SAULO ROBERTO SODRÉ DOS REIS

**Um método iterativo e escalonável para super-resolução de
imagens usando a interpolação DCT e representação esparsa**

São Paulo

2014

SAULO ROBERTO SODRÉ DOS REIS

**Um método iterativo e escalonável para super-resolução de
imagens usando a interpolação DCT e representação esparsa**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do título de Doutor em Ciências

Orientadora: Prof^a. Dra Graça Bressan

São Paulo

2014

SAULO ROBERTO SODRÉ DOS REIS

**Um método iterativo e escalonável para super-resolução de
imagens usando a interpolação DCT e representação esparsa**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do título de Doutor em Ciências

Área de Concentração:
Engenharia de Computação

Orientadora: Prof^ª. Dra. Graça Bressan

São Paulo

2014

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 23 de junho de 2014.

Assinatura do autor _____

Assinatura do orientador _____

Catlogação-na-publicação

Reis, Saulo Roberto Sodré dos

**Um método iterativo e escalonável para super-resolução de imagens usando a interpolação DCT e representação esparsa / S.R.S. dos Reis. -- versão corr. -- São Paulo, 2014.
130 p.**

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Algoritmos para imagens 2.Imagem digital 3.Processamento de imagens I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

Dedico este Trabalho a minha esposa Rubia e aos meus filhos Victor e Julia.

AGRADECIMENTOS

Agradeço primeiramente a DEUS pela tua grandeza, pelo seu amor incondicional, por nunca desistir de mim e por me amparar nos momentos mais difíceis.

À minha Orientadora, **Profa. Dra. Graça Bressan**, na condução deste trabalho. Suas sugestões e reflexões foram fundamentais para o desenvolvimento do trabalho.

Aos membros da banca de qualificação, **Prof. Dr Wilson Ruggiero** e **Profa. Dra. Regina Melo**, pelos conselhos e sugestões feitos durante a qualificação do trabalho.

Aos coordenadores do Dinter, **Profa. Dra. Patricia Souza (UFMT)**, **Prof. Dr. Cristiano Maciel (UFMT)**, **Prof. Dr. André Hirakawa (USP)** pelo apoio e esforço na disponibilização de recursos para a realização deste doutorado.

Aos amigos do Dinter (**Alexandre, Clovis, Evandro, Eunice, Obana, Raul, João Paulo, Evandro, Wanessa, Wesley, Soraia, Luciana, Ivairton, João e Mara**), pela amizade, incentivo e convívio harmonioso durante o período em que cursamos as disciplinas em Cuiabá e posteriormente nas viagens e encontros em São Paulo.

Aos colegas professores do Departamento de Engenharia Elétrica da UFMT que deram todo o apoio necessário no período em fiquei afastado para conclusão deste trabalho.

Aos colegas do LARC pela amizade e incentivo na realização deste trabalho.

Não poderiam faltar alguns agradecimentos especiais,

Aos amigos e familiares pela compreensão nos momentos de ausência para realização deste trabalho.

A minha esposa **Rubia** e meus filhos **Victor** e **Julia**, pelo carinho e compreensão nos períodos em que não pude dar a atenção necessária.

A todos o meu muito obrigado.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.
(Marthin Luther King)

RESUMO

Num cenário em que dispositivos de aquisição de imagens e vídeo possuem recursos limitados ou as imagens disponíveis não possuem boa qualidade, as técnicas de super-resolução (SR) apresentam uma excelente alternativa para melhorar a qualidade das imagens. Nesta tese é apresentada uma proposta para super-resolução de imagem única que combina os benefícios da interpolação no domínio da transformada DCT e a eficiência dos métodos de reconstrução baseados no conceito de representação esparsa de sinais. A proposta busca aproveitar as melhorias já alcançadas na qualidade e eficiência computacional dos principais algoritmos de super-resolução existentes. O método de super-resolução proposto implementa algumas melhorias nas etapas de treinamento e reconstrução da imagem final. Na etapa de treinamento foi incluída uma nova etapa de extração de características utilizando técnicas de aguçamento por máscara de nitidez e construção de um novo dicionário. Esta estratégia busca extrair mais informações estruturais dos fragmentos de baixa e alta resolução do conjunto de treinamento e ao mesmo tempo reduzir o tamanho dos dicionários. Outra importante contribuição foi a inclusão de um processo iterativo e escalonável no algoritmo, reinserindo no conjunto de treinamento e na etapa de reconstrução, uma imagem de alta resolução obtida numa primeira iteração. Esta solução possibilitou uma melhora na qualidade da imagem de alta resolução final utilizando poucas imagens no conjunto de treinamento. As simulações computacionais demonstraram a capacidade do método proposto em produzir imagens com qualidade e com tempo computacional reduzido.

Palavras-Chave: Super-Resolução, Interpolação DCT, Representação Esparsa de Sinais, Métodos baseados em exemplos.

ABSTRACT

In a scenario in which the acquisition systems have limited resources or available images do not have good quality, the super-resolution (SR) techniques have become an excellent alternative for improving the image quality. In this thesis, we propose a single-image super-resolution (SR) method that combines the benefits of the DCT interpolation and efficiency of sparse representation method for image reconstruction. Also, the proposed method seeks to take advantage of the improvements already achieved in quality and computational efficiency of the existing SR algorithms. The proposed method implements some improvements in the dictionary training and the reconstruction process. A new dictionary was built by using an unsharp mask technique to characteristics extraction. Simultaneously, this strategy aim to extract more structural information of the low resolution and high resolution patches and reduce the dictionaries size. Another important contribution was the inclusion of an iterative and scalable process by reinserting the HR image obtained of first iteration. This solution aim to improve the quality of the final HR image using a few images in the training set. The results have demonstrated the ability of the proposed method to produce high quality images with reduced computational time.

Keywords: Super-Resolution, DCT Interpolation, Sparse Representation, Learning-based methods.

LISTA DE FIGURAS

Figura 1.1 – Ilustração do conceito de resolução espacial para uma imagem digital.....	20
Figura 2.1 – Ilustração do Processo de aquisição de múltiplas imagens a partir de um único dispositivo. As imagens LR1, LR2 e LR3 representam diferentes aquisições em baixa resolução. A reconstrução da imagem de alta resolução a direita aproveita os deslocamentos espaciais das imagens de baixa resolução	27
Figura 2.2 – Ilustração do processo de aquisição de múltiplas imagens a partir de múltiplos dispositivos. As imagens LR1, LR2 e LR3 representam diferentes aquisições em baixa resolução utilizando diferentes dispositivos. A reconstrução da imagem de alta resolução a direita aproveita os deslocamentos espaciais das imagens de baixa resolução, bem como os diferentes efeitos produzidos pelos dispositivos.....	28
Figura 2.3 – Ilustração do processo de aquisição de uma sequência de Vídeo. Os deslocamentos espaciais e temporais dos pixels dos quadros de baixa resolução N-1, N, N+1 e N+2, são utilizados no processo de super-resolução dos quadros de alta resolução.....	29
Figura 2.4 – Ilustração do processo de aquisição de uma única imagem de baixa resolução. Os componentes de alta frequência da imagem HR final são recuperados a partir de um conjunto de treinamento formado por imagens de baixa e alta resolução.....	30
Figura 2. 5 - Modelo aquisição que relaciona a imagem de alta resolução com a imagem a imagem de baixa resolução.....	31
Figura 2.6 - Ilustração do processo de interpolação não uniforme	33
Figura 2.7 - Ilustração do Método IBP- Iterative Back Projection	35
Figura 2. 8 - Ilustração de um método típico de super-resolução baseado em exemplos	40
Figura 2. 9 - Ilustração geral dos algoritmos k-SVD e OMP. O dicionário D na parte superior da figura é inicialmente utilizado e posteriormente são realizadas as etapas iterativas de alternadas de codificação esparsa e atualização do dicionário.....	45
Figura 2.10 - Representação gráfica do núcleo da transformada discreta do cosseno para um bloco de tamanho 8×8	47

Figura 2.11 – Ilustração do Método de Interpolação do Domínio da Transformada DCT para um bloco (8x8) pixels da imagem de baixa resolução	49
Figura 4.1 - Diagrama Geral do Método de Super-Resolução Proposto.	59
Figura 4.2 - Construção do conjunto de treinamento formado por imagens de alta resolução e seu equivalente de baixa resolução.	61
Figura 4.3 - Representação dos fragmentos (patches) de alta e baixa resolução da imagem.	62
Figura 4.4 - Visão geral da etapa de treinamento e construção dos dicionários.	66
Figura 4.5 - Algumas imagens do conjunto de treinamento.	70
Figura 4.6 - Qualidade visual da imagem “Man” para fator ampliação $s = 2$ e dicionário com 512 átomos. (a) Imagem Original, (b) Interpolação Bicúbica, (c) SR baseada em representação esparsa (ZEYDE; ELAD; PROTTER, 2012), (d) Método Proposto.	78
Figura 4.7 - Qualidade visual na imagem “Bridge” para fator interpolação $s = 2$ e dicionário com 512 átomos. (a) Imagem Original, (b) Interpolação Bicúbica, (c) SR baseada em representação esparsa (ZEYDE; ELAD; PROTTER, 2012), (d) Método Proposto.	79
Figura 4.8 - Testes de qualidade visual na imagem “Pepper” para fator interpolação $s = 2$ e dicionário com 512 átomos. (a) Imagem Original, (b) Interpolação Bicúbica, (c) SR baseada em representação esparsa (ZEYDE; ELAD; PROTTER, 2012), (d) Método Proposto.	80
Figura 5.1 - Ilustração unidimensional do aguçamento por máscara de Nitidez (a) sinal Original, (b) Sinal borrado, (c) Máscara de Nitidez e (d) Sinal realçado pelo aguçamento.	84
Figura 5.2 - Diagrama geral do treinamento e construção dos dicionários para o método iterativo proposto de super-resolução.	87
Figura 5.3 - Diagrama geral do método proposto de super-resolução.	90
Figura 5.4 - Método Iterativo proposto usando o mesmo fator de ampliação. A imagem $IHRF(1)$ é inserida no conjunto de treinamento com o mesmo fator de interpolação da imagem $IHRF(F)$	92

Figura 5.5 - Método Iterativo proposto usando fatores de ampliação diferentes. A imagem $IHRF(1)$ é inserida no conjunto de treinamento com um fator de interpolação da imagem menor que $IHRF(F)$.	93
Figura 5.6 - Exemplos de dois Dicionários de alta resolução construídos na etapa de treinamento. (a) Dicionário (D_{H1}), (b) Dicionário (D_{H2})	94
Figura 5.7 - Conjunto de imagens escolhidas para o conjunto reduzido de treinamento (5 imagens), considerando as imagens de teste Man, Lena, Bridge, Pepper e Babbon	96
Figura 5.8 - Qualidade Visual da imagem “Pepper” para o Método Proposto com fator interpolação $s = 2$ e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT ; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT	103
Figura 5.9 - Qualidade visual do método proposto para a imagem “CoastGuard”, fator de ampliação $s = 2$ e dicionários de tamanho 512 átomos. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT.	104
Figura 5.10 - Qualidade visual do método proposto para a imagem “Lena”, fator de ampliação $s = 3$ e dicionários de tamanho 512 átomos. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT	105
Figura 5.11 - Qualidade visual do método proposto para a imagem “Foreman”, fator de ampliação $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT.	106
Figura 5.12 - Tempo execução com fator de ampliação $s = 2$	108
Figura 5.13 - Tempo execução com fator de ampliação $s = 3$	108
Figura 5.14 - Implementação do algoritmo com a alteração da imagem $YHRF(1)$ no processo de reconstrução da imagem.	110

- Figura 5.15 - Qualidade visual para a imagem “Lake” considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem *IHRF1*. 112
- Figura 5.16 - Qualidade visual para a imagem *Jetplane* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem *IHRF1*. 113
- Figura 5.17 - Qualidade visual para a imagem *Cameraman* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem *IHRF1*. 114
- Figura 5.18 - Qualidade visual para a imagem *Living room* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem *IHRF1*. 115
- Figura 5.19 - Qualidade visual para a imagem *car plate* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem *IHRF1*. 116

LISTA DE TABELAS

Tabela 4.1 - Comparação em PSNR [dB] do método SR proposto, interpolação bicúbica, método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s=2$ e dicionário com 512 átomos.	74
Tabela 4.2 - Comparação em SSIM do Método SR proposto, Interpolação Bicúbica, Método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s=2$ e dicionário com 512 átomos.	75
Tabela 4.3 - Comparação em PSNR [dB] do método SR proposto, Interpolação Bicúbica, método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s=3$ e dicionário com 512 átomos.	76
Tabela 4.4 - Comparação em SSIM do Método SR proposto, Interpolação Bicúbica, Método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s=3$ e dicionário com 512 átomos.	77
Tabela 5.1 - Comparação em PSNR [dB] e SSIM, do Método Proposto com os métodos de interpolação bicúbica e DCT, para um conjunto treinamento Completo, e fator de ampliação $s = 2$	97
Tabela 5.2 - Comparação em PSNR [dB] e SSIM, do método proposto com os métodos de interpolação bicúbica e DCT, para um conjunto treinamento reduzido (5 imagens), e fator de ampliação $s = 2$	98
Tabela 5.3 - Comparação em PSNR [dB] e SSIM, do método proposto com os métodos de interpolação bicúbica e DCT, para um conjunto completo de treinamento (92 imagens), e fator de ampliação $s = 3$	100
Tabela 5.4 - Comparação em PSNR [dB] e SSIM, do método proposto com os métodos de interpolação bicúbica e DCT, para um conjunto reduzido de treinamento (5 imagens), e fator de ampliação $s = 3$	101
Tabela 5.5 - Comparação do tempo de execução do método proposto com os métodos de interpolação bicúbica, método SR proposto por Zeyde et al (2012) e o método proposto nesta tese no capítulo 4.....	107

Tabela 5.6 - Comparação em PSNR [dB] e SSIM, do método proposto com a substituição da imagem $Y_{HRF}(1)$, e outros métodos de super-resolução propostos na literatura, para um conjunto treinamento composto por 5 imagens, e fator de ampliação de $s = 3$	111
---	-----

LISTA DE ABREVIATURA E SIGLAS

Sigla	Descrição
SR	Super-Resolution
BCGD	Block Coordinate Gradient Descent
CCD	Charge Coupled Device
CMOS	Complementary Metal Oxide Semiconductor
DCT	Discrete Cosine Transform
HR	High Resolution
IBP	Iterative Back Propagation
KLT	Karhunen-Loève Transform
LASSO	Least Absolute Shrinkage and Selection Operator
LLE	Locally Linear Embedding
LR	Low Resolution
MAP	Maximum a Posteriori
OMP	Orthogonal Matching Pursuit
PCA	Principal Components Analysis
PDF	Probability Density Function
PMN	Parametric Markov Network
POCS	Projection onto convex set
PSNR	Peak Signal to Noise
SSIM	Structural Similarity
SVD	Singular Value Decomposition
WLD	Weber Local Descriptor
ASDS	Adaptive Sparse Domain Selection
MPEG	Moving Picture Expert Group

SUMÁRIO

Capítulo 1	19
Introdução	19
1.1 Motivações e Justificativas	21
1.2 Objetivos	22
1.3 Metodologia	23
1.4 Contribuições	24
1.5 Organização da Tese	24
Capítulo 2	26
Fundamentação Teórica	26
2.1 Super-Resolução	26
2.2 Formas de aquisição das imagens de Baixa Resolução	27
2.3 Modelo de Observação da Imagem	30
2.4 Métodos de Super-Resolução	33
2.4.1 Métodos baseados em Interpolação	33
2.4.2 Iterative Back Projection - IBP	34
2.4.3 Projection Onto Convex Set - POCS	36
2.4.4 Métodos Regularizados	36
2.4.5 Métodos baseados em Exemplos	39
2.5 Métodos de Super-Resolução Baseados em Representação Esparsa de Sinais	41
2.6 Algoritmo para Treinamento de Dicionários k -SVD	42
2.7 Transformada Discreta de Cossenos - DCT	45
2.8 Interpolação no Domínio da Transformada DCT	47
Capítulo 3	50
Trabalhos Relacionados	50
3.1 Introdução	50
3.2 Principais Trabalhos Usando Métodos de Representação Esparsa de Sinais	51
3.3 Principais Trabalhos de Super-Resolução Usando a Transformada DCT	53
3.4 Análise das Propostas Estudadas	55

3.5 Conclusões do Capítulo	57
Capítulo 4	58
Super-Resolução de Imagem Usando Interpolação DCT e o Modelo de	
Representação Esparsa	58
4.1 Introdução	58
4.2 Treinamento e construção dos dicionários usando o método de	
Representação esparsa de sinais.....	60
4.2.1 Construção do Conjunto de imagens de alta e baixa resolução.....	60
4.2.2 Extração de Características a partir dos Fragmentos da Imagem.....	61
4.2.3 Redução da Dimensionalidade com o Algoritmo PCA	63
4.2.4 Treinamento dos Dicionários com os Algoritmos K-SVD e OMP	64
4.2.5 Construção do Dicionário de Alta Resolução	65
4.3 Reconstrução da Imagem de Alta Resolução	66
4.3.1 Interpolação da Imagem de Baixa Resolução	66
4.3.2 Filtragem Passa-Alta e Extração dos Fragmentos	67
4.3.3 Codificação Esparsa dos Fragmentos.....	68
4.3.4 Construção dos Fragmentos de Alta Resolução	68
4.3.5 Construção da Imagem com as Componentes de Alta Frequência	68
4.3.6 Construção da Imagem de Alta Resolução Final	69
4.4 Experimentos Realizados com o Algoritmo Proposto	69
4.4.1 Parâmetros do Experimento.....	69
4.4.2 Algoritmos Implementados	71
4.4.3 Métricas Utilizadas	71
4.4.4 Resultados Obtidos	73
4.5 Conclusões do Capítulo	81
Capítulo 5	82
Implementação Iterativa e Escalonável do Método Proposto	82
5.1 Introdução	82
5.2 Treinamento e Construção dos Dicionários	82
5.3 Reconstrução da Imagem de Alta Resolução Final	87
5.4 Etapa Iterativa e Escalonável do Método Proposto.....	91
5.4.1 Imagem de alta resolução parcial com o mesmo fator de ampliação.....	91
5.4.2 Imagem Alta Resolução Parcial com o Fator de Ampliação Diferente	92
5.5 Experimentos Realizados com o Algoritmo Proposto	94

5.5.1 Parâmetros do Experimento.....	94
5.5.2 Comparações com Algoritmos de Interpolação.....	95
5.5.3 Métricas Utilizadas	95
5.5.4 Resultados Obtidos para o Fator de Ampliação $s = 2$	95
5.5.5 Resultados Obtidos para o Fator de Ampliação $s = 3$	99
5.5.6 Qualidade Visual do Método Proposto	102
5.5.7 Comparativo em relação ao tempo de execução dos algoritmos	107
5.6 Implementação do algoritmo com a inclusão da Imagem Alta	
Resolução parcial no processo de reconstrução	109
5.6.1 Comparativo com outros métodos.....	110
5.6.2 Resultados experimentais de Qualidade Visual	111
5.7 Conclusões do Capítulo	117
Capítulo 6	118
Conclusão	118
Referências	121
APÊNDICE A – Normas Vetoriais	128
APÊNDICE B – Argumentos máximos e mínimos de uma função	129

Capítulo 1

Introdução

Os últimos anos trouxeram um grande avanço e popularidade para a área de processamento de imagens e vídeos com o desenvolvimento rápido e crescente de novas tecnologias e dispositivos de maior capacidade de processamento e baixo custo. Os dispositivos digitais como câmeras, filmadoras, celulares, tablets se tornaram itens obrigatórios e necessários em muitas atividades cotidianas.

Um efeito natural deste processo é a necessidade cada vez maior de produção e utilização de conteúdos digitais de alta qualidade em diversas áreas, como por exemplo, o entretenimento, o sensoriamento remoto, os sistemas de segurança, a TV Digital, a medicina e os serviços de multimídia através da internet.

Na medicina, as imagens e vídeos de alta resolução têm sido utilizados para a realização de diagnósticos mais precisos, diminuição do tempo de exames laboratoriais, realização de ultrassonografias, entre outros. A utilização das imagens médicas de boa resolução ajuda os médicos na interpretação das imagens, e consequentemente na elaboração de um diagnóstico mais preciso para o paciente (GREENSPAN, 2008, BODENSTEINER; DAROLTI; SCHWEIKARD, 2009, ETTER; JOVANOVIC; VETTERLI, 2011). No sensoriamento remoto as imagens são importantes para o monitoramento de áreas ambientais, áreas urbanas, agricultura, cartografia, mineração, energia e defesa. O uso de imagens de melhor resolução espacial proporciona uma melhor interpretação e análise da imagem da área sensoriada (SHEN et al., 2009, LU; INAMURA, 2009).

