.; TAVARES, J. M. R. S. Towards the Automatic Study of the Vocal Tract From Magnetic Resonance Images. *Journal of Voice*, v. In Press, Corrected Proof, p. –, 2010.
- VIOLA, P. A. *Alignment by Maximization of Mutual Information*. Tese de Doutorado, Massachusetts Institute of Technology, 1995.
- WAHBA, G. *Spline Models for Observational Data*. SIAM: Society for Industrial and Applied Mathematics, 1990.
- WANG, Z.; QI, F. On ambiguities in super-resolution modeling. *IEEE Signal Processing Letters*, v. 11, n. 8, p. 678–681, 2004.
- WANG, Z.; QI, F. Analysis of multiframe super-resolution reconstruction for image anti-aliasing and deblurring. *Image Vision Comput.*, v. 23, n. 4, p. 393–404, 2005.
- WHEELER, F. W.; HOCTOR, R.; BARRETT, E. Super-resolution image synthesis using projections onto convex sets in the frequency domain. In: BOUMAN, C. A.; MILLER, E. L., eds. *Computational Imaging III. Proceedings of the SPIE*, 2005, p. 479–490.
- YEH, S.; STARK, H. Iterative and One-step Reconstruction from Nonuniform Samples by Convex Projections. *Journal of the Optical Society of America A*, v. 7, p. 491–499, 1990.
- ZIBETTI, M. V. W. *Super-resolução Simultânea para Sequências de Imagens*. Tese de Doutorado em Engenharia Elétrica, Universidade Federal de Santa Catarina, 2007.
- ZITOVÁ, B.; FLUSSER, J. Image registration methods: a survey. *Image and Vision Computing*, v. 21, n. 11, p. 977–1000, 2003.