Na segurança, a utilização os vídeos e imagens de melhor qualidade podem ajudar as autoridades policiais na identificação de suspeitos em ações de vandalismos e crimes. (GUNTURK et al., 2003; LIN et al., 2007). Com a popularização da TV digital, a programação em alta-definição tem sido ampliada nas emissoras, distribuindo também estes conteúdos para dispositivos móveis (NG et al., 2007; BOOM et al., 2006)

A qualidade de uma imagem ou vídeo está diretamente relacionada à sua *resolução*, ou seja, ao nível de detalhe da representação visual definidos na forma de resolução espacial e de resolução de intensidade. De acordo com (GONZALEZ; WOODS, 2010), a resolução espacial é considerada uma medida do menor parâmetro perceptível da imagem e pode ser definida quantitativamente como pares de linha por unidade de área ou pixels por unidade de área. A Figura 1.1 ilustra o conceito de resolução espacial para uma imagem digital.

Figura 1.1 – Ilustração do conceito de resolução espacial para uma imagem digital.

Fonte: imagesurvey.com.br

Os dispositivos de aquisição são equipados com sensores de imagem matriciais construídos a partir de duas tecnologias: CCD (*charge-coupled device*) e CMOS (*complementary metal-oxide semiconductor*). Os sensores do tipo CCD têm sido amplamente utilizados em câmeras digitais, pois permite uma maior aproximação de pontos, produzindo imagens com maior qualidade. Entretanto, estes sensores apresentam um consumo maior de energia e um custo de fabricação elevado (GONZALEZ; WOODS, 2010). Os sensores do tipo CMOS apresentam uma qualidade ainda inferior aos dispositivos CCD, porém, apresentam uma resposta mais rápida e uma menor dissipação de energia no chip (MAGMAN, 2003; BAKER; KANADE, 2002).

Podemos observar que o aumento do número de sensores de imagem acarreta numa maior resolução e melhor qualidade da imagem digital. Entretanto, este aumento na densidade nos sensores, ou seja, o aumento do número de pixels por unidade de área, pode aumentar o ruído no processo de aquisição e elevar significativamente o custo do dispositivo, o que não é desejável em muitas aplicações (ZIBETTI, 2007; GONZALEZ; WOODS, 2010).

Nos últimos anos diversos estudos tem abordado o uso de técnicas de processamento de imagem para aumentar a resolução considerando os dispositivos de aquisição existentes de baixa resolução. As técnicas de super-resolução (do inglês, *Super-Resolution-SR*) permitem melhorar a resolução de uma imagem de baixa resolução ou mesmo um conjunto de imagens, utilizando técnicas de pós-processamento e aproveitando características presentes de outras imagens de baixa-resolução capturadas (PARK; PARK; KANG, 2003).

As estratégias existentes empregam técnicas de interpolação, técnicas baseadas na reconstrução a partir de multiplas imagens de baixa resolução, ou utilizando técnicas baseadas em exemplos. Essas estratégias serão detalhadas nos capítulos seguintes, a partir da análise dos principais trabalhos da literatura.

1.1 Motivações e Justificativas

Dentro de um cenário em que os dispositivos de aquisição possuem recursos limitados ou as imagens disponíveis não possuem boa qualidade, as técnicas de super-resolução apresentam uma excelente alternativa para melhorar a qualidade das imagens.

Os métodos de super-resolução propostos na literatura conseguem reconstruir imagens com qualidade superior aos métodos tradicionais de interpolação bicúbica e bilinear. Os algoritmos baseados em reconstrução, como os apresentados em (PARK; PARK; KANG, 2003), garantem um bom compromisso entre eficiência computacional e qualidade na reconstrução da imagem de alta resolução para fatores de ampliação de até duas vezes o tamanho da imagem de baixa resolução capturada. Os algoritmos baseados em exemplos buscam extrair os detalhes mais

finos da imagem, a partir de um conjunto de treinamento formado por imagens de alta e baixa resolução. Os resultados recentes apresentados na literatura tem demonstrado uma qualidade superior na reconstrução quando considerados fatores de ampliação maiores que duas vezes o tamanho da imagem de baixa resolução (FREEMAN; JONES; PASZTOR, 2002; ELAD; DATSENKO, 2009).

Ainda assim, alguns desafios precisam ser enfrentados para tornar os algoritmos mais eficientes e aplicáveis em diversas áreas. Dentre eles podemos citar:

- Os métodos de super-resolução devem prover um conteúdo digital de alta qualidade utilizando recursos limitados de processamento e armazenamento. Os algoritmos atuais conseguem reconstruir uma imagem de melhor resolução que algoritmos de interpolação, porém com um tempo computacional ainda elevado;
- Os métodos de super-resolução devem possibilitar a melhoria na qualidade para fatores de ampliação maiores. Os algoritmos atuais têm mostrado eficiência e qualidade na super-resolução para fatores de ampliação da imagem de baixa resolução de até duas vezes. Entretanto, para algumas aplicações é fundamental que as técnicas de super-resolução consigam reconstruir uma imagem de qualidade para um fator de ampliação maior que duas vezes.

1.2 Objetivos

Considerando as motivações e justificativas expostas, esta tese propõe um novo método de super-resolução iterativo e escalonável para imagem única que busca um compromisso entre qualidade na reconstrução da imagem de alta resolução final e eficiência computacional.

1.3 Metodologia

Para a realização e validação deste trabalho, as seguintes etapas foram contempladas de acordo com a sequência a seguir:

- Inicialmente foi realizada uma ampla revisão de literatura sobre o tema super-resolução de imagens, de forma a identificar algumas oportunidades de melhoria no aspecto de qualidade e complexidade computacional;
- Foi realizado um estudo e análise dos principais trabalhos relacionados a super-resolução de imagens usando algoritmos baseados em exemplos e algoritmos de interpolação no domínio da Transformada Discreta de Cossenos (DCT);
- A partir desta análise, foi possível identificar os problemas e dificuldades associadas à melhoria da qualidade da imagem e redução do tempo computacional, considerando as diversas etapas do processo de reconstrução da imagem de alta resolução;
- Foi elaborado um algoritmo de super-resolução para imagem única baseado numa estratégia de uso conjunto da transformada DCT e a representação esparsa de sinais;
- Foram realizadas várias simulações computacionais para diferentes cenários de aplicação, utilizando a ferramenta MATLAB;
- A partir da análise dos resultados obtidos, foi elaborado um novo algoritmo com melhorias e aperfeiçoamentos de forma a tornar o algoritmo mais eficiente e com tempo computacional reduzido;
- Novamente, foram realizados novos testes para diferentes cenários de aplicação em MATLAB.

1.4 Contribuições

A primeira contribuição desta tese é o desenvolvimento de um novo método de super-resolução para imagem única que combina a interpolação no domínio da transformada DCT e a representação esparsa de sinais. O método que será apresentado no capítulo 4 se divide em duas principais etapas: a primeira de treinamento e construção de dicionários, e uma segunda de reconstrução da imagem final de alta resolução. As contribuições foram publicadas em (REIS; BRESSAN, 2013a; REIS; BRESSAN, 2013b).

A segunda contribuição é a implementação de melhorias no método proposto para diminuir o tempo computacional e melhorar a qualidade da imagem final de alta resolução. As melhorias propostas foram:

- treinamento e construção de um segundo dicionário compacto de baixa resolução utilizando um conjunto de filtros de aguçamento por *máscara de Nitidez*, com o objetivo de extrair mais informações do conjunto de treinamento;
- implementação de uma estratégia iterativa e escalonável. A imagem de alta resolução construída na primeira iteração com um fator de ampliação diferente da imagem final é reinserida no conjunto de treinamento para uma nova etapa de treinamento e reconstrução da imagem final.

1.5 Organização da Tese

Esta tese está organizada em seis capítulos. O capítulo 1 apresenta a introdução ao trabalho. O capítulo 2 apresenta a fundamentação teórica necessária para o entendimento do trabalho. O capítulo 3 descreve os principais trabalhos de super-resolução de imagens usando representação esparsa e a interpolação no domínio da transformada DCT, buscando correlacionar e contextualizar com o tema proposto nesta tese. No Capítulo 4, é apresentado o algoritmo para super-resolução da imagem combinando a interpolação no domínio da transformada DCT e o método de representação esparsa de sinais. Também são realizadas simulações do algoritmo proposto considerando diferentes parâmetros e comparando com os outros métodos propostos na literatura, utilizando métricas objetivas e subjetivas. No Capítulo 5, são

apresentadas as melhorias e aperfeiçoamentos propostos no algoritmo. Também são realizadas simulações computacionais em diferentes cenários e comparados com as principais técnicas de interpolação e super-resolução. No Capítulo 6, são apresentadas as conclusões e pontos para futuras pesquisas.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica dos principais conceitos relacionados ao tema desta tese. Será abordado o conceito de super-resolução de imagens e vídeos, o modelo de aquisição adotado nos algoritmos de super-resolução e os principais métodos estudados na literatura. Também é descrito o conceito de representação esparsa de sinais e sua aplicação em super-resolução de imagens. Por fim, será descrita brevemente a Transformada Discreta de Cossenos – DCT e sua aplicação no processo de interpolação de imagens.

2.1 Super-Resolução

A super-resolução (*SR*) de imagens ou vídeos é definida como uma técnica de processamento utilizada para melhorar a resolução de uma única imagem ou de uma sequência de imagens, a partir da aquisição de uma imagem ou sequência de baixa resolução. Os algoritmos aproveitam a correlação entre as imagens capturadas e extraem pequenas diferenças, como o movimento de subpixel ou distorções na aquisição para construção da imagem ou sequência de imagens de alta-resolução (RAJAN; CHAUDHURI, 2002; PARK; PARK; KANG, 2003; HUNG, 2012).

A ideia principal do algoritmo é reduzir o custo no processamento da imagem de alta resolução aproveitando as características dos sistemas de aquisição existentes para imagens de baixa resolução.

As técnicas de super-resolução tem uma formulação semelhante às técnicas de restauração de imagens. Entretanto, na restauração de imagens o objetivo não é aumentar a resolução, mas remover as distorções que degradam a qualidade da imagem. Por outro lado, a super-resolução procura remover as distorções e ao mesmo tempo ampliar ou aumentar a resolução da imagem (BANHAM; KATSAGGELOS, 1997; ZIBETTI, 2007).

2.2 Formas de aquisição das imagens de Baixa Resolução

Nas técnicas de super-resolução de imagens, duas informações são importantes no processo de reconstrução: a forma como as imagens de baixa resolução são capturadas e o modelo de observação de imagem adotado. O formato de aquisição das imagens de baixa resolução define qual a melhor estratégia a ser adotada na reconstrução da imagem de alta resolução. As principais formas de aquisição são resumidas a seguir:

a) Aquisição de múltiplas imagens por um dispositivo único: neste formato de aquisição, as imagens de baixa resolução são adquiridas em pequenos intervalos de tempo entre cada imagem ou com pequenos deslocamentos espaciais. Nesta situação, ilustrada na Figura 2.1, os efeitos oriundos do processo de aquisição como, por exemplo, distorção óptica e subamostragem são considerados os mesmos para todas as imagens adquiridas. Sendo assim, a informação necessária para o processo de reconstrução da imagem de alta resolução está presente nos deslocamentos de pixels, ou movimentos de subpixels entre as imagens de baixa resolução (PARK; PARK; KANG, 2003).

Figura 2.1 – Ilustração do Processo de aquisição de múltiplas imagens a partir de um único dispositivo. As imagens LR1, LR2 e LR3 representam diferentes aquisições em baixa resolução. A reconstrução da imagem de alta resolução a direita aproveita os deslocamentos espaciais das imagens de baixa resolução.

Fonte: (PARK; PARK; KANG, 2003).

b) Aquisição de múltiplas imagens usando múltiplos dispositivos: Nesta forma de aquisição as imagens de baixa resolução da cena são capturadas por dispositivos diferentes de aquisição, conforme pode ser observado na Figura 2.2. As informações necessárias para o processo de super-resolução estão presentes nos diferentes efeitos de distorção óptica e subamostragem produzidos pelos dispositivos de aquisição e nos movimentos entre os pixels das imagens de baixa resolução capturadas.

Figura 2.2 – Ilustração do processo de aquisição de múltiplas imagens a partir de múltiplos dispositivos. As imagens LR1, LR2 e LR3 representam diferentes aquisições em baixa resolução utilizando diferentes dispositivos. A reconstrução da imagem de alta resolução a direita aproveita os deslocamentos espaciais das imagens de baixa resolução, bem como os diferentes efeitos produzidos pelos dispositivos.

Fonte: (PARK; PARK; KANG, 2003).

c) Aquisição de uma sequência de vídeo: outra possibilidade de super-resolução de imagens e vídeos ocorre quando uma sequência de vídeo é capturada, conforme pode ser observado na Figura 2.3. Nesta situação, a informação necessária para a super-resolução está presente nos deslocamentos espaciais e

temporais entre os pixels dos quadros. Alguns trabalhos, como em (SCHULTZ; STEVENSON, 1996; SEGALL; MOLINA; KATSAGGELOS, 2003; HUNG, 2011; GARCIA; DOREA; QUEIROZ, 2012), tem explorado bastante esta alternativa para super-resolução de quadros de baixa resolução, utilizando as informações de quadros de alta resolução, ou quadro-chaves, em aplicações com vídeos de resolução mista.

Figura 2.3 – Ilustração do processo de aquisição de uma sequência de Vídeo. Os deslocamentos espaciais e temporais dos pixels dos quadros de baixa resolução N-1, N, N+1 e N+2, são utilizados no processo de super-resolução dos quadros de alta resolução.

Fonte: (PARK; PARK; KANG, 2003).

d) Aquisição de uma única imagem: Neste formato de aquisição apenas uma imagem de baixa resolução é usada no processo de reconstrução, conforme pode ser visto na Figura 2.4. As informações complementares necessárias para a construção da imagem de alta resolução estão presentes em um conjunto de treinamento formado por imagens de alta e baixa resolução, podendo ter ou não conteúdo similar à imagem de baixa resolução adquirida. Utilizando técnicas de aprendizagem de máquina, as imagens do conjunto de treinamento são correlacionadas com a imagem de baixa resolução adquirida, e características não presentes são incorporadas. O método proposto nesta tese se baseia neste formato de aquisição e será detalhado mais adiante.

Figura 2.4 – Ilustração do processo de aquisição de uma única imagem de baixa resolução. Os componentes de alta frequência da imagem HR final são recuperados a partir de um conjunto de treinamento formado por imagens de baixa e alta resolução.

Fonte: Autor

2.3 Modelo de Observação da Imagem

As formas de aquisição apresentadas na seção anterior adotam um modelo de observação da imagem para o processo de super-resolução. O modelo tem como objetivo representar os efeitos e as degradações ocorridas durante a aquisição (PARK; PARK; KANG, 2003; FARSIU et al., 2004; VAN OUWERKERK, 2006; ZIBETTI, 2007). Um dos modelos mais adotados na literatura pode ser representado pela seguinte expressão:

$$Y_L = SHX_H + \eta \quad (2.1)$$

onde Y_L representa a imagem de baixa resolução $Y_L = [y_{L1}, y_{L2}, \dots, y_{LN}]^T$ ordenada lexicograficamente por um vetor de tamanho $N = N_1 \times N_2$ pixels. X_H representa a

imagem de alta resolução $X_H = [x_{H1}, x_{H2}, \dots, x_{HM}]^T$, ordenada lexicograficamente por um vetor de tamanho $M = M_1 \times M_2$, com $M > N$. A matriz H de tamanho $M \times M$ representa as degradações sofridas durante o processo de aquisição. Mais especificamente, no armazenamento de uma imagem digital há uma perda natural de resolução espacial causada pela distorção óptica nas lentes e fotossensores e distorções de foco. Também podem ocorrer algumas degradações devido ao borramento causado por movimento. A matriz S representa os efeitos do processo de subamostragem da imagem capturada. Esta degradação é causada pela digitalização dos valores de amplitude da tensão de saída dos fotossensores, o que implica numa redução do número de amostras ou pixels da imagem. Também devem ser considerados os efeitos do ruído presente no processo de aquisição η , que pode ser modelado como um ruído branco gaussiano de média zero e variância σ_η^2 . A figura 2.5 ilustra os efeitos do sistema de aquisição.

Figura 2. 5 - Modelo aquisição que relaciona a imagem de alta resolução com a imagem a imagem de baixa resolução.

Fonte: (PARK; PARK; KANG, 2003).

A super-resolução é tratada como um problema inverso, ou seja, a Imagem alta resolução é estimada a partir da observação dos dados ou imagens de baixa resolução (FARSIU et al., 2004).

Os algoritmos de super-resolução buscam resolver o problema inverso sem amplificar os efeitos do sistema de aquisição, representado pelas matrizes S e H na equação (2.1). Para resolver o problema inverso, os algoritmos definem uma função custo $J(x)$ que assegura a fidelidade e a proximidade da solução final com os dados observados. Para minimizar a função custo e obter uma solução ótima que resulte no melhor desempenho possível, pode ser utilizada uma solução por mínimos quadrados (do inglês, *Least Squares- LS*) como a seguir:

$$\hat{X}_H = \arg \min_{X_H} J(X_H) = \arg \min_{X_H} \|Y_L - SHX_H\|_2^2 \quad (2.2)$$

onde $J(X_H)$ representa a função custo a ser minimizada. $\|\cdots\|_2$ é uma norma eucliana l_2 adotada para o problema de minimização proposto. O Apêndice A apresenta mais detalhes sobre as principais normas vetoriais. O Apêndice B apresenta mais detalhes sobre mínimo de funções e argumentos.

Considerando que a equação (2.2) pode resultar em mais de uma solução possível, os algoritmos de super-resolução buscam encontrar uma solução única e estável, utilizando informações adicionais das imagens. Nos algoritmos determinísticos, esta informação adicional sobre a imagem é definida como uma penalidade de regularização (ZIBETTI, 2007). Já nos algoritmos estatísticos Bayesianos, a informação adicional é definida na forma de uma distribuição *a priori*. Nas seções seguintes alguns dos principais algoritmos serão apresentados.

2.4 Métodos de Super-Resolução

Conforme abordado no capítulo 1, os métodos de super-resolução podem empregar técnicas de interpolação, técnicas de reconstrução da imagem ou técnicas baseadas em exemplos. Nas próximas seções serão revisados os principais algoritmos aplicados para imagem única ou sequência de imagens.

2.4.1 Métodos baseados em Interpolação

Os primeiros métodos utilizados para super-resolução foram baseados em interpolação não uniforme. O método apresenta três estágios: registro ou estimação de movimentos, interpolação não uniforme e restauração para correção de possíveis distorções ópticas, conforme pode ser observado na Figura 2.6.

As informações de movimento ou deslocamento espacial da sequência de imagens de baixa resolução capturadas são utilizadas para criar uma imagem interpolada com as dimensões da imagem de alta resolução. Como a imagem interpolada possui distorções ópticas que precisam ser corrigidas, é necessária a restauração para correção das distorções ópticas.

Figura 2.6 - Ilustração do processo de interpolação não uniforme.

Fonte: Uma adaptação de (PARK; PARK; KANG, 2003).

Alguns trabalhos propuseram melhorias na etapa de interpolação não uniforme. Em (UR; GROSS, 1992), é proposto um algoritmo de interpolação utilizando o teorema da amostragem multicanal, considerando múltiplas imagens LR com pequenas diferenças espaciais. Em (KOMATSU et al., 1993), foi proposto um algoritmo de Landweber para aplicações de super-resolução de múltiplas imagens capturadas a partir de múltiplas câmeras.

Outras propostas de interpolação com o objetivo de melhorar regiões específicas das imagens, como contornos e bordas, também foram apresentadas na literatura. Em (LI; ORCHARD, 2001), os autores propuseram um método de interpolação direcionado aos contornos das imagens empregando a dualidade geométrica entre coeficientes de covariância das imagens de baixa e alta resolução. Em (ZHANG; XIAOLING, 2006), também foi proposto um algoritmo de interpolação direcionado a contornos, usando filtros direcionais para construção de subconjuntos específicos de pixels.

Os algoritmos de interpolação são vantajosos para situações em que são necessários métodos rápidos e de baixo custo computacional. Entretanto, as imagens interpoladas apresentam um aspecto borrado e com pouca definição dos detalhes mais finos, como texturas e contornos.

2.4.2 Iterative Back Projection - IBP

O método IBP (do inglês, *Iterative Back Projection*), inicialmente proposto por (IRANI; PELEG, 1991) faz uma estimativa da imagem de alta resolução, projetando um erro entre imagens de baixa resolução simuladas e as imagens de baixa resolução realmente capturadas. As imagens de baixa resolução simuladas são obtidas aplicando os efeitos do processo de aquisição, como borramento e subamostragem, numa imagem de alta resolução. Por meio de um processo iterativo, o erro é minimizado e a imagem final de alta resolução é reconstruída conforme pode ser observado na Figura 2.7.

O método IBP pode ser representado pela seguinte expressão:

$$X_H^{n+1} = X_H^n + \sum_{i=1}^k (Y_{Li} - \hat{Y}_{Li}^n) h^{BP} \quad (2.3)$$

onde $\hat{Y}_{Li}^n = SHX_H^n$ representam as projeções em baixa resolução da imagem de alta resolução e h^{BP} é um parâmetro que determina a contribuição do erro no processo iterativo de reconstrução. Outras contribuições, que incluíram o movimento entre imagens no modelo de aquisição, foram apresentadas posteriormente em (MANN; PICARD, 1994).

Figura 2.7 - Ilustração do Método IBP- Iterative Back Projection.

Fonte: Uma adaptação de (IRANI; PELEG, 1991).

2.4.3 Projection Onto Convex Set - POCS

O método POCS (do inglês, *Projection Onto Convex Set*) foi primeiramente proposto por (STARK; OSKOUI, 1989), com outras adaptações propostas posteriormente em (GEVREKCI et al., 2007). O método utiliza informações *a priori* disponíveis das imagens de baixa resolução capturadas para compor uma série de conjuntos convexos. A imagem de alta resolução X_H se situa na intersecção de todos os conjuntos convexos. A busca por X_H é feita por um processo iterativo, cujo resultado é a projeção sobre os conjuntos convexos, conforme a expressão:

$$X_H^{n+1} = P_m P_{m-1} \cdots P_1 P_0 X_H^n \quad (2.4)$$

onde P_m é um operador de projeção para o m -ésimo conjunto convexo. No método POCS, se os conjuntos forem adequadamente especificados, a solução do problema está na intersecção dos conjuntos (HUANG et al., 2005; ZIBETTI, 2007).

Entre as vantagens do método POCS está a utilização conveniente das informações *a priori* e a possibilidade de incorporar novas informações *a priori* no processo de reconstrução. Entretanto, o método apresenta dificuldades na determinação dos operadores de projeção P_i , que requer o cálculo de matrizes inversas, resultando em um elevado custo computacional do algoritmo e a uma solução não única.

2.4.4 Métodos Regularizados

Os métodos regularizados buscam uma solução única e estável através da minimização do erro entre as imagens de baixa resolução observadas e a imagem de alta resolução, acrescentando uma penalidade de regularização (ZIBETTI, 2007; ZIBETTI et al., 2011).

Os métodos regularizados se dividem em determinísticos e estatísticos. Nas duas situações o objetivo é restringir o conjunto de soluções possíveis para o problema da super-resolução. Os métodos regularizados determinísticos consideram que as imagens de baixa resolução capturadas são naturalmente suaves e com acuidade visual limitada, e utilizam esta informação *a priori* para tornar o problema bem-posto. A formulação típica do método regularizado determinístico é mostrada a seguir:

$$\hat{X}_H = \arg \min_{X_H} \sum_{k=1}^p \|Y_{Lk} - SHX_H\|_2^2 + \alpha \|CX_H\|_2^2 \quad (2.5)$$

onde C representa um operador passa-alta, $\|\cdots\|_2$ representa a norma l_2 , α é o multiplicador de lagrange do parâmetro de regularização, que controla os dados de entrada representados pelo termo $\|Y_{Lk} - SHX_H\|_2^2$ e a suavidade da solução representada pelo termo $\|CX_H\|_2^2$. Os valores elevados de α são utilizados quando há um número reduzido de imagens disponíveis, pois o problema torna-se indeterminado devido aos erros na aquisição e o ruído. Quando se tem uma grande quantidade de imagens de baixa resolução disponíveis e ruído reduzido, podem ser utilizados valores menores de α para estabilizar a solução. A função custo da equação (2.5) pode ser resolvida utilizando a técnica iterativa a seguir:

$$\hat{X}_H^{n+1} = \hat{X}_H^n + \beta \left[\sum_{k=1}^p (SH)^T (Y_{Lk} - SHX_H^n) - \alpha C^T CX_H^n \right] \quad (2.6)$$

em que β representa um parâmetro de convergência da técnica iterativa.

Os métodos regularizados estatísticos, também conhecidos na literatura como métodos bayesianos, são utilizados quando a função densidade de probabilidade (*PDF*) *a posteriori* pode ser estabelecida (ZIBETTI, 2007; SHEN, 2007). O estimador de máxima probabilidade *a posteriori* (*MAP*) que maximiza a função de densidade de probabilidade, com relação à imagem de alta resolução X_H pode ser expresso por:

$$\hat{X}_H = \arg \max_{X_H} \{P(X_H/Y_{L1}, Y_{L2}, \dots, Y_{LP})\} \quad (2.7)$$

Considerando o teorema de Bayes (PAPOULIS, 1991), onde $p(x|y) = \frac{P(y|x)P(x)}{P(y)}$, e aplicando na equação (2.7) têm-se:

$$\hat{X}_H = \arg \max_{X_H} \left\{ \frac{P(Y_{Li}/X_H)P(X_H)}{P(Y_{Li})} \right\} \quad i = 1, 2, \dots, p \quad (2.8)$$

Como as imagens de baixa resolução são conhecidas, a probabilidade *a priori* $P(Y_{Li})$ na equação (2.8) não influencia na maximização e pode ser eliminada. Aplicando o logaritmo na equação (2.8), temos:

$$\hat{X}_H = \arg \max_{X_H} \{P(Y_{Li}/X_H)P(X_H)\} \quad (2.9)$$

$$\hat{X}_H = \arg \max_{X_H} \{\ln P(Y_{Li}/X_H) + \ln P(X_H)\} \quad (2.10)$$

Observa-se na equação (2.10), que o modelo da imagem $P(X_H)$ e a função densidade condicional $P(Y_{Li}|X_H)$ são definidos considerando a imagem de alta resolução X_H . Os métodos estatísticos são flexíveis quanto ao modelamento do ruído presente no processo de aquisição e ao modelamento das informações *a priori* das imagens adquiridas.

2.4.5 Métodos baseados em Exemplos

Os métodos de super-resolução baseados em exemplos buscam reconstruir uma imagem de alta resolução a partir de características extraídas de um conjunto de treinamento formado por imagens de alta e baixa resolução. O conjunto de imagens utilizadas para aumentar a resolução pode ser diferentes fotos de uma mesma cena, diferentes quadros de um mesmo vídeo, ou mesmo um conjunto de imagens que não guardam nenhuma relação com a imagem de baixa resolução observada (FREEMAN et al., 2002; DATSENKO; ELAD, 2007; BARRETO, 2007; HUNG, 2012).

O uso de imagens completas do conjunto de treinamento para o processo de super-resolução pode resultar em um aumento considerável na complexidade computacional do algoritmo, e consequentemente, na impossibilidade de utilização de um conjunto de treinamento com muitas imagens. Para melhorar a eficiência computacional, os algoritmos trabalham com pequenos blocos retangulares de imagem ou fragmentos (do inglês, *Patches*), que implica num procedimento realizado localmente entre os pixels. A Figura 2.8 ilustra de método típico de super-resolução baseado em exemplos.

Assim, seja um conjunto de treinamento formado por imagens de alta e baixa resolução, representado por $\{x_k, y_k\}_{k=1}^K$, as imagens do conjunto são subdivididas em pequenos blocos retangulares sobrepostos. O resultado da subdivisão produz dois conjuntos de fragmentos, sendo o de baixa resolução $P_L = \{p_{Lk}\}_{k=1}^{K_s}$, de tamanho $n \times n$, e seu equivalente em alta resolução $P_H = \{p_{Hk}\}_{k=1}^{K_s}$, de tamanho $m \times m$, com $K_s \gg K$. É importante ressaltar que o tamanho dos fragmentos influencia no desempenho do algoritmo de super-resolução. A escolha por fragmentos menores podem causar uma redundância desnecessária no conjunto de treinamento e aumentar sua complexidade computacional. Por outro lado, a escolha por fragmentos maiores pode causar artefatos na imagem reconstruída devido aos erros no processo de regularização.

Figura 2.8 - Ilustração de um método típico de super-resolução baseado em exemplos

Fonte: Autor

Na reconstrução, a imagem de baixa resolução capturada também é subdividida em pequenos fragmentos de imagem p_L^i para cada posição i da imagem. Em seguida, o algoritmo faz uma busca por fragmentos próximos considerando o conjunto $\{P_L\} = \{p_{Lk}\}_{k=1}^{K_s}$, de acordo com a seguinte expressão:

$$\|p_L^i - p_{Lk}\|_2^2 \leq T \quad (2.11)$$

onde o limiar T depende do tamanho do fragmento e da variância do ruído considerados. Assim, o algoritmo encontra um subconjunto de fragmentos de baixa resolução correlacionados com p_L^i , denotado por $P_{L[i,j]} = \{p_{Lk}^{[i,j]}\}_{k \in \Omega[i,j]}$ e o subconjunto equivalente em alta resolução $P_{H[i,j]} = \{p_{Hk}^{[i,j]}\}_{k \in \Omega[i,j]}$. Considerando então que todos os fragmentos pertencentes ao subconjunto $P_{H[i,j]}$ são candidatos ao fragmento de alta resolução da imagem final, a tarefa do algoritmo de super-resolução é encontrar o mais correlacionado.

Nos últimos anos várias estratégias foram propostas para reconstrução de imagens de alta resolução a partir de exemplos. Um dos primeiros trabalhos apresentados na literatura foi proposto por (FREEMAN; JONES; PASZTOR, 2002), que adotou uma estratégia de treinamento e reconstrução usando uma Rede de Markov Paramétrica (do inglês, *Parametric Markov Network*). Em (QIANG; XIAOOU; SHUM, 2005), foi proposto um modelo probabilístico que combina uma restrição de reconstrução global com um algoritmo adaptativo de localização de fragmentos similares. Em (HONG; DIT-YAN E YIMIN, 2004), foi proposto um algoritmo baseado na busca de múltiplos fragmentos vizinhos de alta resolução sobrepostos usando um método denominado LLE (do inglês, *Locally Linear Embedding*). Em (JIAN et al., 2003), os autores propuseram uma estratégia de reconstrução a partir do método *Primal Sketch Prior*, considerando as primitivas das imagens de alta e baixa resolução do conjunto de treinamento. Outras abordagens usando as transformadas Wavelet e Contourlet, também foram propostas em (JIJI; JOSHI; CHAUDHURI, 2004; JIJI; CHAUDHURI, 2006).

2.5 Métodos de Super-Resolução Baseados em Representação Esparsa de Sinais

Os modelos baseados em representação esparsa têm sido utilizados em diversas classes de sinais como áudio, imagens e vídeos, para tarefas de aquisição, compressão, visão computacional e reconhecimento de padrões (AHARON; ELAD; BRUCKSTEIN, 2006; WRIGHT et al., 2010; DONOHO, 2006).

A representação esparsa considera um sinal como uma combinação linear esparsa de átomos a partir de um dicionário redundante. Seja um sinal $x \in \mathcal{R}^n$ e um dicionário $D \in \mathcal{R}^{n \times K}$ que contém k átomos do sinal em cada coluna, o modelo de representação esparsa pode ser definido como uma aproximação na forma:

$$\min_{\alpha} \|\alpha\|_0 \quad \text{sujeito a} \quad \|x - D\alpha\|_2 \leq \epsilon \quad (2.12)$$

Na equação (2.12), α representa o vetor esparso do sinal x e $\|\cdots\|_0$ é a norma l^0 que conta o número de elementos diferentes de zero do vetor esparso α . O dicionário D pode ser construído a partir de funções pré-definidas como Fourier, Wavelets ou Curvelets, ou construído a partir do treinamento de um conjunto de amostras do sinal x , (AHARON, ELAD; BRUCKSTEIN, 2006).

No contexto da super-resolução de imagens, a representação esparsa utiliza o conjunto de fragmentos da imagem para construção do vetor esparso α e dos dicionários de baixa e alta resolução. Seja um fragmento da imagem de baixa resolução, definido como $p_L^i = R_i \cdot Y_L$, a combinação esparsa em relação a um dicionário de baixa resolução D_L , é expressa na forma:

$$\arg \min_{\alpha} \|\alpha\|_0 \quad \text{sujeito a} \quad \|R_i \cdot Y_L - D_L \alpha\|_2^2 \leq \varepsilon \quad (2.13)$$

onde R_i representa a operação de extração dos fragmentos da imagem na posição i e ε está relacionado ao ruído presente no processo.

Uma vantagem importante do método de representação esparsa é a possibilidade de utilizar um conjunto reduzido de imagens se comparado com outros métodos tradicionais baseados em exemplos.

2.6 Algoritmo para Treinamento de Dicionários *k*-SVD

O algoritmo de treinamento de dicionários *k*-SVD, considerado uma generalização do algoritmo *k-means*, tem sido muito utilizado devido à flexibilidade e possibilidade de trabalhar em conjunto com qualquer algoritmo de perseguição (do inglês, *pursuit algorithm*) (AHARON, ELAD; BRUCKSTEIN, 2006). A implementação do algoritmo é realizada iterativamente em dois estágios: um estágio de codificação esparsa dos sinais e outro de atualização de um dicionário. O algoritmo *k*-SVD também possui dois modos de operação: um modo baseado em esparsidade e outro baseado em

erro. O algoritmo baseado em erro procura resolver o problema de otimização proposto de acordo com as equações (2.12) e (2.13). O algoritmo baseado em esparsidade procura resolver o problema de otimização proposto de acordo com a seguinte equação:

$$\min_{D,X} \{\|Y - DX\|_2^2\} \quad \text{sujeito a} \quad \forall_i \|x_i\|_0 \leq T_0 \quad (2.14)$$

onde X é uma matriz que contém todos os vetores esparsos e T_0 é um parâmetro que determina o número de elementos diferentes de zero do vetor esparsos x .

O objetivo é encontrar um dicionário D que represente da melhor forma possível um conjunto de exemplos $Y = \{y_i\}_{i=1}^N$. O primeiro estágio de codificação esparsa busca calcular os coeficientes de representação esparsa x , considerando um sinal de entrada y . Assumindo um dicionário D fixo, a função penalidade $\|Y - DX\|_2^2$ pode ser reescrita na forma:

$$\|Y - DX\|_2^2 = \sum_{i=1}^N \|y_i - Dx_i\|_2^2 \quad (2.15)$$

O problema de otimização proposto em (2.15) pode agora ser decomposto em N problemas distintos, na forma:

$$\min_{x_i} \{\|y_i - Dx_i\|_2^2\} \quad \text{sujeito a} \quad \forall_i \|x_i\|_0 \leq T_0 \quad \text{para} \quad i = 1, 2, \dots, N \quad (2.16)$$

O problema proposto em (2.16) pode ser resolvido por qualquer método de perseguição encontrado na literatura. O algoritmo OMP (do inglês, *Orthogonal*

Matching Pursuit) é considerando um dos mais eficientes para codificação esparsa e que trabalha muito bem com o algoritmo k-SVD. Basicamente, o algoritmo OMP seleciona a cada iteração um elemento do vetor esparsa mais correlacionado com um vetor esparsa residual obtido em uma iteração anterior, considerando o problema proposto pela equação (2.16).

O segundo estágio do processo consiste na atualização do dicionário D e dos coeficientes da matriz esparsa X . Assumindo X e D fixos, o algoritmo realiza a atualização por iteração de uma coluna d_k do dicionário D , e dos coeficientes esparsos da k -ésima linha da matriz X , representada por x_k . A partir da equação (2.16), o termo de penalidade pode ser reescrito da seguinte forma:

$$\begin{aligned} \|Y - DX\|_2^2 &= \left\| Y - \sum_{j=1}^K d_j x_T^j \right\|_2^2 = \left\| \left(Y - \sum_{j \neq k} d_j x_T^j \right) - d_k x_T^k \right\|_2^2 \\ &= \|E_k - d_k x_T^k\|_2^2 \end{aligned} \quad (2.17)$$

onde E_k representa uma matriz de erros do conjunto de amostras Y quando um elemento é removido na iteração. Assim, a equação (2.17) pode ser resolvida usando decomposição de valor singular (SVD). Uma visão geral do processo de codificação esparsa e treinamento dos dicionários com os algoritmos k-SVD e OMP é apresentada na figura 2.9.

Figura 2. 9 - Ilustração geral dos algoritmos k-SVD e OMP. O dicionário D na parte superior da figura é inicialmente utilizado e posteriormente são realizadas as etapas iterativas de alternadas de codificação esparsa e atualização do dicionário.

Fonte: (Aharon, Elad e Bruckstein, 2006)

Conforme será apresentado nos capítulos seguintes, os algoritmos k-SVD e OMP foram escolhidos para os estágios de codificação esparsa e treinamento de dicionários por serem eficientes computacionalmente, quando aplicados em super-resolução de imagens.

2.7 Transformada Discreta de Cossenos - DCT

A transformada discreta de cossenos (do inglês, *Discrete Cosine Transform-DCT*) foi apresentada inicialmente na década de 70 e impulsionou as aplicações em processamento de imagens, principalmente nas áreas de compressão, filtragem e extração de características. Considerada uma transformada sub-ótima em relação à Transformada KLT (*Karhunen-Loève Transform*), a transformada DCT possui como característica a geração de coeficientes descorrelacionados, concentrando a maior parte da energia do sinal em um reduzido número de coeficientes. A transformada DCT é definida em quatro tipos: DCT- tipo I, DCT-tipo II, DCT-tipo III e DCT-tipo IV.

A transformada DCT- tipo II é a mais conhecida e utilizada em processamento de imagens devido à sua maior capacidade de compactação de energia (RAO; YIP, 1990; MATOS, 2008; PEDRINI; SCHWARTZ, 2008; GONZALEZ; WOODS, 2010).

Considerando um sinal original $f(x, y)$, a transformada DCT tipo II é representada por:

$$F(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left[\frac{(2x+1)u\pi}{2N} \right] \cos \left[\frac{(2y+1)v\pi}{2N} \right] \quad (2.18)$$

sendo:

$$\alpha(x) = \begin{cases} \sqrt{\frac{1}{N}} & \text{para } x = 0 \\ \sqrt{\frac{2}{N}} & \text{para } x = 1, 2, \dots, N-1 \end{cases} \quad (2.19)$$

onde N representa o tamanho do bloco. Um exemplo ilustrando a aplicação da transformada DCT para um bloco de tamanho (8×8) pixels é mostrada na Figura 2.10.

Os coeficientes obtidos após a aplicação da transformada refletem a importância das frequências presentes no sinal original. Considerando a figura 2.10, os primeiros coeficientes a partir do canto superior esquerdo, referem-se às frequências mais baixas e representam em geral, a maior parte da informação da imagem original. Os últimos coeficientes, localizados no canto inferior direito do bloco transformado, referem-se às frequências mais altas, que geralmente representam as informações mais detalhadas ou finas da imagem, ou que em muitos casos representam predominantemente ruído.

Figura 2.10 - Representação gráfica do núcleo da transformada discreta do cosseno para um bloco de tamanho (8×8) .

Fonte: (Pedrini e Schwartz, 2008).

A maioria dos sistemas de codificação por transformada se baseia na DCT, devido ao melhor custo-benefício entre a capacidade de compressão e o custo computacional. A transformada DCT se tornou o padrão internacional para sistemas de codificação por transformada e compressão de imagens (SULLIVAN; WIEGAND, 2005; GONZALEZ; WOODS, 2010), e também tem sido utilizada com sucesso em outras áreas de processamento de imagem como interpolação, deblurring e super-resolução.

2.8 Interpolação no Domínio da Transformada DCT

O processo de interpolação de uma imagem usando a transformada DCT adiciona zeros em regiões de alta frequência de um bloco e em seguida aplica a transformada DCT inversa em todo o bloco, de forma a obter a imagem interpolada.

Para entender melhor o processo, vamos considerar inicialmente a imagem Y_L da equação (2.1) sendo dividida em blocos não sobrepostos de tamanho $n \times n$, onde

$n \ll N_1, N_2$ (tamanho da imagem). Para cada bloco b_L da imagem Y_L é aplicada a transformada DCT do tipo II, resultando um bloco transformado $B_{L(DCT)}$:

$$B_{L(DCT)} = DCT\{b_L\} \quad (2.20)$$

cada bloco $B_{L(DCT)}$ é redimensionado de acordo com um fator de ampliação do tamanho da imagem s , acrescentando zeros nas regiões de alta frequência do bloco DCT:

$$\hat{B}_{L(DCT)} = \begin{bmatrix} \{B_{L(DCT)}\}_{n \times n} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{sn \times sn} \quad (2.21)$$

Em seguida, para cada bloco $\hat{B}_{L(DCT)}$ é aplicada a transformada DCT inversa resultando em um bloco interpolado $\hat{b}_L$ de tamanho $sn \times sn$:

$$\hat{b}_L = IDCT\{\hat{B}_{L(DCT)}\} \quad (2.22)$$

A Figura 2.11 ilustra um exemplo de aplicação da transformada DCT para o caso de $s = 2$. O processo de interpolação tem início com a aplicação da transformada DCT em um bloco de (4×4) pixels. Em seguida, os coeficientes de valor zero são adicionados a cada bloco, formando um bloco de (8×8) . A transformada DCT inversa é aplicada em todos os blocos (8×8) resultando na imagem interpolada.

Figura 2.11 – Ilustração do Método de Interpolação do Domínio da Transformada DCT para um bloco (8x8) pixels da imagem de baixa resolução.

Fonte: Autor

O processo de interpolação usando a transformada DCT tem apresentado resultados superiores em relação a técnicas de interpolação no domínio espacial, como por exemplo, as interpolações bicúbica e bilinear (HUNG, 2011; ZHANG; CHAM, 2011). Entretanto, a interpolação no domínio da Transformada DCT apresenta como desvantagem os artefatos de blocos na imagem final, principalmente em regiões de contornos e textura, que pioram a qualidade visual da imagem interpolada. Conforme será visto no capítulo 5, o método proposto nesta tese tem como um dos objetivos diminuir estes efeitos de blocos no processo de super-resolução.

Capítulo 3

Trabalhos Relacionados

Este capítulo apresenta uma revisão dos principais trabalhos relacionados ao tema proposto nesta tese. Alguns trabalhos abordam técnicas de super-resolução de imagens utilizando a representação esparsa de sinais e outros descrevem as principais estratégias utilizadas para interpolação e super-resolução de imagens usando a transformada discreta de cossenos (DCT). Também é realizada uma análise dos trabalhos buscando correlacionar e contextualizar os trabalhos com a proposta desta tese, considerando as diferentes estratégias adotadas para melhorar a qualidade na reconstrução e redução do custo computacional.

3.1 Introdução

Os métodos de super-resolução que utilizam o conceito de representação esparsa de sinais buscam encontrar uma relação linear entre os fragmentos das imagens de baixa e alta resolução do conjunto de treinamento em relação a um dicionário, considerando princípios de esparsidade. Com a utilização de um dicionário compacto, os algoritmos de super-resolução conseguem reduzir significativamente o custo computacional das etapas de treinamento e reconstrução da imagem de alta resolução.

Conforme já abordado no item 2.7, o uso da Transformada DCT é bem conhecido na compressão de imagens dos principais padrões como MPEG-2, MPEG-4, H.264/AVC (CHRISTOPOULOS; SKODRAS; EBRAHIMI, 2000; WIEGAND et al., 2003). A sua utilização no processo de super-resolução de imagens e vídeos é algo recente na literatura. O interesse na utilização da transformada DCT se deve principalmente à capacidade de preservar os coeficientes de baixa frequência da imagem sem afetar consideravelmente a qualidade final da imagem reconstruída.

Os trabalhos que serão apresentados a seguir abordam diversas estratégias de reconstrução de imagens de alta resolução que buscam melhorar a eficiência nas diversas etapas do processo como: classificação dos fragmentos, treinamento de

dicionários de acordo com as características estruturais dos fragmentos, treinamento de dicionários acoplados, utilização de algoritmos de otimização eficientes e interpolação e reconstrução dos coeficientes de alta frequência por meio da transformada DCT.

3.2 Principais Trabalhos Usando Métodos de Representação Esparsa de Sinais

Em (YANG et al., 2010) é proposta uma estratégia de treinamento conjunto de dicionários a partir da extração de fragmentos de imagens de baixa e alta resolução de um conjunto de treinamento. No algoritmo proposto, as etapas de treinamento dos dicionários e codificação esparsa são realizadas com os algoritmos *LASSO* (do inglês, *Least Absolute Shrinkage and Selection Operator*) e OMP. Segundo os autores, o método é robusto ao ruído presente no processo de aquisição da imagem de baixa resolução e pode realizar o processo de “*deblurring*” e super-resolução simultaneamente. Uma das desvantagens do método é o aumento no custo computacional com o aumento do tamanho dos dicionários utilizados na etapa de treinamento.

O trabalho de (WANG; ZHU; GONG, 2010) apresenta um método de super-resolução de imagens baseado na aprendizagem por associação esparsa entre os fragmentos da imagem de baixa resolução capturada e os fragmentos de imagem do conjunto de treinamento. O método proposto utiliza o algoritmo modificado de busca por características (do inglês, *Feature-sign search*) para obter uma convergência rápida na etapa de codificação esparsa. Entretanto, o desempenho do algoritmo depende da qualidade dos fragmentos extraídos e da sua associação com vetores-bases pertencentes a um dicionário redundante.

O trabalho de (DONG et al., 2011) apresenta uma proposta de super-resolução e *deblurring* a partir da seleção de fragmentos de baixa e alta resolução do conjunto de treinamento utilizando um algoritmo de domínio esparsa adaptativo (do inglês, *Adaptive Sparse Domain Selection-ASDS*). A etapa de treinamento dos dicionários é realizada a partir de uma subdivisão dos fragmentos da imagem de acordo com o seu conteúdo estrutural. Na reconstrução são propostos dois novos termos de

regularização para melhorar a qualidade da imagem final. O primeiro termo é baseado em modelos autoregressivos e o segundo termo utiliza médias não locais para melhorar a classificação dos fragmentos e eliminar possíveis artefatos gerados no processo de reconstrução. Os resultados demonstraram que a estratégia é eficiente na redução do efeito de “*deblurring*” das imagens, mas ainda insatisfatórios no aspecto de qualidade da imagem super-resolvida.

O trabalho de (ZEYDE; ELAD; PROTTER, 2012) apresenta uma proposta de super-resolução de imagem única implementada em duas etapas: uma etapa de treinamento e construção de dicionários e uma etapa de reconstrução da imagem de alta resolução. A partir de um conjunto de treinamento composto por imagens de alta e baixa resolução, os algoritmos *K-SVD* e *OMP* são utilizados para o treinamento dos dicionários e geração dos vetores esparsos. Na etapa de reconstrução, a imagem de baixa resolução capturada é interpolada bicubicamente e passa por um processo de filtragem passa-alta utilizando os filtros *Laplaciano* e *Gradiente* para extração de características. No trabalho ainda são apresentadas outras contribuições com o objetivo de reduzir o tempo computacional do algoritmo como: construção do dicionário de alta resolução por meio de uma expressão pseudo-inversa e redução da dimensionalidade dos fragmentos após a extração de características com os filtros passa-alta. Os resultados computacionais demonstraram a efetividade do método para fatores de ampliação da imagem maiores. Considerando os resultados relevantes apresentados, este método foi utilizado como parâmetro de comparação com o método proposto nesta tese.

Em (YANG et al., 2012) é apresentada uma estratégia de treinamento e construção de múltiplos dicionários a partir da extração e agrupamento de fragmentos de imagens considerando a sua orientação geométrica. Os fragmentos extraídos do conjunto são agrupados em três diferentes grupos: fragmentos suaves, fragmentos com orientação dominante e fragmentos estocásticos. Os algoritmos de codificação esparsa e treinamento de dicionários são aplicados para os fragmentos que representam regiões de borda, contornos e texturas das imagens. Para os demais fragmentos de regiões planas são aplicados os algoritmos tradicionais de interpolação tradicionais, sem comprometer a qualidade da imagem final. Na reconstrução da imagem de alta resolução utiliza-se uma estratégia de agregação

de fragmentos baseada na similaridade não local. Além disso, é proposto um esquema de compensação residual para melhorar a qualidade da imagem de alta resolução final na etapa de reconstrução.

O trabalho de (ZHOU; YANG; LIAO, 2012) apresenta um método para super-resolução de imagens usando um conceito de treinamento de subdicionários incoerentes. A primeira etapa do processo consiste na subdivisão dos fragmentos de alta e baixa resolução em subconjuntos de mesma característica estrutural usando um descritor local de Weber (do inglês, *Weber Local Descriptor-WLD*). Em seguida, são treinados sub-dicionários para cada conjunto formado. Na reconstrução, os fragmentos de entrada são super-resolvidos, utilizando os sub-dicionários de treinamento.

Em (JIA; WANG; TANG, 2013) é apresentada uma estratégia de treinamento de dicionários acoplados para diversas aplicações envolvendo transformação de imagens. O acoplamento é realizado forçando os coeficientes esparsos dos fragmentos de alta e baixa resolução do conjunto de treinamento a ter o mesmo número de elementos diferentes de zero. O processo de codificação esparsa com acoplamento é realizado usando um algoritmo BCGD (do inglês, *block coordinate gradient descent*) e os vetores esparsos obtidos são agrupados em diferentes subconjuntos. Na reconstrução é utilizada uma estratégia de regressão paramétrica local, considerando os fragmentos esparsos da imagem de baixa resolução capturada e os subconjuntos de fragmentos. Os resultados do algoritmo para super-resolução apresentaram uma leve melhora quando comparado com outras propostas de super-resolução da literatura. Entretanto, alguns aspectos relacionados à eficiência computacional do algoritmo ainda permanecem em aberto.

3.3 Principais Trabalhos de Super-Resolução Usando a Transformada DCT

Em (WU; YU; CHEN, 2010) é apresentado um método de interpolação de quadros de baixa resolução de um vídeo utilizando um esquema híbrido de interpolação no domínio da transformada DCT e a filtragem de Wiener. Os coeficientes de baixa frequência dos quadros são obtidos pela transformada DCT e os coeficientes de alta

frequência são obtidos por filtragem de Wiener. O filtro de Wiener utilizado é o mesmo utilizado no padrão H.264/AVC. Uma das vantagens do método proposto é a preservação dos coeficientes de baixa frequência por meio da transformada DCT, o que melhora a qualidade da imagem final ampliada.

Em (KWOK-WAI; WAN-CHI, 2013) também é proposto um esquema híbrido de interpolação de quadros de vídeo por meio da transformada DCT e filtros de Wiener. Diferente da proposta anterior, que usava o filtro de Wiener com coeficientes fixos, o mesmo utilizado no padrão H.264/AVC, neste trabalho os coeficientes do filtro são calculados de acordo com as imagens presentes no conjunto de treinamento. Os resultados demonstram uma leve superioridade do método.

O trabalho de (SANG-JUN; JECHANG, 2010) apresenta uma estratégia de redimensionamento da imagem usando uma representação do bloco livre da transformada inversa DCT. O objetivo é diminuir a complexidade computacional do algoritmo sem a utilização da transformada inversa DCT para o retorno ao domínio espacial. Os resultados demonstraram uma melhoria na qualidade da imagem, principalmente na reconstrução das bordas horizontais e verticais presentes na imagem.

O trabalho de (GARCIA; DOREA; QUEIROZ, 2012) apresenta um método de super-resolução para aplicações de vídeos de resolução mista. O método usa o conceito de interpolação no domínio da transformada DCT, a partir de múltiplas vistas obtidas de uma mesma cena. A técnica proposta recupera o conteúdo de alta frequência dos quadros de baixa resolução, utilizando o conteúdo de alta frequência de imagens de alta resolução vizinhas, também denominadas quadros-chave. O método proposto preserva os componentes de baixa frequência da imagem e complementa a imagem final com os coeficientes de alta frequência obtidos das múltiplas vistas projetadas.

O trabalho de (HUNG, 2011) apresenta uma proposta de super-resolução de quadros de baixa resolução em vídeos de resolução mista, usando a transformada DCT. Os coeficientes de alta frequência dos quadros de baixa resolução são recuperados utilizando os quadros adjacentes de alta resolução (Quadros-chave). A interpolação baseada na transformada DCT preserva os componentes de baixa

frequência da imagem no processo de subamostragem e interpolação, porém, a imagem interpolada pode gerar o efeito de blocos devido ao acréscimo de coeficientes de valor zero no bloco transformado.

Em (ZHANG; CHAM, 2011) é apresentada uma proposta de super-resolução de rostos baseada numa estratégia de treinamento e inferência de coeficientes de alta frequência no domínio da transformada DCT. Os Coeficientes de baixa frequência podem ser estimados por um processo mais simples de interpolação, enquanto os demais componentes de alta frequência são estimados usando um modelo probabilístico com blocos adjacentes correlacionados. O método proposto é dividido em duas etapas: uma primeira de pré-filtragem para separação dos coeficientes de baixa e alta frequência e uma segunda de reconstrução no domínio da transformada DCT a partir da imagem pré-filtrada.

3.4 Análise das Propostas Estudadas

Esta seção faz uma análise dos algoritmos apresentados nas seções anteriores a partir das diferentes estratégias adotadas para a melhoria do processo de super-resolução. Foram adotados cinco parâmetros para análise a seguir:

- **Agrupamento de fragmentos das imagens:** um aspecto importante dos métodos de representação esparsa é a correlação entre os vetores esparsos gerados a partir dos fragmentos de baixa e alta resolução com os dicionários. Conforme pôde ser observado, alguns trabalhos buscaram resolver este problema através de diferentes estratégias de agrupamento de fragmentos de acordo com a sua orientação geométrica. Esta solução melhora o processo de construção dos dicionários e consequentemente, a reconstrução da imagem de alta resolução final.
- **Treinamento dos dicionários:** um dicionário compacto e que possa representar uma classe específica ou uma ampla classe de imagens é importante para melhorar a qualidade da imagem de alta resolução. Os trabalhos estudados apresentaram estratégias de construção de dicionários de baixa e alta resolução de forma simultânea, como em (YANG et al., 2010), ou

mesmo utilizando métodos indiretos de construção como em (ZEYDE; ELAD; PROTTER, 2012). Alguns trabalhos adotaram o treinamento e construção de múltiplos dicionários a partir dos fragmentos agrupados de mesma característica estrutural, como em (DONG et al., 2011; ZHOU; YANG; LIAO, 2012). Em (JIA; WANG; TANG, 2013) é apresentado um conceito de treinamento de dicionários acoplados, considerando a utilização dos fragmentos de baixa e alta resolução conjuntamente.

- **Utilização de algoritmos mais eficientes:** a utilização de algoritmos de otimização mais eficientes tanto para a etapa de treinamento quanto para a reconstrução da imagem de alta resolução tendem a reduzir o custo computacional do processo de super-resolução. Os algoritmos citados e utilizados para o treinamento de dicionários e codificação esparsa foram: k-SVD, LASSO, OMP, BCGD. Entretanto, podemos observar que os algoritmos k-SVD e OMP são os mais populares e eficientes algoritmos utilizados para o treinamento e codificação esparsa.
- **Implementação de novos termos de regularização:** para melhorar a qualidade das imagens e acelerar o processo de reconstrução, alguns trabalhos propuseram incluir novos termos de regularização nos algoritmos. Dentre eles podemos citar os Modelos autoregressivos (DONG et al., 2011), e redes neurais em (YANG et al., 2012).
- **Utilização da transformada DCT:** alguns trabalhos analisados propuseram estratégias de interpolação e super-resolução no domínio da transformada DCT. A utilização da transformada DCT busca melhorar a qualidade das imagens interpolação preservando os coeficientes de baixa frequência do bloco transformado.

Apesar dos trabalhos terem apresentado importantes contribuições, foi possível identificar algumas estratégias ainda não utilizadas para a melhoria da qualidade da imagem final e redução do custo computacional, como:

- Possibilidade de utilização de dicionários mais compactos de forma a diminuir o tempo computacional do algoritmo na etapa de treinamento;

- Melhoria no processo de extração de características dos fragmentos de alta e baixa resolução utilizando diferentes filtros passa-alta;
- Possibilidade de utilização de estratégias de ampliação da imagem de forma escalonável e iterativa para melhorar a qualidade da imagem final;
- Utilização do processo de interpolação DCT para melhorar a qualidade da imagem, preservando os coeficientes de baixa frequência;

Desta forma, o método proposto nesta tese busca contemplar as estratégias acima identificadas num método de super-resolução para imagem única.

3.5 Conclusões do Capítulo

Este capítulo fez uma revisão e análise dos principais trabalhos de super-resolução de imagens usando representação esparsa de sinais. Também foram revisados e analisados os principais trabalhos de interpolação e super-resolução usando a transformada DCT.

Com base na análise realizada foi possível identificar algumas estratégias ainda não adotadas nos principais trabalhos relacionados, de forma a melhorar o processo de treinamento e também a etapa de reconstrução das imagens. Estas contribuições serão apresentadas nos capítulos seguintes.

Capítulo 4

Super-Resolução de Imagem Usando Interpolação DCT e o Modelo de Representação Esparsa

Este capítulo apresenta uma proposta para super-resolução de imagem única que combina a interpolação no domínio da transformada DCT e o método de aprendizagem baseado na representação esparsa de sinais. O método proposto é dividido em duas principais etapas: uma primeira de treinamento e construção de dois dicionários de alta e baixa resolução, e uma segunda etapa de reconstrução da imagem final de alta resolução. As etapas são descritas em detalhes ao longo do capítulo, mostrando as vantagens da utilização destes dois métodos combinados. No final do capítulo são realizados alguns experimentos computacionais com o objetivo de avaliar o método proposto em relação a outros métodos tradicionais de super-resolução, utilizando as principais métricas objetivas e subjetivas para análise. Essas contribuições foram publicadas em (REIS; BRESSAN, 2013a; REIS; BRESSAN, 2013b, REIS; BRESSAN, 2013c).

4.1 Introdução

Conforme abordado no capítulo 3, a interpolação de imagens utilizando a transformada DCT tem sido amplamente estudada na literatura nos últimos anos (HUNG, 2011), (ZHANG; CHAM, 2011), (SUN; CHAM, 2007; WU; YU; CHEN, 2010), (SUN; CHAM, 2007; WU; YU; CHEN, 2010). A melhoria na qualidade da imagem final interpolada com a transformada DCT está relacionada à preservação das componentes de baixa frequência dentro de um bloco transformado.

Os métodos de super-resolução baseados em representação esparsa de sinais buscam recuperar os componentes de alta frequência da imagem por meio de um processo de treinamento e construção de um dicionário que represente todas as informações contidas no conjunto de treinamento. Os resultados obtidos com estes métodos tem se mostrado superiores a outros métodos de super-resolução baseados em exemplos, e a métodos de super-resolução tradicionais baseados em reconstrução com múltiplas imagens. Entretanto, o tempo computacional gasto para

o treinamento e a construção de um dicionário, e a necessidade de utilizar um conjunto de treinamento com muitas imagens, tem limitado as aplicações destes métodos.

O método proposto neste capítulo explora as vantagens da interpolação DCT e dos métodos de super-resolução baseados em representação esparsa, no processo de reconstrução da imagem de alta resolução.

Considerando uma imagem de baixa resolução capturada, uma imagem intermediária é construída representando os coeficientes de baixa frequência. Os componentes de alta frequência não presentes são obtidos por meio da utilização do algoritmo de representação esparsa. A Figura 4.1 ilustra o método proposto.

Nas seções seguintes são descritas as etapas de treinamento dos dicionários e a etapa de reconstrução da imagem de alta resolução final, a partir da imagem interpolada no domínio DCT e da imagem obtida pelo método de representação esparsa.

Figura 4.1 - Diagrama Geral do Método de Super-Resolução Proposto.

Fonte: Autor

4.2 Treinamento e construção dos dicionários usando o método de Representação esparsa de sinais

Esta seção descreve a etapa de treinamento e construção dos dicionários de baixa e alta resolução do método proposto utilizando o modelo de representação esparsa de sinais.

4.2.1 Construção do Conjunto de imagens de alta e baixa resolução

Os algoritmos de super-resolução baseados em exemplos utilizam para o treinamento um conjunto de imagens de alta resolução e um conjunto equivalente de imagens de baixa resolução. As imagens utilizadas para composição destes dois conjuntos podem estar correlacionadas com a imagem de baixa resolução capturada (*Imagem LR*), ou podem conter características distintas. O método proposto neste capítulo para super-resolução pode utilizar as duas formas, porém neste capítulo, os experimentos serão conduzidos usando a segunda forma.

Uma vez escolhidas as imagens de alta resolução para composição do conjunto de treinamento, é formado um conjunto de imagens de baixa resolução equivalente, por meio de um processo de degradação das imagens, de acordo com o modelo apresentado na equação (2.1). O modelo de observação contém os efeitos de borramento (*Blurring*) e subamostragem (*downsampling*), conforme ilustrado na Figura 4.2.

Figura 4.2 - Construção do conjunto de treinamento formado por imagens de alta resolução e seu equivalente de baixa resolução.

Fonte: Autor

4.2.2 Extração de Características a partir dos Fragmentos da Imagem

O conjunto de treinamento de alta resolução e seu equivalente em baixa resolução podem conter imagens de conteúdo complexo e distinto. Para reduzir o custo computacional no processo de super-resolução, as imagens são subdivididas em pequenos fragmentos ou blocos retangulares (do inglês, *patches*). Cada fragmento representa um conjunto de pixels extraídos na vizinhança de uma posição i da imagem, conforme apresentado na Figura 4.3.

Figura 4.3 - Representação dos fragmentos (patches) de alta e baixa resolução da imagem.

Fonte: Autor

O processo de extração de características busca obter o máximo de detalhes das imagens como bordas, contornos ou texturas, por meio da utilização de filtros passa-alta. Muitos algoritmos de super-resolução propostos na literatura tem usado com sucesso os filtros de aguçamento baseados em derivadas de primeira e segunda ordem, também denominados *Gradiente* e o *Laplaciano* (FREEMAN; PASZTOR; CARMICHAEL, 2000; HONG; DIT-YAN; YIMIN, 2004; YANG et al., 2010; ZEYDE; ELAD; PROTTER, 2012). Os filtros passa-alta Laplaciano e Gradiente são isotrópicos, ou seja, possuem uma resposta independente da direção das descontinuidades da imagem, assim são muito eficientes no processo de extração de características.

No algoritmo proposto foi adotado o mesmo conjunto de filtros Laplaciano e Gradiente, cujos coeficientes são apresentados abaixo:

$$F_1 = [-1, 0, 1]; \quad F_2 = F_1^T \quad (4.1)$$

$$F_3 = [1, 0, -2, 0, 1]; \quad F_4 = F_3^T \quad (4.2)$$

O processo de extração de características por meio da filtragem passa-alta resulta em dois conjuntos de fragmentos de alta e baixa resolução $\tilde{p}_l^i$ e $\tilde{p}_h^i$, de acordo com as seguintes expressões:

$$\tilde{p}_l^i = F_r^i * I_L \quad (4.3)$$

$$\tilde{p}_h^i = F_r^i * I_H \quad (4.4)$$

onde F_r ($r = 1, \dots, 4$) representa o conjunto de filtros passa-alta utilizado, o símbolo “*” representa a operação de convolução, I_L representa a imagem de baixa resolução do conjunto de treinamento de tamanho $N_1 \times N_2$, I_H a imagem de alta resolução do conjunto de treinamento de tamanho $M_1 \times M_2$. O processo de extração de características representados pelas equações (4.3) e (4.4) resulta nos fragmentos de imagem $\tilde{p}_l^i$, e $\tilde{p}_h^i$. É importante ressaltar que os vetores resultantes do processo de extração possuem dimensão maior que os vetores originais p_l^i e p_h^i , de tamanho $L_1 \times L_2$ e $sL_1 \times sL_2$, com $L_1 \ll N_1, M_1$ e $L_2 \ll N_2, M_2$ respectivamente, e s representando o fator de ampliação da imagem.

4.2.3 Redução da Dimensionalidade com o Algoritmo PCA

O uso de um algoritmo de redução da dimensionalidade nos vetores $\tilde{p}_l^i$ e $\tilde{p}_h^i$ é importante para melhorar o desempenho computacional nas etapas de treinamento e reconstrução da imagem de alta resolução.

O método de Análise de Componentes Principais (PCA), (do inglês, *Principal Analysis Components – PCA*) é um método clássico e amplamente utilizado para redução de dimensionalidade. O método consiste na aplicação de transformações

lineares sobre os dados de espaço de maior dimensão projetando em espaços de menor dimensão, de forma que as projeções com maior variância coincidam com os eixos representados pelas componentes principais. A operação tem como objetivo obter um espaço de menor dimensão mantendo ao máximo a informação (JOLLIFFE, 2002).

Considerando os fragmentos de imagem $\tilde{p}_l^i$ e $\tilde{p}_h^i$ obtidos pelas equações (4.3) e (4.4), o algoritmo PCA busca reduzir a dimensionalidade vetor para a dimensão original, mantendo ao máximo a informação adquirida pelo processo de extração de característica (WANG; TANG, 2005; CHAKRABARTI, RAJAGOPALAN; CHELLAPPA, 2007; JIJI; CHAUDHURI; CHATTERJEE, 2007). A operação de redução de dimensionalidade pode ser representada pelas seguintes expressões:

$$p_l^i = B \cdot \tilde{p}_l^i \quad (4.5)$$

$$p_h^i = B \cdot \tilde{p}_h^i \quad (4.6)$$

onde B representa um operador de projeção do fragmento $\tilde{p}_l^i$ em p_l^i e $\tilde{p}_h^i$ em p_h^i , com p_l^i e p_h^i de dimensão reduzida.

4.2.4 Treinamento dos Dicionários com os Algoritmos K-SVD e OMP

Para a construção do dicionário de baixa resolução D_L e dos vetores esparsos α_i que representam os fragmentos de baixa resolução do conjunto de treinamento, são utilizados os algoritmos de treinamento de dicionários k-SVD e codificação esparsa OMP. Estes dois algoritmos trabalham de forma iterativa e se alternam entre um processo de atualização do dicionário D_L e o processo de construção dos vetores esparsos α_i , conforme já abordado na seção 2.5. Considerando o fragmento p_l^i , os algoritmos k-SVD e OMP buscam resolver o seguinte problema de otimização:

$$\arg \min \|p_l^i - D_L \alpha_i\|_2^2 \quad \text{sujeito a} \quad \|\alpha_i\|_0 \leq T_o \quad (4.7)$$

onde $\|\cdots\|_2$ é a norma l_2 (Apêndice A), α_i é o vetor esparsos que representa o fragmento p_l^i na posição i da image e T_o é o parâmetro de esparsidade adotado, ou seja, o número mínimo de elementos diferentes de zero do vetor esparsos.

4.2.5 Construção do Dicionário de Alta Resolução

Para reduzir o tempo de processamento do algoritmo, o dicionário de alta-resolução D_H é obtido aplicando-se uma expressão pseudo-inversa, da forma:

$$D_H = P_H Q^T (Q Q^T)^{-1} \quad (4.8)$$

onde P_H é a matriz que representa os fragmentos extraídos do conjunto de imagens de alta-resolução, e a matriz Q contém todos os vetores esparsos obtidos do processo de codificação esparsa utilizando o algoritmo OMP.

A Figura 4.4 apresenta uma visão geral de todo o processo de treinamento e construção dos dicionários descritos nas seções anteriores.

Figura 4.4 - Visão geral da etapa de treinamento e construção dos dicionários.

Fonte: Autor

4.3 Reconstrução da Imagem de Alta Resolução

Esta seção descreve o processo de reconstrução da imagem de alta-resolução, a partir da imagem de baixa-resolução capturada e do processo de treinamento descrito nas seções anteriores.

4.3.1 Interpolação da Imagem de Baixa Resolução

A imagem de baixa-resolução capturada Y_L , de tamanho $N_1 \times N_2$, é inicialmente interpolada bicubicamente e pela interpolação DCT, conforme as expressões abaixo:

$$Y_{BIC}^{LR} = Q_{BIC} \cdot Y_L \quad (4.9)$$

$$Y_{DCT}^{LR} = Q_{DCT} \cdot Y_L \quad (4.10)$$

onde as imagens Y_{BIC}^{LR} e Y_{DCT}^{LR} , ambas de tamanho $M_1 \times M_2$, com $M_1 > N_1$ e $M_2 > N_2$, representam as imagens interpoladas. Os operadores Q_{BIC} e Q_{DCT} representam as operações de interpolação bícubica e no domínio da transformada DCT.

4.3.2 Filtragem Passa-Alta e Extração dos Fragmentos

A imagem interpolada bicubicamente Y_{BIC}^{LR} , passa por um processo de filtragem passa-alta utilizando filtros Laplaciano e Gradiente das equações (4.1) e (4.2), para extração de fragmentos. O processo de filtragem passa-alta e extração dos fragmentos resulta em um fragmento de imagem de acordo com a equação:

$$\tilde{p}_l^i = F_r^i * Y_{BIC}^{LR} \quad (4.11)$$

onde $\tilde{p}_l^i$ representa um fragmento extraído da imagem Y_{BIC}^{LR} na posição i . Observa-se também aqui que o fragmento $\tilde{p}_l^i$ possui uma dimensionalidade maior em função do conjunto de filtros utilizados.

Novamente é empregada a técnica baseada em análise de Componentes Principais para redução da dimensionalidade:

$$p_l^i = B \cdot \tilde{p}_l^i \quad (4.12)$$

onde B representa o operador de projeção de $\tilde{p}_l^i$ em p_l^i .

4.3.3 Codificação Esparsa dos Fragmentos

Esta etapa tem como objetivo calcular os vetores esparsos que representam os fragmentos resultantes p_l^i extraídos da imagem interpolada bicubicamente. Considerando o dicionário de baixa resolução D_L obtido na etapa de treinamento, e os fragmentos p_l^i da seção anterior, o algoritmo de codificação esparsa *OMP* é aplicado para obtenção dos vetores esparsos α_i para cada posição i da imagem:

$$\min_{\hat{\alpha}_i} \|p_l^i - D_L \alpha_i\|_2^2 \quad \text{sujeito a} \quad \|\alpha_i\|_0 \leq T_0 \quad (4.13)$$

O algoritmo OMP busca encontrar os coeficientes esparsos α_i que minimizam a função $\|p_l^i - D_L \alpha_i\|_2^2$, considerando a restrição dada pelo parâmetro de esparsidade T_0 . Para o algoritmo proposto, foi adotado $T_0 = 3$.

4.3.4 Construção dos Fragmentos de Alta Resolução

Os fragmentos de alta-resolução p_h^i são obtidos a partir do vetor esparsa α_i e do dicionário de alta-resolução D_H da etapa de treinamento conforme a expressão:

$$p_h^i = D_H \alpha_i \quad (4.14)$$

4.3.5 Construção da Imagem com as Componentes de Alta Frequência

A imagem com os componentes de alta frequência do processo de representação esparsa é obtida reunindo os fragmentos de alta resolução $\hat{p}_h^i$ numa solução por mínimos quadrados, como segue:

$$I_{HR} = \left[\sum_i F_r^{iT} F_r^i \right]^{-1} \sum_i F_r^{iT} p_h^i \quad (4.15)$$

onde F_r é o operador passa-alta de extração de características da imagem definido por meio das equações (4.1) e (4.2) da seção 4.2.2.

4.3.6 Construção da Imagem de Alta Resolução Final

A imagem final é obtida pela soma da imagem que representa as componentes de baixa frequência (Y_{DCT}^{LR}) e a imagem que representa as componentes de alta frequência (I_{HR}):

$$I_{HRF} = I_{HR} + Y_{DCT}^{LR} \quad (4.16)$$

4.4 Experimentos Realizados com o Algoritmo Proposto

4.4.1 Parâmetros do Experimento

Para ilustrar o desempenho do método de super-resolução proposto, foram realizadas simulações computacionais em MATLAB. As imagens utilizadas para os testes do algoritmo foram: Lena, Bridge e Man, todas de tamanho (512×512) pixels. As imagens foram degradadas por subamostragem e borramento usando fatores de redução de resolução $s = 2$ e $s = 3$ e depois retornadas para o seu tamanho original.

O conjunto de imagens utilizado para o treinamento e construção dos dicionários D_L e D_H na etapa de treinamento, foi o mesmo utilizado nos experimentos de (YANG et al., 2010; ZEYDE; ELAD; PROTTER, 2012). Este conjunto é composto por imagens de tamanho e conteúdo variados, que incluem paisagens, plantas,

automóveis, pessoas e construções. Algumas das imagens do conjunto de treinamento são mostradas na Figura 4.5.

O conjunto de quatro filtros Laplaciano e gradiente, descritos na seção 4.2.2, foram utilizados para extração das componentes de alta frequência das imagens durante as etapas de treinamento e reconstrução da imagem de alta resolução.

Os algoritmos k-SVD e OMP foram utilizados para o treinamento e construção do dicionário D_L . Considerando que os algoritmos possuem uma convergência rápida, os testes foram realizados para 01 iteração e depois utilizando 20 iterações. Todos os testes foram realizados utilizando dicionários de tamanho igual a 512 átomos.

Figura 4.5 - Algumas imagens do conjunto de treinamento.

Fonte: (YANG et al., 2010; ZEYDE; ELAD; PROTTER, 2012)

4.4.2 Algoritmos Implementados

O algoritmo de super-resolução proposto foi testado e comparado com os métodos de interpolação bicúbica, interpolação DCT, e com o algoritmo proposto em (ZEYDE; ELAD; PROTTER, 2012), adaptados para utilização dos mesmos parâmetros apresentados na seção anterior.

4.4.3 Métricas Utilizadas

Para as avaliações objetivas e subjetivas foram adotadas três medidas: Relação sinal-ruído de pico-PSNR (do inglês, *Peak Signal to Noise Ratio-PSNR*), Similiaridade Estrutural-SSIM (do inglês, *Structural Similarity-SSIM*) e qualidade visual das imagens.

A relação sinal-ruído de pico *PSNR*, expressa em decibéis (*dB*), é uma das métricas mais utilizadas na literatura para avaliar a qualidade de uma imagem ou vídeo. A métrica avalia a diferença global entre a imagem original e a imagem obtida pelo algoritmo, segundo a expressão:

$$PSNR (dB) = 20 \log_{10} \frac{MN(255)^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x,y) - g(x,y)]^2} \quad (4.17)$$

onde $f(x, y)$ representa a imagem original e $g(x, y)$ a imagem obtida pelo algoritmo, ambas de tamanho $M \times N$ pixels. Quanto maior o valor da métrica PSNR em *dB*, melhor é a qualidade da imagem de alta resolução obtida.

A métrica definida como Índice de Similaridade Estrutural (SSIM), foi proposta com o objetivo de avaliar a qualidade objetiva de uma imagem de acordo com as suas alterações estruturais, independente das alterações de contraste e luminância, (WANG et al., 2004; BROOKS; ZHAO; PAPPAS, 2008; RODRIGUEZ; BRESSAN,

2012). Nesta métrica, os pixels possuem uma forte dependência entre si, e esta dependência carrega informações estruturais importantes sobre os objetos. A métrica SSIM separa a tarefa de comparação da similaridade em 03 comparações independentes: luminância, contraste e estrutura.

A função $l(x, y)$ compara a luminância em função da intensidade de brilho entre duas imagens x e y , dada pela expressão:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4.18)$$

onde μ_x e μ_y representam a média de intensidades de brilho das imagens x e y e C_1 é uma constante incluída na divisão para evitar a instabilidade quando a soma $\mu_x^2 + \mu_y^2$ for muito próxima de zero. A função $c(x, y)$ compara o contraste usando o desvio padrão σ entre duas imagens x e y :

$$c(x, y) = \frac{(2\sigma_x\sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4.19)$$

onde σ_x e σ_y representam os desvios padrão das imagens x e y e a constante C_2 na expressão também é utilizada para estabilizar a divisão. A função $s(x, y)$ compara a estrutura em termos de uma covariação σ_{xy} entre duas amostras de luminosidade, dado por:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x^2 + \sigma_y^2 + C_3} \quad (4.20)$$

onde σ_{xy} covariância entre as imagens x e y e C_3 tem a mesma finalidade das constantes C_1 e C_2 . A métrica SSIM pode ser obtida a partir da seguinte expressão:

$$SSIM(x, y) = [l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma] \quad (4.21)$$

onde α , β e γ são parâmetros usados para ajustar a importância relativa das 03 funções. Considerando $\alpha = \beta = \gamma = 1$ e $C_3 = \frac{C_2}{2}$, a expressão pode ser simplificada na forma:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(2\sigma_x\sigma_y + C_2)} \quad (4.22)$$

Quanto maior o valor da métrica SSIM, melhor é a qualidade da imagem de alta resolução obtida.

4.4.4 Resultados Obtidos

As Tabelas 4.1 e 4.2 apresentam os resultados em PSNR e SSIM, do algoritmo de super-resolução proposto com os algoritmos usados para comparação e citados na seção 4.4.2. Os testes computacionais foram realizados para o fator de ampliação da imagem de $s = 2$.

O método proposto apresentou resultados superiores em PSNR e SSIM quando comparados com o método de interpolação bicúbica para todas as imagens. Com relação ao algoritmo proposto por (ZEYDE; ELAD; PROTTER, 2012), os resultados apresentam uma leve melhora no índice de similaridade estrutural (SSIM). Esta melhoria está relacionada a uma melhor qualidade visual da imagem interpolada DCT com a preservação dos coeficientes de baixa frequência.

Observamos que a utilização da interpolação DCT somente na etapa de reconstrução da imagem de alta resolução e não na etapa de treinamento para a construção do conjunto de imagens de baixa resolução equivalente, manteve os resultados em termos de PSNR muito próximos dos demais algoritmos de super-resolução e interpolação. As melhorias propostas no capítulo 5 melhoraram o desempenho do método de super-resolução proposto.

Tabela 4.1 - Comparação em PSNR [dB] do método SR proposto, interpolação bicúbica, método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s = 2$ e dicionário com 512 átomos.

Imagem	Nº Iterações Algoritmo k-SVD	PSNR (dB)		
		Interpolação Bicúbica	SR com Representação Esparsa (ZEYDE; ELAD; PROTTER, 2012)	Método Proposto
Lena	1	34,697	36,167	35,554
Lena	20	34,697	36,219	35,574
Bridge	1	26,523	27,399	27,149
Bridge	20	26,523	27,444	27,440
Man	1	27,938	29,065	28,770
Man	20	27,938	29,118	28,808

Tabela 4.2 - Comparação em SSIM do Método SR proposto, Interpolação Bicúbica, Método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s = 2$ e dicionário com 512 átomos.

Imagem	Nº Iterações Algoritmo k-SVD	SSIM		
		Interpolação Bicúbica	SR com Representação Esparsa (ZEYDE; ELAD; PROTTER, 2012)	Método Proposto
Lena	1	0,990	0,996	0,997
Lena	20	0,990	0,996	0,997
Bridge	1	0,973	0,991	0,992
Bridge	20	0,973	0,991	0,992
Man	1	0,980	0,992	0,993
Man	20	0,980	0,993	0,993

As Tabelas 4.3 e 4.4 apresentam os resultados do método proposto para um fator de ampliação $s = 3$. Observa-se que o algoritmo proposto produziu resultados superiores em termos de PSNR e SSIM quando comparado com o algoritmo de interpolação bicúbica. O algoritmo proposto por (ZEYDE; ELAD; PROTTER, 2012) produziu os melhores resultados objetivos para todas as imagens testadas. Além do exposto anteriormente, observamos que os efeitos de bloco (*Blocking*) da imagem interpolada DCT para este fator de interpolação, não contribuiu para melhoria do algoritmo proposto em relação à proposta de (ZEYDE; ELAD; PROTTER, 2012).

Tabela 4.3 - Comparação em PSNR [dB] do método SR proposto, Interpolação Bicúbica, método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s = 3$ e dicionário com 512 átomos.

Imagem	Nº Iterações Algoritmo k-SVD	PSNR (dB)		
		Interpolação Bicúbica	SR com Representação Esparsa (ZEYDE; ELAD; PROTTER, 2012)	Método Proposto
Lena	1	31,704	32,941	32,756
Lena	20	31,704	33,021	32,818
Bridge	1	24,354	24,925	24,880
Bridge	20	24,354	24,943	24,892
Man	1	25,715	26,559	26,464
Man	20	25,715	26,595	26,490

Tabela 4.4 - Comparação em SSIM do Método SR proposto, Interpolação Bicúbica, Método SR proposto por (ZEYDE; ELAD; PROTTER, 2012), para um fator de ampliação $s = 3$ e dicionário com 512 átomos.

Imagem	Nº Iterações Algoritmo k-SVD	SSIM		
		Interpolação Bicúbica	SR com Representação Esparsa (ZEYDE; ELAD; PROTTER, 2012)	Método Proposto
Lena	1	0,953	0,966	0,965
Lena	20	0,953	0,966	0,965
Bridge	1	0,866	0,899	0,896
Bridge	20	0,866	0,900	0,896
Man	1	0,903	0,928	0,923
Man	20	0,903	0,929	0,926

As Figuras 4.6, 4.7 e 4.8 ilustram as diferenças entre a imagem original e as imagens de alta resolução obtidas pelo método de interpolação bicúbica, pelo método de super-resolução proposto em (ZEYDE; ELAD; PROTTER, 2012) e o método de super-resolução proposto neste capítulo. Os testes foram realizados para o fator de interpolação $s = 2$ com as imagens Man, Bridge e Pepper. Considerando a qualidade visual das imagens obtidas, podemos observar nas figuras 4.6(d), 4.7(d) e 4.8(d), que o algoritmo proposto apresenta uma imagem final menos borrada e com textura mais próxima da imagem original, representada pelas figuras 4.6(a), 4.7(a) e 4.8(a). Este aspecto está relacionado a uma melhor preservação dos coeficientes de baixa frequência da imagem pela interpolação DCT, em comparação com os outros métodos.

Figura 4.6 - Qualidade visual da imagem “Man” para fator ampliação $s = 2$ e dicionário com 512 átomos. (a) Imagem Original, (b) Interpolação Bicúbica, (c) SR baseada em representação esparsa (ZEYDE; ELAD; PROTTER, 2012), (d) Método Proposto.

Figura 4.7 - Qualidade visual na imagem “Bridge” para fator interpolação $s = 2$ e dicionário com 512 átomos. (a) Imagem Original, (b) Interpolação Bicúbica, (c) SR baseada em representação esparsa (ZEYDE; ELAD; PROTTER, 2012), (d) Método Proposto.

Fonte: Autor

Figura 4.8 - Testes de qualidade visual na imagem “Pepper” para fator interpolação $s = 2$ e dicionário com 512 átomos. (a) Imagem Original, (b) Interpolação Bicúbica, (c) SR baseada em representação esparsa (ZEYDE; ELAD; PROTTER, 2012), (d) Método Proposto.

Fonte: Autor

4.5 Conclusões do Capítulo

Neste capítulo foi apresentado o método proposto para super-resolução de imagem única numa estratégia que combina a interpolação no domínio da transformada DCT com o método de representação esparsa de sinais. Simulações computacionais foram realizadas comparando o método proposto com o método de interpolação bicúbica e com o método proposto em (ZEYDE; ELAD; PROTTER, 2012). Os resultados obtidos, principalmente utilizando a métrica SSIM e pela qualidade visual das imagens finais, demonstraram a eficiência do método proposto para o fator de ampliação $s = 2$.

O próximo capítulo apresenta algumas contribuições no algoritmo proposto, com vistas à melhoria da qualidade final da imagem de alta resolução, bem como a redução tempo computacional do algoritmo.

Capítulo 5

Implementação Iterativa e Escalonável do Método Proposto

Este capítulo apresenta as contribuições para o aprimoramento do método proposto de super-resolução apresentado no capítulo anterior. Para melhorar a qualidade da imagem de alta resolução final e o tempo computacional do algoritmo, algumas melhorias na construção dos dicionários e extração de características são apresentadas. Também é proposta uma implementação iterativa e escalonável na etapa de reconstrução da imagem de alta resolução. No final do capítulo são realizados alguns experimentos computacionais com o objetivo de avaliar o método proposto em relação aos métodos de interpolação, e também avaliar a melhoria na qualidade das imagens com as melhorias propostas.

5.1 Introdução

A implementação do método de super-resolução proposto neste capítulo também é dividida em duas principais etapas: uma etapa de treinamento e construção de dicionários e uma segunda etapa de reconstrução da imagem de alta resolução.

A primeira contribuição proposta é a construção de dois dicionários de baixa resolução, utilizando filtros distintos de extração de características, conforme será abordado na seção 5.2. Na etapa de reconstrução, uma das contribuições mais importantes para a melhoria da qualidade final da imagem, foi a reinserção de uma imagem parcial no conjunto de treinamento para a realização de uma nova etapa de treinamento e reconstrução. Estas contribuições serão detalhadas na seção 5.3.

5.2 Treinamento e Construção dos Dicionários

Os procedimentos de construção do conjunto de treinamento e a extração dos fragmentos das imagens são muito parecidos com os descritos nas seções 4.2.1 e 4.2.2. No entanto, uma alteração importante ocorre na etapa de extração de

características dos fragmentos por meio de filtragem passa-alta para a construção de um segundo dicionário de baixa resolução.

Para o treinamento e construção do primeiro dicionário D_{L1} , os fragmentos das imagens de baixa e alta resolução do conjunto de treinamento são extraídos de acordo com as expressões:

$$\tilde{p}_{L1}^i = F_r^i * I_L \quad (5.1)$$

$$\tilde{p}_{H1}^i = F_r^i * I_H \quad (5.2)$$

onde $\tilde{p}_{L1}^i$ e $\tilde{p}_{H1}^i$ representam os fragmentos das imagens de baixa e alta resolução extraídos na posição i , I_L e I_H correspondem as imagens de baixa e alta resolução do conjunto de treinamento, respectivamente, e F_r ($r = 1,2,3,4$) representa o conjunto de filtros passa-alta Laplaciano e gradiente, conforme já descritos no capítulo 4.

Para o segundo dicionário D_{L2} , as características são extraídas utilizando filtros de aguçamento por máscara de nitidez (do inglês, *Unsharp Mask*), (GONZALEZ; WOODS, 2010). O processo de aguçamento da imagem utilizando a máscara de nitidez é definido como:

$$I_{mask} = I_L - I_{L(LP)} \quad (5.3)$$

onde,

$$I_{L(LP)} = H_{LP} * I_L \quad (5.4)$$

Na equação (5.3), I_{mask} representa a máscara de nitidez para a imagem de baixa resolução do conjunto de treinamento I_L . Na equação (5.4), $I_{L(LP)}$ é a imagem de baixa resolução suavizada por meio de filtragem passa-baixa com o filtro H_{LP} . Este processo utilizando a máscara de nitidez permite o aguçamento das componentes de alta frequência presentes na imagem de baixa resolução I_L , e a extração de características não presentes no dicionário D_{L1} . A figura 5.1 ilustra o processo de aguçamento usando a máscara de nitidez.

Figura 5.1 - Ilustração unidimensional do aguçamento por máscara de Nitidez (a) sinal Original, (b) Sinal borrado, (c) Máscara de Nitidez e (d) Sinal realçado pelo aguçamento.

Fonte: Autor

Assim, com a máscara de nitidez é possível extrair novos tipos de fragmentos de baixa e alta resolução de acordo com as expressões:

$$\tilde{p}_{L2}^i = F_r^i * [I_L + k \cdot I_{mask}] \quad (5.5)$$

$$\tilde{p}_{H2}^i = F_r^i * [I_H + k \cdot I_{mask}] \quad (5.6)$$

onde I_{mask} corresponde a máscara de nitidez para cada imagem I_L ou I_H e k é uma constante utilizada para enfatizar a contribuição da máscara de nitidez, sendo adotado um valor de $k = 1$. Novamente, Podemos observar que a utilização dos filtros Laplaciano e Gradiente produz fragmentos com alta dimensionalidade. Desta forma, o algoritmo PCA é utilizado para reduzir a dimensionalidade dos fragmentos. Os fragmentos resultantes são dados por:

$$p_{L1}^i = B\tilde{p}_{L1}^i \quad (5.7)$$

$$p_{H1}^i = B\tilde{p}_{H1}^i \quad (5.8)$$

$$p_{L2}^i = B\tilde{p}_{L2}^i \quad (5.9)$$

$$p_{H2}^i = B\tilde{p}_{H2}^i \quad (5.10)$$

Assim, da mesma forma que no capítulo anterior, os dicionários de baixa resolução D_{L1} e D_{L2} são treinados iterativamente com os algoritmos k-SVD e OMP, (AHARON; ELAD; BRUCKSTEIN, 2006). A diferença aqui está no treinamento e construção de dois dicionários, a partir da solução dos seguintes problemas de otimização:

$$\arg \min \|p_{L1}^i - D_{L1}\alpha_{1i}\|_2^2 \quad \text{sujeito a} \quad \forall_i \|\alpha_{1i}\|_0 \leq T_0 \quad (5.11)$$

$$\arg \min \|p_{L2}^i - D_{L2}\alpha_{2i}\|_2^2 \quad \text{sujeito a} \quad \forall_i \|\alpha_{2i}\|_0 \leq T_0 \quad (5.12)$$

em que α_{1i} e α_{2i} representam os vetores esparsos obtidos e T_0 é o parâmetro de esparsidade. Os dicionários de alta resolução D_{H1} e D_{H2} também são obtidos aplicando-se a expressão pseudo-inversa, da forma:

$$D_{H1} = P_{H1}Q_1^T(Q_1Q_1^T)^{-1} \quad (5.13)$$

$$D_{H2} = P_{H2}Q_2^T(Q_2Q_2^T)^{-1} \quad (5.14)$$

onde P_{H1} e P_{H2} são as matrizes que contém os fragmentos de alta resolução extraídos do conjunto de treinamento, utilizando o mesmo conceito explicado nas equações (5.1) e (5.5). As matrizes Q_1 e Q_2 contém todos os vetores esparsos obtidos na etapa de codificação esparsa, considerando as equações (5.11) e (5.12). A Figura 5.2 apresenta uma visão geral da etapa de treinamento e construção dos dicionários.

Figura 5.2 - Diagrama geral do treinamento e construção dos dicionários para o método iterativo proposto de super-resolução.

Fonte: Autor

5.3 Reconstrução da Imagem de Alta Resolução Final

A segunda etapa do processo consiste na reconstrução da imagem de alta resolução final. Inicialmente, a imagem de baixa resolução Y_L , de tamanho $N_1 \times N_2$, é interpolada no domínio da transformada DCT, como segue:

$$Y_{DCT}^{LR} = Q_{DCT} Y_L \quad (5.15)$$

onde Q_{DCT} representa a operação de interpolação da imagem Y_L no domínio da transformada DCT. A imagem interpolada Y_{DCT}^{LR} , de tamanho $M_1 \times M_2$, com $M_1 > N_1$ e $M_2 > N_2$, representa as componentes de baixa frequência da imagem de baixa resolução na composição da imagem de alta resolução final.

Em seguida, a imagem Y_L passa por um processo de filtragem e extração de características utilizando os mesmos filtros passa-alta da etapa de treinamento e construção dos dicionários da seção 5.2. Desta forma, são obtidos dois tipos de fragmentos de baixa resolução que representam diferentes características da imagem de baixa resolução Y_L , conforme a seguir:

$$\tilde{p}_{Y_L1}^i = F_r^i * Y_L \quad (5.16)$$

$$\tilde{p}_{Y_L2}^i = F_r^i * [Y_L + k \cdot Y_{Lmask}] \quad (5.17)$$

onde $\tilde{p}_{Y_L1}^i$ na equação (5.16), é um fragmento da imagem de baixa resolução extraído da posição i , após o processo de filtragem passa-alta utilizando o conjunto de quatro filtros Laplaciano e Gradiente das equações (4.1) e (4.2). O fragmento $\tilde{p}_{Y_L2}^i$ na equação (5.17) é obtido após o processo de aguçamento por máscara de nitidez, de acordo com as equações (5.3) e (5.4).

O processo de extração de características usando o conjunto de filtros Gradiente e Laplaciano produzem dois fragmentos $\tilde{p}_{Y_L1}^i$ e $\tilde{p}_{Y_L2}^i$ com alta dimensionalidade. Para reduzir a dimensionalidade e consequentemente o custo computacional do algoritmo, aqui também é utilizada a redução de dimensionalidade baseada em Análise de Componentes principais (PCA):

$$p_{Y_L1}^i = B \cdot \tilde{p}_{Y_L1}^i \quad (5.18)$$

$$p_{Y_L2}^i = B \cdot \tilde{p}_{Y_L2}^i \quad (5.19)$$

onde B representa a operação de projeção PCA. A partir dos fragmentos $p_{Y_{L1}}^i$ e $p_{Y_{L2}}^i$, e dos dicionários D_{L1} e D_{L2} obtidos na etapa de treinamento e construção dos dicionários, é utilizado o algoritmo de codificação esparsa OMP para calcular os coeficientes esparsos $\alpha_{Y_{L1}}^i$ e $\alpha_{Y_{L2}}^i$ de representação dos fragmentos $p_{Y_{L1}}^i$ e $p_{Y_{L2}}^i$ em relação aos dicionários, conforme já descrito na seção anterior.

Com os coeficientes esparsos $\alpha_{Y_{L1}}^i$ e $\alpha_{Y_{L2}}^i$ e os dicionários de alta resolução D_{H1} e D_{H2} , a próxima etapa é a obtenção dos fragmentos de alta resolução, de acordo com as seguintes expressões:

$$p_{H1}^i = D_{H1} \alpha_{Y_{L1}}^i \quad (5.20)$$

$$p_{H2}^i = D_{H2} \alpha_{Y_{L2}}^i \quad (5.21)$$

Os fragmentos p_{H1}^i e p_{H2}^i são reunidos, utilizando uma solução por mínimos quadrados de forma a obter as imagens I_{HR1} e I_{HR2} , que representam as componentes de alta frequência da imagem:

$$I_{HR1} = \left[\sum_i F_r^{iT} F_r^i \right]^{-1} \sum_i F_r^{iT} p_{H1}^i \quad (5.22)$$

$$I_{HR2} = \left[\sum_i F_r^{iT} F_r^i \right]^{-1} \sum_i F_r^{iT} p_{H2}^i \quad (5.23)$$

A imagem de alta resolução final do método é composta pela imagem interpolada DCT Y_{DCT}^{LR} , e as imagens obtidas pelo processo de representação esparsa I_{HR1} e I_{HR2} , por meio da expressão abaixo:

$$I_{HRF} = Y_{DCT}^{LR} + \gamma(I_{HR1} + I_{HR2}) \quad (5.24)$$

onde $\gamma = 1/2$ para dar a mesma contribuição das imagens I_{HR1} e I_{HR2} na composição da imagem final. A Figura 5.3 apresenta um diagrama geral do método proposto.

Figura 5.3 - Diagrama geral do método proposto de super-resolução.

Fonte: Autor

5.4 Etapa Iterativa e Escalonável do Método Proposto

Outra contribuição importante para melhorar a qualidade da imagem final foi tornar o método de super-resolução proposto iterativo e escalonável a partir da segunda iteração. Na prática, a imagem de alta resolução $I_{HRF}^{(1)}$ obtida na primeira iteração do algoritmo é inserida no conjunto de treinamento para a realização de uma nova etapa de treinamento e construção dos dicionários. A partir da segunda iteração do algoritmo, a imagem de alta resolução final tem a forma:

$$I_{HRF}^{(n)} = I_{HRF}^{(n-1)} + \frac{1}{2} \left(I_{HR1}^{(n)} + I_{HR2}^{(n)} \right), \quad n = 2, \dots, N \quad (5.25)$$

onde N é o número de iterações do algoritmo. A imagem $I_{HRF}^{(1)}$ obtida na primeira iteração pode ser inserida de duas formas, conforme será visto nas seções seguintes.

5.4.1 Imagem de alta resolução parcial com o mesmo fator de ampliação

Nesta configuração, a imagem $I_{HRF}^{(1)}$ é gerada com o mesmo fator de ampliação da imagem de alta resolução final, sendo inserida no conjunto de treinamento para uma nova etapa de treinamento e construção dos dicionários, de acordo com o número de iterações desejado.

Por exemplo, se a ampliação final desejada é de três vezes o valor da imagem de baixa resolução de entrada, $I_{HRF}^{(1)}$ é construída na primeira iteração do algoritmo com um fator de ampliação igual a três ($s = 3$), sendo inserida no conjunto de treinamento para as próximas iterações do algoritmo, até obter a imagem final de alta resolução. A Figura 5.4 ilustra um exemplo do método iterativo para esta configuração.

Figura 5.4 - Método Iterativo proposto usando o mesmo fator de ampliação. A imagem $I_{HRF}^{(1)}$ é inserida no conjunto de treinamento com o mesmo fator de interpolação da imagem $I_{HRF}^{(F)}$.

Fonte: Autor

5.4.2 Imagem Alta Resolução Parcial com o Fator de Ampliação Diferente

Nesta segunda configuração, a imagem $I_{HRF}^{(1)}$ é gerada com um fator de ampliação menor que a imagem HR final, sendo inserida também no conjunto de treinamento para uma nova etapa de treinamento e construção dos dicionários, conforme o número de iterações desejado.

Voltando ao exemplo da seção anterior, se a ampliação final desejada é de três vezes o valor da imagem de baixa resolução de entrada, $I_{HRF}^{(1)}$ é construída na primeira iteração do algoritmo com um fator de ampliação igual a dois ($s = 2$), sendo inserida no conjunto de treinamento para as próximas iterações do algoritmo,

até obter a imagem final super-resolvida com $s = 3$. A ideia do algoritmo para esta configuração é construir a imagem final a partir da inserção das imagens parciais no conjunto de treinamento de forma escalonável, iniciando o processo com um fator de ampliações $s = 2$. A figura 5.5 ilustra um exemplo do método iterativo para esta configuração.

Figura 5. 5 - Método Iterativo proposto usando fatores de ampliação diferentes. A imagem $I_{HRF}^{(1)}$ é inserida no conjunto de treinamento com um fator de interpolação da imagem menor que $I_{HRF}^{(F)}$.

Fonte: Autor

5.5 Experimentos Realizados com o Algoritmo Proposto

5.5.1 Parâmetros do Experimento

Para ilustrar o desempenho do algoritmo proposto, foram realizadas diversas simulações computacionais em MATLAB. As imagens utilizadas para os testes do algoritmo foram: *Lena* (512×512), *Bridge* (512×512), *Man* (512×512), *Pepper* (512×512), *Baboon* (500×480), *Barbara* (720×576), *Coastguard* (352×288), e *Foreman* (352×288). As imagens foram degradadas por subamostragem e borrimento e depois ampliadas utilizando fatores iguais a $s = 2$ e $s = 3$.

O conjunto de imagens utilizado para o treinamento dos dicionários D_{L1} e D_{L2} foi o mesmo utilizado no capítulo 4 e pelos autores (YANG et al., 2010; ZEYDE; ELAD; PROTTER, 2012). Foi adotado um total de 20 iterações dos algoritmos k-SVD e OMP para o treinamento e construção dos dicionários D_{L1} e D_{L2} . Os testes foram realizados variando o tamanho dos dicionários para valores iguais a 256, 512 e 1024 átomos. A figura 5.6 ilustra um exemplo de dois dicionários de alta resolução D_{H1} e D_{H2} construídos de tamanho 512 átomos.

Figura 5.6 - Exemplos de dois Dicionários de alta resolução construídos na etapa de treinamento. (a) Dicionário (D_{H1}), (b) Dicionário (D_{H2})

Fonte: Autor

5.5.2 Comparações com Algoritmos de Interpolação

O algoritmo proposto neste capítulo foi comparado com os algoritmos de interpolação bicúbica e interpolação DCT. A ideia foi comparar o algoritmo proposto com técnicas de ampliação de imagens e vídeos, de custo computacional reduzido e amplamente utilizadas comercialmente.

5.5.3 Métricas Utilizadas

Os testes foram realizados usando as mesmas métricas do capítulo 4, a relação sinal ruído de pico (PSNR), o índice de similaridade estrutural (SSIM) e a qualidade visual da imagem de alta resolução reconstruída em relação à imagem original.

5.5.4 Resultados Obtidos para o Fator de Ampliação $s = 2$

Esta seção apresenta os resultados obtidos do algoritmo proposto usando um fator de ampliação da imagem $s = 2$. Os dicionários D_{L1} e D_{L2} foram treinados sequencialmente com os algoritmos k-SVD e OMP. A imagem de alta resolução $I_{HRF}^{(1)}$, obtida na primeira iteração do algoritmo, foi inserida no conjunto de treinamento para a execução de duas outras iterações, de acordo com a configuração apresentada na seção 5.4.1.

A primeira etapa das simulações computacionais foi realizada utilizando todas as 92 imagens do conjunto de treinamento, como já mencionada na seção anterior. Para fins de comparação dos resultados, foi adotada a nomenclatura “conjunto de treinamento completo”.

Os resultados numéricos obtidos, em PSNR e SSIM, do método de super-resolução proposto e um comparativo com os métodos de interpolação bícubica e interpolação DCT estão apresentados na tabela 5.1.

Na segunda etapa das simulações computacionais foram selecionadas 05 (cinco) imagens do conjunto de treinamento para cada imagem testada, usando como critério seu conteúdo estrutural. Para fins de comparação dos resultados, foi adotada a nomenclatura “conjunto de treinamento reduzido”

Na Figura 5.7 são apresentadas algumas imagens testadas e o seu respectivo conjunto de imagens selecionadas para o treinamento e reconstrução. A tabela 5.2 apresenta os resultados numéricos obtidos, em PSNR e SSIM, do método de super-resolução proposto, e compara com os métodos de interpolação, considerando o conjunto reduzido de imagens.

Figura 5.7 - Conjunto de imagens escolhidas para o conjunto reduzido de treinamento (5 imagens), considerando as imagens de teste Man, Lena, Bridge, Pepper e Babbon

Imagens testadas	Conjunto de Imagens Escolhidas

Fonte: Autor

Tabela 5.1 - Comparação em PSNR [dB] e SSIM, do Método Proposto com os métodos de interpolação bicúbica e DCT, para um conjunto treinamento Completo, e fator de ampliação $s = 2$.

Imagem	Métrica	Interpolação Bicúbica	Interpolação DCT	Método Proposto		
				$D = 256$	$D = 512$	$D = 1024$
Man	PSNR	27,925	28,412	28,585	28,625	28,625
	SSIM	0,980	0,989	0,989	0,989	0,989
Lena	PSNR	34,669	35,216	35,900	35,422	35,422
	SSIM	0,990	0,990	0,995	0,995	0,995
Bridge	PSNR	26,518	26,995	27,074	27,07	27,065
	SSIM	0,974	0,986	0,987	0,987	0,987
Pepper	PSNR	33,924	34,979	35,679	35,790	35,807
	SSIM	0,994	0,997	0,997	0,997	0,997
Baboon	PSNR	25,061	25,518	25,574	25,581	25,574
	SSIM	0,957	0,976	0,976	0,976	0,976
Barbara	PSNR	27,932	28,353	28,416	28,432	28,424
	SSIM	0,963	0,975	0,975	0,975	0,975
Coast Guard	PSNR	29,132	29,661	29,778	29,783	29,786
	SSIM	0,790	0,828	0,831	0,831	0,831
Foreman	PSNR	32,662	33,783	34,237	34,310	34,343
	SSIM	0,950	0,958	0,959	0,960	0,960

Fonte: Autor

Tabela 5.2 - Comparação em PSNR [dB] e SSIM, do método proposto com os métodos de interpolação bicúbica e DCT, para um conjunto treinamento reduzido (5 imagens), e fator de ampliação $s = 2$.

Imagem	Métrica	Interpolação Bicúbica	Interpolação DCT	Método Proposto		
				$D = 256$	$D = 512$	$D = 1024$
Man	PSNR	27,925	28,412	28,466	28,465	28,481
	SSIM	0,980	0,989	0,989	0,989	0,988
Lena	PSNR	34,669	35,216	35,268	35,245	35,200
	SSIM	0,990	0,990	0,995	0,995	0,995
Bridge	PSNR	26,518	26,995	27,020	27,007	26,962
	SSIM	0,974	0,986	0,987	0,986	0,987
Pepper	PSNR	33,924	34,979	35,403	35,366	35,397
	SSIM	0,994	0,997	0,997	0,997	0,997
Baboon	PSNR	25,061	25,518	25,540	25,526	25,494
	SSIM	0,957	0,976	0,976	0,976	0,976
Barbara	PSNR	27,932	28,353	28,411	28,340	28,407
	SSIM	0,963	0,975	0,975	0,975	0,975
Coast Guard	PSNR	29,132	29,661	29,738	29,701	29,619
	SSIM	0,790	0,828	0,830	0,829	0,825
Foreman	PSNR	32,662	33,783	33,870	33,863	33,839
	SSIM	0,950	0,958	0,957	0,957	0,955

Fonte: Autor

Os resultados numéricos apresentados nas tabelas 5.1 e 5.2 indicam que o método proposto utilizando o conjunto completo para treinamento, produziu imagens com maior qualidade do que os métodos de interpolação testados.

Para as simulações computacionais utilizando o conjunto reduzido de imagens para o treinamento, o método proposto de super-resolução apresentou resultados superiores aos métodos de interpolação, porém inferiores aos testes com o conjunto completo de treinamento. Observou-se também uma redução significativa do tempo computacional do algoritmo, conforme será visto mais adiante.

Os resultados não apresentaram uma melhora significativa quando se aumentou o tamanho dos dicionários nas simulações. Se considerarmos a métrica SSIM, os resultados praticamente não se alteram.

5.5.5 Resultados Obtidos para o Fator de Ampliação $s = 3$

Para as simulações computacionais utilizando o fator de interpolação $s = 3$, o conjunto de treinamento foi o mesmo utilizado na seção anterior. Entretanto, a imagem de alta resolução $I_{HRF}^{(1)}$, foi construída usando um fator de interpolação $s = 2$ na primeira iteração do algoritmo, e foi inserida no conjunto de treinamento para a execução de duas outras iterações com fator de interpolação $s = 3$.

As tabelas 5.3 e 5.4 apresentam os resultados numéricos obtidos, em PSNR e SSIM, do método de super-resolução proposto, e compara com os métodos de interpolação.

Tabela 5.3 - Comparação em PSNR [dB] e SSIM, do método proposto com os métodos de interpolação bicúbica e DCT, para um conjunto completo de treinamento (92 imagens), e fator de ampliação $s = 3$.

Imagem	Métrica	Interpolação Bicúbica	Interpolação DCT	Método Proposto		
				$D = 256$	$D = 512$	$D = 1024$
Man	PSNR	25,705	25,777	26,055	26,093	26,090
	SSIM	0,904	0,916	0,920	0,920	0,921
Lena	PSNR	31,695	31,622	32,131	32,178	32,221
	SSIM	0,953	0,958	0,961	0,961	0,961
Bridge	PSNR	24,352	24,418	24,556	24,552	24,543
	SSIM	0,867	0,886	0,890	0,889	0,889
Pepper	PSNR	31,260	31,408	32,228	32,233	32,307
	SSIM	0,971	0,972	0,975	0,975	0,975
Baboon	PSNR	24,055	24,205	24,349	24,345	24,341
	SSIM	0,818	0,843	0,848	0,848	0,848
Barbara	PSNR	26,197	26,333	26,516	26,521	26,502
	SSIM	0,877	0,891	0,894	0,894	0,893
Coast Guard	PSNR	26,541	26,594	26,788	26,847	26,829
	SSIM	0,621	0,646	0,653	0,655	0,653
Foreman	PSNR	29,603	30,065	30,830	30,936	30,984
	SSIM	0,905	0,897	0,909	0,910	0,911

Fonte: Autor

Tabela 5.4 - Comparação em PSNR [dB] e SSIM, do método proposto com os métodos de interpolação bicúbica e DCT, para um conjunto reduzido de treinamento (5 imagens), e fator de ampliação $s = 3$.

Imagem	Métrica	Interpolação Bicúbica	Interpolação DCT	Método Proposto		
				$D = 256$	$D = 512$	$D = 1024$
Man	PSNR	25,705	25,777	25,726	25,525	25,951
	SSIM	0,904	0,916	0,914	0,910	0,918
Lena	PSNR	31,695	31,622	32,628	32,940	33,436
	SSIM	0,953	0,958	0,964	0,966	0,970
Bridge	PSNR	24,352	24,418	24,652	24,727	24,852
	SSIM	0,867	0,886	0,892	0,895	0,900
Pepper	PSNR	31,260	31,408	32,568	32,842	33,036
	SSIM	0,971	0,972	0,975	0,976	0,976
Baboon	PSNR	24,055	24,205	24,294	24,212	24,065
	SSIM	0,818	0,843	0,847	0,845	0,840
Barbara	PSNR	26,197	26,333	26,929	27,020	27,155
	SSIM	0,877	0,891	0,901	0,904	0,908
Coast Guard	PSNR	26,541	26,594	26,723	26,707	26,686
	SSIM	0,621	0,646	0,649	0,647	0,641
Foreman	PSNR	29,603	30,065	31,549	31,698	31,617
	SSIM	0,905	0,897	0,912	0,910	0,904

Fonte: Autor

Os resultados numéricos apresentados nas tabelas 5.3 e 5.4 demonstraram que o método proposto produziu imagens com maior qualidade do que os métodos de interpolação testados, tanto para o conjunto completo quanto para o conjunto reduzido de imagens

Os resultados demonstraram a efetividade do método proposto para o fator de ampliação $s = 3$ e utilizando um conjunto reduzido de imagens para o treinamento e construção dos dicionários. Os resultados em termos de qualidade visual e uma análise do tempo computacional do algoritmo serão apresentados nas próximas seções.

5.5.6 Qualidade Visual do Método Proposto

A qualidade visual do método iterativo proposto também foi testada e comparada com os métodos de interpolação. Os parâmetros testados foram os mesmos descritos nas seções anteriores e os resultados para as imagens *Pepper*, *Coast Guard*, *Lena* e *Foreman*, são ilustrados nas figuras 5.8, 5.9, 5.10 e 5.11.

Figura 5.8 - Qualidade Visual da imagem “Pepper” para o Método Proposto com fator interpolação $s = 2$ e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT ; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT

Fonte: Autor

Figura 5.9 - Qualidade visual do método proposto para a imagem “CoastGuard”, fator de ampliação $s = 2$ e dicionários de tamanho 512 átomos. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT.

Fonte: Autor

Conforme podemos observar nas figuras 5.8(d), 5.8(e), 5.9(d), e 5.9(e), as porções das imagens *pepper* e *Coast Guard* testadas com o método proposto apresentam uma melhora em relação ao método de interpolação bicúbica, figuras 5.8(c), 5.9(c).

Considerando o fator de interpolação $s = 3$, os melhores resultados foram obtidos com o conjunto reduzido de imagens, conforme podemos observar nas figuras 5.10(e) 5.11(e), para as porções das imagens *Lena* e *Foreman*.

Figura 5.10 - Qualidade visual do método proposto para a imagem “Lena”, fator de ampliação $s = 3$ e dicionários de tamanho 512 átomos. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT

Fonte: Autor

Figura 5.11 - Qualidade visual do método proposto para a imagem “Foreman”, fator de ampliação $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Interpolação DCT; (d) Proposta Conj. 92 imgs; (e) Proposta Conj. 05 imgs + IT.

Fonte: Autor

5.5.7 Comparativo em relação ao tempo de execução dos algoritmos

As simulações computacionais também foram realizadas com o objetivo de avaliar a tempo de execução dos algoritmos. Para fins de comparação, os mesmos algoritmos da seção 5.5.7 foram analisados, exceto a interpolação bicúbica que não exige a execução da etapa de treinamento. As simulações foram realizadas usando um computador com processador Intel dual core T4300 2.10 GHz, memória RAM de 3 GB, e as rotinas executadas em MATLAB 7.0, e os algoritmos foram testados para um fator de ampliação da imagem $s = 3$.

Os resultados apresentados na tabela 5.5 confirmam a redução no tempo de execução com o uso da estratégia iterativa e escalonável proposta e com o uso do conjunto reduzido de imagens no treinamento.

Tabela 5.5 - Comparação do tempo de execução do método proposto com os métodos de interpolação bicúbica, método SR proposto por Zeyde et al (2012) e o método proposto nesta tese no capítulo 4

Imagem	Tempo Execução do Algoritmo (min)		
	SR com Representação Esparsa (Zeyde, Elad e Protter, 2012)	Método Proposto (capítulo 4)	Método Proposto (Capítulo 5)
Man	9,50	9,53	4,76
Bridge	9,88	9,92	9,48
Lena	9,30	9,34	4,61

Fonte: Autor

As figuras 5.12 e 5.13 apresentam os tempos de execução do algoritmo proposto para as imagens Lena, Bridge, Man, Pepper, Baboon, Barbara, Coastguard e Foreman, considerando o tamanho dos dicionários e o fator de ampliação da imagem.

Figura 5.12 - Tempo execução com fator de ampliação $s = 2$

Fonte: Autor

Figura 5.13 - Tempo execução com fator de ampliação $s = 3$

Fonte: Autor

5.6 Implementação do algoritmo com a inclusão da Imagem Alta Resolução parcial no processo de reconstrução

Esta seção apresenta outra importante contribuição no algoritmo proposto na seção 5.4 com o objetivo de melhorar a qualidade da imagem final. Na proposta apresentada na seção 5.4, a imagem $I_{HRF}^{(1)}$, obtida na primeira iteração era somente inserida no conjunto de treinamento para a realização de uma nova etapa de treinamento. Após a realização das etapas de treinamento e reconstrução de forma iterativa e escalonável, a imagem de alta resolução final era formada pela soma da imagem ampliada por interpolação DCT e uma imagem com os coeficientes de alta frequência obtidos no processo de treinamento, conforme definida pela equação (5.24).

Na alteração proposta aqui, os coeficientes de alta frequência obtidos na primeira iteração, tornam-se os coeficientes de baixa frequência a partir da segunda iteração do algoritmo, conforme pode ser observado na figura 5.14. Como exemplo, considerando um fator de interpolação $s = 3$, o algoritmo constrói uma imagem parcial de fator $s = 2$, insere no conjunto de treinamento para uma nova etapa de treinamento e utiliza esta mesma imagem redimensionada para o fator $s = 3$ no processo de reconstrução, em substituição a imagem Y_{DCT}^{LR} da equação (5.24). Todas as outras etapas do processo de super-resolução são semelhantes as apresentadas nas seções anteriores.

Figura 5.14 - Implementação do algoritmo com a alteração da imagem $Y_{HRF}^{(1)}$ no processo de reconstrução da imagem.

Fonte: Autor

5.6.1 Comparativo com outros métodos

O desempenho do algoritmo proposto considerando as melhorias apresentadas na seção anterior foi comparado com o algoritmo desenvolvido na seção 5.4 e com outros algoritmos de super-resolução propostos recentemente na literatura como (ZEYDE; ELAD; PROTTER, 2012) e (YANG *et al.*, 2010).

Foram utilizadas as imagens *Cameraman*, *Jetplane*, *Lake*, *Lena* e *living room* e considerado um fator de ampliação da imagem final de $s = 3$ para todos os experimentos. Os resultados em termos de PSNR e SSIM são apresentados na tabela 5.6.

Tabela 5.6 - Comparação em PSNR [dB] e SSIM, do método proposto com a substituição da imagem $Y_{HRF}^{(1)}$, e outros métodos de super-resolução propostos na literatura, para um conjunto treinamento composto por 5 imagens, e fator de ampliação de $s = 3$.

Imagem	Mét.	Interp. Bicúbica	Interp. DCT	Método (YANG et al, 2010)	Método (ZEYDE; ELAD; PROTTER, 2012)	Método Proposto (Seção 5.4)	Método Proposto (Substituição $I_{HRF}^{(1)}$)
Cameraman	PSNR	30,367	30,299	32,018	32,620	32,0563	33,987
	SSIM	0,9587	0,9640	0,959	0,974	0,969	0,984
Jetplane	PSNR	29,429	29,313	30,383	30,980	29,976	31,311
	SSIM	0,957	0,961	0,956	0,971	0,965	0,968
Lake	PSNR	27,419	27,414	28,063	28,556	27,917	29,599
	SSIM	0,938	0,937	0,944	0,956	0,949	0,971
Lenna	PSNR	31,704	31,622	32,668	33,029	33,436	33,565
	SSIM	0,953	0,958	0,956	0,966	0,970	0,977
Living Room	PSNR	26,840	26,834	27,441	27,608	27,327	28,537
	SSIM	0,893	0,904	0,912	0,920	0,914	0,956

Fonte: Autor

Conforme pode ser observado, o desempenho do algoritmo proposto com as melhorias propostas foram os melhores em termos de PSNR e SSIM para todas as imagens testadas.

5.6.2 Resultados experimentais de Qualidade Visual

A qualidade visual considerando as alterações propostas também foi testada e comparada com os diferentes métodos. Os parâmetros testados foram os mesmos descritos nas seções anteriores e os resultados estão ilustrados nas figuras 5.15, 5.16, 5.17, 5.18 e 5.19. Considerando que para os testes realizados, o critério de escolha adotado para as imagens do conjunto de treinamento foi visual e de acordo com as características da imagem testada, houve uma melhoria significativa na qualidade visual das imagens testadas, conforme verificado nas imagens 5.15(g), 5.16(g), 5.17(g), 5.18(g) e 5.19(g).

Figura 5.15 - Qualidade visual para a imagem “Lake” considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem $I_{HRF}^{(1)}$.

Fonte: Autor

Figura 5.16 - Qualidade visual para a imagem *Jetplane* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem $I_{HRF}^{(1)}$.

Fonte: Autor

Figura 5.17 - Qualidade visual para a imagem *Camerman* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem $I_{HRF}^{(1)}$.

Figura 5.18 - Qualidade visual para a imagem *Living room* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem $I_{HRF}^{(1)}$.

Fonte: Autor

Figura 5.19 - Qualidade visual para a imagem *car plate* considerando as alterações propostas na seção 5.6, fator de ampliação da imagem $s = 3$, e dicionários de tamanho 512. (a) Imagem Original; (b) Interpolação bicúbica; (c) Método SR (YANG et al, 2010); (d) Método SR (ZEYDE;ELAD;PROTTER, 2012); (e) Método proposto Seção 5.4; (f) Método Proposto com substituição da imagem $I_{HRF}^{(1)}$.

Fonte: Autor

5.7 Conclusões do Capítulo

Neste capítulo foram apresentadas as principais contribuições desta tese. O algoritmo proposto no capítulo 4 foi aprimorado, com a incorporação de algumas melhorias na etapa de treinamento e na etapa de reconstrução da imagem de alta resolução.

A primeira contribuição foi a inclusão de uma nova etapa de extração de características utilizando técnicas de aguçamento por máscara de nitidez, para a construção de um segundo dicionário. Esta solução buscou extrair diferentes características do conjunto de treinamento e ao mesmo tempo diminuir o tamanho dos dicionários treinados.

A segunda contribuição importante foi a inclusão de um processo iterativo e escalonável no algoritmo, inserindo no conjunto de treinamento e no processo de reconstrução, a imagem de alta resolução obtida na primeira iteração. As simulações computacionais realizadas comparando o método proposto com os métodos de interpolação, demonstraram que as soluções propostas melhoraram significativamente a qualidade objetiva e subjetiva das imagens de alta resolução. Verificou-se também que a utilização do algoritmo iterativo combinado com a utilização de um conjunto reduzido de imagem contribuiu para a redução do tempo computacional do algoritmo.

Capítulo 6

Conclusão

O desenvolvimento de um método de super-resolução capaz de buscar um compromisso entre eficiência computacional e qualidade na reconstrução da imagem de alta resolução é fundamental para o avanço e utilização destas técnicas em substituição as tradicionais técnicas de interpolação de imagens.

Esta tese apresentou uma proposta de super-resolução para imagem única que incorpora os benefícios da interpolação no domínio da transformada DCT e a eficiência dos métodos de reconstrução de imagens baseados no conceito de representação esparsa de sinais. A proposta procurou aproveitar alguns benefícios já conseguidos com os principais algoritmos de super-resolução existentes, aperfeiçoando algumas etapas para tornar o processo de reconstrução mais eficiente.

Uma das contribuições da tese foi a inclusão de uma nova etapa de extração de características utilizando técnicas de aguçamento por *máscara de nitidez* para a construção de um segundo dicionário compacto. Esta estratégia busca extrair mais informações estruturais dos fragmentos de baixa e alta resolução do conjunto de treinamento e ao mesmo tempo reduzir o tamanho dos dicionários.

Outra importante contribuição foi a inclusão de um processo iterativo e escalonável no algoritmo, inserindo no conjunto de treinamento e no processo de reconstrução, uma imagem de alta resolução obtida numa primeira iteração. Esta solução possibilitou uma melhora na qualidade da imagem final de alta resolução utilizando poucas imagens no conjunto de treinamento, e ao mesmo tempo resultou na diminuição no tempo de execução do algoritmo.

Para atingir o objetivo deste trabalho foi necessário cumprir algumas etapas. A primeira delas consistiu num estudo e análise dos principais trabalhos de super-resolução de imagens que propuseram o uso de técnicas baseadas em exemplos e técnicas de interpolação no domínio da transformada DCT. A partir desta análise, foram identificados alguns problemas e dificuldades associadas à melhoria da

qualidade da imagem e redução do custo computacional, considerando as diversas etapas do processo de reconstrução da imagem de alta resolução.

Em seguida, foi elaborado um algoritmo de super-resolução para imagem única que incorpora a interpolação no domínio da transformada DCT e a representação esparsa de sinais. A etapa seguinte dos estudos consistiu na implementação algumas melhorias e aperfeiçoamentos para tornar o algoritmo mais eficiente, utilizando um menor número de imagens do conjunto de treinamento.

Na etapa de simulações computacionais ficou demonstrada a capacidade do método proposto em produzir imagens com qualidade e com custo computacional reduzido. As simulações foram realizadas considerando diferentes parâmetros de análise e utilizando métricas de avaliação objetivas e subjetivas.

6.1 Trabalhos Futuros

Os trabalhos desenvolvidos nesta tese e os resultados apresentados abriram novas questões para trabalhos futuros. Dentre as oportunidades, podemos citar:

- Desenvolvimento de um modelo de super-resolução para sequências de imagens ou vídeos, incorporando no modelo de representação esparsa a informação de movimento entre os quadros em sequência;
- Desenvolvimento do modelo proposto com outros tipos de filtros de extração de características, como por exemplo, a utilização de filtros adaptativos de acordo com a característica estrutural das imagens pertencentes ao conjunto de treinamento;
- Utilização do modelo proposto em aplicações específicas para melhoria de imagens onde os recursos disponíveis possuem recursos limitados;
- Estudo e desenvolvimento do hardware necessário que contemple todas as melhorias e aperfeiçoamento propostos nesta tese.

6.2 Produção Científica

A partir dos estudos e resultados obtidos com as contribuições desta tese, alguns trabalhos foram publicados em eventos nacionais e internacionais a seguir:

REIS, S. R. S.; BRESSAN, G., Método de super-resolução de imagem digital usando interpolação DCT e representação esparsa. In: **II Workshop de Pós Graduação da Área de Concentração Engenharia de Computação do PPGE-EPUSP**, 10 de outubro de 2013, São Paulo-SP, Brasil.

REIS, S. R. S.; BRESSAN, G. Image super-resolution using a hybrid scheme with DCT interpolation and sparse representation method. In: **XXXI Simpósio Brasileiro de Telecomunicações - SBrT 2013**. 01 a 04 setembro de 2013, Fortaleza-CE, Brasil.

REIS, S. R. S.; BRESSAN, G. Image super-resolution using DCT interpolation and sparse learning-based method. In: **Fifth International Conference on Digital Image Processing (ICDIP 2013)**. Beijing (2013). v. 8878, p. 88784C-88784C-4.

Referências

- AHARON, M.; ELAD, M.; BRUCKSTEIN, A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. **IEEE Transactions on Signal Processing**, v. 54, n. 11, 2006.
- BAKER, S.; KANADE, T. Limits on super-resolution and how to break them. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 9, p. 1167-1183, Sep 2002.
- BANHAM, M. R.; KATSAGGELOS, A. K. Digital image restoration. **IEEE Signal Processing Magazine**, v. 14, n. 2, p. 24-41, 1997.
- BARRETO, D., ALVAREZ, L. D., MOLINA, R., KATSAGGELOS, A. K., CALLICO, G.M. Region-based super-resolution for compression. **Multidimensional Systems and Signal Processing**, v. 18, n. 2-3, p. 59-81, 2007.
- BODENSTEINER C.; DAROLTI C.; SCHWEIKARD A., Achieving super-resolution X-ray Imaging with mobile C-arm devices, **The international journal of Medical and Computer Assisted Surgery**, v. 5, p. 243-256, 2009.
- BOON, C. S.; GULERYUZ, O. G.; KAWAHARA, T.; SUZUKI, Y.i, Sparse super-resolution reconstructions of video from mobile devices in digital TV broadcast applications, Conference on Applications of Digital Image Processing XXIX, **Proceedings**, v. 6312, M3120-M3120, 2006.
- BROOKS, A. C.; ZHAO, X.; PAPPAS, T. N. Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions. **IEEE Transactions on Image Processing**, v. 17, n. 8, p. 1261-1273, 2008.
- CHAKRABARTI, A.; RAJAGOPALAN, A. N.; CHELLAPPA, R. Super-resolution of face images using kernel PCA-based prior. **IEEE Transactions on Multimedia**, v. 9, n. 4, p. 888-892, 2007.
- CHRISTOPOULOS, C.; SKODRAS, A.; EBRAHIMI, T. The JPEG2000 still image coding system: an overview. **IEEE Transactions on Consumer Electronics**, v. 46, n. 4, p. 1103-1127, 2000.
- DATSENKO, D.; ELAD, M. Example-based single document image super-resolution: a global MAP approach with outlier rejection. **Multidimensional Systems and Signal Processing**, v. 18, n. 2-3, p. 103-121, 2007.
- DONG, W., ZHANG, L., SHI, GUANGMING, WU, X. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. **IEEE Transactions on Image Processing**, v. 20, n. 7, p. 1838-1857, 2011.
- DONOHU, D. L. Compressed sensing. **IEEE Transactions on Information Theory**, v. 52, n. 4, p. 1289-1306, 2006.

ELAD, M.; DATSENKO, D. Example-based regularization deployed to super-resolution reconstruction of a single image. **Computer Journal**, v. 52, n. 1, p. 15-30, 2009.

ETTER, V.; JOVANOVIĆ I.; VETTERLI M., Use of Learned Dictionaries in Tomographic Reconstruction, XIV Wavelets and Sparsity, **Proceedings**, Vol. 8138-81381C-1, 2011.

FARSIU, S., ROBINSON, D., ELAD, M., MILANFAR, P.. Advances and challenges in super-resolution. **International Journal of Imaging Systems and Technology**, v. 14, n. 2, p. 47-57, 2004.

FREEMAN, W. T.; JONES, T. R.; PASZTOR, E. C. Example-based super-resolution. **IEEE Computer Graphics and Applications**, v. 22, n. 2, p. 56-65, 2002.

FREEMAN, W. T.; PASZTOR, E. C.; CARMICHAEL, O. T. Learning low-level vision. **International Journal of Computer Vision**, v. 40, n. 1, p. 25-47, 2000.

GARCIA, D. C.; DOREA, C.; QUEIROZ, R. L. Super Resolution for Multiview Images Using Depth Information. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 22, n. 9, p. 1249-1256, 2012.

GEVREKCI, M., GUNTURK, B. K., ALTUNBASAK, Y. POCS-based restoration of Bayer-sampled image sequences. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. **Proceedings**, Vol I, p. 753-756, 2007.

GLASNER, D., BAGON, S., IRANI, M. **Super-Resolution from a Single Image**. 12th IEEE International Conference on Computer Vision. **Proceedings**, Sep 29-Oct 02. p.349-356, Kyoto, JAPAN. 2009.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. Pearson, 2010.

GREENSPAN H., Super-resolution in Medical Imaging, **The Computer Journal**, v. 52, n. 1, p. 43-63, 2009.

GUNTURK, B. K.; BATUR, A. U.; ALTUNBASAK, Y.; HAYES, M. H.; MERSEREAU, R. M., Eigenface-domain super-resolution for face recognition, **IEEE Transactions on Image Processing**, v. 12, n. 5, p. 597-606, 2003.

HEATH, M. T. **Scientific computing**: an introductory survey. 2nd. ed. McGraw-Hill, 2002.

HONG, C.; DIT-YAN, Y.; YIMIN, X. Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04). **Proceedings**, June 27-July 2, 2004.

HUANG, H., FAN, X., QI, C., ZHU, S. H. A learning-based POCS algorithm for face image super-resolution reconstruction. In: International Conference on Machine Learning and Cybernetics, **Proceedings**, v. 1-9, p. 5071-5076, 2005

HUNG, E. M. **Realce de vídeo para sequencias de qualidade e resolução variáveis**. Tese (Doutorado) - Departamento Engenharia Elétrica. Universidade de Brasília, Brasília-DF, Brasil,(140p), 2012..

HUNG, E. M., GARCIA, D.C., QUEIROZ, R.L.D. Transform domain semi-super resolution. In: The International Conference on Image Processing ICIP'11 **Conference**, p. 1193-1196, 2011.

IRANI, M.; PELEG, S. Improving Resolution by Image Registration. In: CVGIP-Graphical Models and Image Processing, **Conference**, v. 53, n. 3, p. 231-239, 1991.

JIA, K.; WANG, X.; TANG, X. Image Transformation Based on Learning Dictionaries across Image Spaces. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 2, p. 367-380, 2013.

JIAN, S., NAN-NING, Z., HAI, T., HEUNG-YEUNG, S. Image hallucination with primal sketch priors. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03), **Proceedings**, p.II-729-36, vol.2, 18-20 June 2003.

JIJL, C. V.; CHAUDHURI, S. **Single-frame image super-resolution through contourlet learning**. Eurasip Journal on Applied Signal Processing, 2006.

JIJL, C. V.; CHAUDHURI, S.; CHATTERJEE, P. Single frame image super-resolution: should we process locally or globally?. **Multidimensional Systems and Signal Processing**, v. 18, n. 2-3, p. 123-152, 2007.

JIJL, C. V.; JOSHI, M. V.; CHAUDHURI, S. Single-frame image super-resolution using learned wavelet coefficients. **International Journal of Imaging Systems and Technology**, v. 14, n. 3, p. 105-112, 2004.

JOLLIFFE, I. T., **Principal Components Analysis**, 2nd ed. Springer, 2002.

JOSHI, M. V.; CHAUDHURI, S.; PANUGANTI, R. A learning-based method for image super-resolution from zoomed observations. **IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics**, v. 35, n. 3, p. 527-537, 2005.

KOMATSU, T., IGARASHI, T., AIZAWA, K., SAITO, T. Very high resolution imaging scheme with multiple different-aperture cameras. **Signal Processing: Image Communication**, v. 5, n. 5-6, p. 511-526, 1993.

KWOK-WAI, H.; WAN-CHI, S. Hybrid DCT-Wiener-based interpolation via learnt Wiener filter. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **Conference**, p.1419-1423, May 26-31 2013.

LI, X.; ORCHARD, M. New Edge-Directed Interpolation. **IEEE Transactions on Image Processing**, v. 10, n. 10, p. 1521-1527, 2001.

LIN, F. ;FOOKES, C. ;CHANDRAN, V. ;SRIDHARAN, S., Super-resolved faces for improved face recognition from surveillance video, **Advances in Biometrics, Proceedings, Lecture Notes in Computer Science**, v. 4642, p. 1-10, 2007.

LU, Y.; INAMURA, M., Spatial resolution improvement of remote sensing images by fusion of subpixel-shifted multi-observation images, **International Journal of Remote Sensing**, v. 24, i. 23, p. 4647-4660, 2003.

MAGMAN, P. Detection of Visible photons in CCD and CMOS: A comparative view. **Nuclear Instruments and Methods in Physics Research**, section A, p.199-212, 2003.

MANN, S.; PICARD, R. W.; Virtual Bellows - Constructing High Quality Stills from Video. In: IEEE International Conference on Image Processing- ICIP-94 **Conference**, vol. I, p. 363-367, 1994.

MATOS, F. M. D. S. **Reconhecimento de faces utilizando a transformada cosseno discreta**. Dissertação (Mestrado) - CCEN, Universidade Federal da Paraíba, João Pessoa-PB, Brasil,(122p) 2008.

NG, M. K.; SHEN H., LAM, E. Y.; ZHANG, L., A total variation regularization based super-resolution reconstruction algorithm for digital video, **Eurasip Journal on Advances in Signal Processing**, N° art. 74585, 2007.

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.

PAPOULIS, A. **Probability, random variables, and stochastics process**, McGraw-Hill, 1991.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de imagens digitais** - princípios, algoritmos e aplicações. São Paulo: Thomson Learning, 2008.

QIANG, W.; XIAOOU, T.; SHUM, H. **Patch based blind image super resolution**. In IEEE International Conference on Computer Vision - ICCV 2005, **Conference**, p.709-716 Vol. 1, 17-21 Oct. 2005.

RAJAN, D.; CHAUDHURI, S. An MRF-based approach to generation of super-resolution images from blurred observations. **Journal of Mathematical Imaging and Vision**, v. 16, n. 1, p. 5-15, Jan 2002.

RAO, K. R.; YIP, P. **Discrete cosine transform: algorithms, advantages, applications**. Academic Press Professional, 1990.

REIS, S. R. S.; BRESSAN, G. Image Super-Resolution Using a Hybrid Scheme with DCT Interpolation and Sparse Representation Method.Em: XXXI Simpósio Brasileiro de Telecomunicações - SBrT 2013, **Anais...** Fortaleza, Brasil 2013a.

_____. Image super-resolution using DCT interpolation and sparse learning-based method. In: V International **Conference** on Digital Image Processing - ICDIP 2013. Beijing. 8878: 88784C-88784C-4 p. 2013b.

_____. **Método de Super-Resolução de Imagem Digital usando Interpolação DCT e Representação Esparsa, II Workshop de Pós Graduação da Área de Concentração Engenharia de Computação do PPGEE-EPUSP**, 10 de outubro de 2013, São Paulo-SP, Brasil. 2013c.

RODRIGUEZ, D. Z.; BRESSAN, G. Video quality assessments on digital tv and video streaming services using objective metrics. **IEEE Transacions Latin America**, v. 10, n. 1, p. 1184- 1189, 2012.

SANG-JUN, P.; JECHANG, J. Hybrid image upsampling method in the discrete cosine transform domain. **IEEE Transactions on Consumer Electronics**, v. 56, n. 4, p. 2615-2622, 2010.

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.

SEGALL, C. A.; MOLINA, R.; KATSAGGELOS, A. K. High-resolution images from low-resolution compressed video. **IEEE Signal Processing Magazine**, v. 20, n. 3, p. 37-48, 2003.

SHEN, H. F., ZHANG, L. P., HUANG, B., LI, P.X. A MAP approach for joint motion estimation, segmentation, and super resolution. **IEEE Transactions on Image Processing**, v. 16, n. 2, p. 479-490, 2007.

SHEN, H.; NG, M. K.; LI, P.; ZHANG, L., Super-Resolution Reconstruction Algorithm To MODIS Remote Sensing Images, **Computer Journal**, v. 52, i. 1, p. 90-100, 2009.

STARK, H.; OSKOU, P. High-Resolution Image Recovery From Image-Plane Arrays, Using Convex Projections. **Journal of the Optical Society of America a-Optics Image Science and Vision**, v. 6, n. 11, p. 1715-1726, 1989.

SULLIVAN, G. J.; WIEGAND, T. Video compression - from concepts to the h.264/avc standard. Proceedings of the IEEE, **Proceedings**, v. 93, n. 1, p. 18-31, 2005.

SUN, D.; CHAM, W.-K. Postprocessing of low bit-rate block DCT coded images based on a fields of experts prior. **IEEE Transactions on Image Processing**, v. 16, n. 11, Nov 2007.

UR, H.; GROSS, D. Improved resolution from subpixel shifted pictures. CVGIP: Graph. Models Image Process., **Conference**, v. 54, n. 2, p. 181-186, 1992.

VAN OUWERKERK, J. D. Image super-resolution survey. **Image and Vision Computing**, v. 24, n. 10, p. 1039-1052, Oct 1 2006.

VOGEL, C. R. **Computational Methods for Inverse Problems**. SIAM, 2002. (Frontiers in Applied Mathematics).

WANG, J.; ZHU, S.; GONG, Y. Resolution enhancement based on learning the sparse association of image patches. **Pattern Recognition Letters**, v. 31, n. 1, p. 1-10, Jan 1 2010.

WANG, X. G.; TANG, X. O. Hallucinating face by eigentransformation. **IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews**, v. 35, n. 3, p. 425-434, 2005.

WANG, Z., BOVIK, A. C., SHEIKH, H. R., SIMONCELLI, E. P. Image quality assessment: From error visibility to structural similarity. **IEEE Transactions on Image Processing**, v. 13, n. 4, p. 600-612, 2004.

WIEGAND, T., SULLIVAN, G. J., BJONTEGAARD, G., LUTHRA, A. et al. Overview of the H.264/AVC video coding standard. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 13, n. 7, p. 560-576, 2003.

WRIGHT, J., MA, Y., MAIRAL, J., SAPIRO, G., HUANG, T. S., SHUICHENG, Y. Sparse Representation for Computer Vision and Pattern Recognition. In: Proceedings of the IEEE, **Proceedings**, v. 98, n. 6, p. 1031-1044, 2010.

WU, Z.; YU, H.; CHEN, C. W. A New Hybrid DCT-Wiener-Based Interpolation Scheme for Video Intra Frame Up-Sampling. **IEEE Signal Processing Letters**, v. 17, n. 10, 2010.

YANG, J. WRIGHT, J., HUANG, T. S., MA, Y. Image Super-Resolution Via Sparse Representation. **IEEE Transactions on Image Processing**, v. 19, n. 11, Nov 2010.

YANG, S. WANG, M., CHEN Y., SUN Y. Single-Image Super-Resolution Reconstruction via Learned Geometric Dictionaries and Clustered Sparse Coding. **IEEE Transactions on Image Processing**, v. 21, n. 9, p. 4016-4028, Sep 2012.

ZEYDE, R.; ELAD, M.; PROTTER, M. On single image scale-up using sparse-representations. In: 7th international conference on Curves and Surfaces. **Proceedings**, Avignon, France: Springer-Verlag 2012.

ZHANG, L.; XIAOLING, W. An Edge-guided image interpolation algorithm via directional filtering and data fusion. **IEEE Transactions on Image Processing**, v. 15, n. 8, p. 2226-2238, 2006.

ZHANG, W.; CHAM, W.-K. Hallucinating Face in the DCT Domain. **IEEE Transactions on Image Processing**, v. 20, n. 10, 2011.

ZHOU, F.; YANG, W.; LIAO, Q. **Single Image Super-Resolution Using Incoherent Sub-dictionaries Learning**. **IEEE Transactions on Consumer Electronics**, v. 58, n. 3, p. 891-897, 2012.

ZIBETTI, M. V. W. Super-Resolução Simultânea para Sequência de Imagens. 2007, 167p Tese (Doutorado) - Engenharia Elétrica. Universidade Federal de Santa Catarina- UFSC, Florianópolis, 2007.

ZIBETTI, M. V. W.; BAZAN, F. S. V.; MAYER, J. Estimation of the parameters in regularized simultaneous super-resolution. **Pattern Recognition Letters**, v. 32, n. 1, p. 69-78, 2011.

APÊNDICE A – Normas Vetoriais

Este apêndice apresenta algumas informações complementares sobre as principais normas utilizadas nesta tese. Para mais detalhes, sugerimos consultar as referências (HEATH, 2002; VOGEL, 2002) ou referências complementares.

1) Definição: Seja V um espaço vetorial real ou complexo. Uma norma vetorial em V é uma função $|\cdot|: V \rightarrow \mathcal{R}$ que satisfaz as seguintes propriedades:

- a) $|x| > 0$ para todo $x \neq 0$ e $|x| = 0$ se $x = 0$;
- b) $\|\alpha x\| = |\alpha| \|x\|$ para todo $x \in V$ e para todo $\alpha \in \mathcal{R}$;
- c) $\|x + y\| \leq \|x\| + \|y\|$ para todo $x, y \in V$;

2) Norma l_1 (norma da soma):

$$\|A\|_1 = \sum_{i,j=1}^n |a_{i,j}| \quad (\text{A.1})$$

3) Norma l_2 (norma euclidiana):

$$\|A\|_2 = \left(\sum_{i,j=1}^n |a_{i,j}|^2 \right)^{1/2} \quad (\text{A.2})$$

4) Norma l_∞ (norma euclidiana):

$$\|A\|_\infty = \max_{1 \leq i \leq n} |x_i| \quad (\text{A.3})$$

APÊNDICE B – Argumentos máximos e mínimos de uma função

Este apêndice apresenta algumas informações complementares sobre Argumentos máximos e mínimos de uma função. Para mais detalhes, sugerimos consultar as referências (HEATH, 2002; VOGEL, 2002) ou informações complementares.

Um problema de otimização pode ser representado da seguinte forma:

1) Dada uma função $f : A \rightarrow \mathbb{R}$ de algum conjunto A de números reais,

2) Busca-se:

2.1) um elemento x_0 em A tal que $f(x_0) \leq f(x)$ para todo x em A (**minimização**), ou,

2.2) um elemento x_0 em A tal que $f(x_0) \geq f(x)$ para todo x em A (**maximização**).

Onde A é algum subconjunto do espaço euclidiano $\mathbb{R}^n$ especificado por um conjunto de *restrições*, igualdades ou desigualdades que os membros de A devem satisfazer.

O domínio A de f é chamado de espaço de busca ou o conjunto de escolha, enquanto os elementos de A são chamados de soluções candidatas ou soluções viáveis. A função f é chamada de função de custo (**minimização**) ou função utilidade (**maximização**).

.3) Valor mínimo e máximo de uma função

Considere a seguinte notação:

$$\min_{x \in \mathbb{R}} (x^2 + 1) \tag{B.1}$$

A expressão (B.1) denota o valor mínimo de uma função objetivo $x^2 + 1$, ao escolher x de um conjunto de números reais $\mathbb{R}$. O valor mínimo neste caso é 1, ocorrendo em $x = 0$.

4) Argumentos ótimos de variáveis

Notação:

$$\min_{x \in (-\infty, -1]} (x^2 + 1) \quad (\text{B.2})$$

ou de forma equivalente

$$\arg \min_x x^2 + 1 \quad \text{sujeito a } x \in (-\infty, -1] \quad (\text{B.3})$$

As expressões (B.2) e (B.3) representam o valor (ou valores) do argumento x no intervalo $(-\infty, -1]$ que minimiza (ou minimizam) a função objetivo $x^2 + 1$.