UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
Departamento de Engenharia Elétrica

Leandro Luiz de Almeida

**Construção automática de imagens de super-resolução a partir
de mosaicos formados por sequências de imagens**

Tese apresentada à Escola de Engenharia de São Carlos da
Universidade de São Paulo, como parte dos requisitos para
obtenção do título de Doutor em Ciências, Programa de
Engenharia Elétrica.

Área de Concentração: Processamento de Sinais e
Instrumentação
Orientadora: Prof^a. Dr^a. Maria Stela Veludo de Paiva

São Carlos
2013

Trata-se da versão corrigida da tese. A versão original se encontra disponível na EESC/USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica.

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTA TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

A447c Almeida, Leandro Luiz
Construção automática de imagens de super-resolução
a partir de mosaicos formados por sequências de imagens
/ Leandro Luiz Almeida; orientadora Maria Stela Veludo
Paiva. São Carlos, 2013.

Tese (Doutorado) - Programa de Pós-Graduação em
Engenharia Elétrica e Área de Concentração em
Processamento de Sinais e Instrumentação -- Escola de
Engenharia de São Carlos da Universidade de São Paulo,
2013.

1. Sequência de Imagens. 2. Mosaico. 3.
Super-resolução. 4. Imagens aéreas. 5. SIFT. I. Título.

FOLHA DE JULGAMENTO

Candidato: Bacharel **LEANDRO LUIZ DE ALMEIDA**.

Título da tese: "Construção automática de imagens de super-resolução a partir de mosaicos formados por sequências de imagens".

Data da defesa: 30/09/2013

Comissão Julgadora:

Resultado:

Profa. Dra. **Maria Stela Veludo de Paiva (Orientadora)**
(Escola de Engenharia de São Carlos/EESC)

APROVADO

Prof. Dr. **Ricardo Luis Barbosa**
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP- Sorocaba)

Aprovado

Dr. **Lúcio André de Castro Jorge**
(EMBRAPA)

APROVADO

Prof. Dr. **Almir Olivette Artero**
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP- Presidente Prudente)

APROVADO

Prof. Dr. **Celso Aparecido de França**
(Universidade Federal de São Carlos/UFSCar)

APROVADO

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica e Presidente da Comissão de Pós-Graduação:
Prof. Titular **Denis Vinicius Coury**

Dedico este trabalho primeiramente a Deus, por ter me concedido a oportunidade, a capacidade e sabedoria para a realização deste.

À Elisane (esposa) e, especialmente à Luíse, minha filha, que é a motivação maior para tudo.

À minha querida Família: Jorge (Pai), Vera (Mãe), Fernando (Irmão) e Belisa (Irmã), que sempre acreditou e me incentivou a não desistir de alcançar aquilo que se deseja.

Agradecimentos

À professora orientadora Dra. Maria Stela Veludo de Paiva, que confiou em mim, demonstrando sempre muita compreensão e generosidade.

Ao meu Amigo Francisco Assis da Silva (Chico) pelo companheirismo, incentivo, ideias e parcerias durante todo o período do doutoramento.

Aos Amigos Claudiney (Mastiga), Emerson Dória, Flávio Pandur, Franz Sérgio, Haroldo, Josemar, João Júnior, Mário Pazoti, Moacir, Robson Siscoutto, Ronaldo Oikawa e Silvio Carro pelo sempre apoio e comentários importantíssimos em momentos cruciais do trabalho.

E, em especial, as diversas pessoas e familiares que sempre souberam exprimir sentimentos verdadeiros de encorajamento em momentos turbulentos.

Ao aluno de Ciência da Computação da Faculdade de Informática de Presidente Prudente (FIPP) Thiago Elias Vicente, que colaborou em vários momentos do processo de doutoramento, participando de trabalhos de iniciação científica vinculados ao tema.

Ao colega de grupo de pesquisa André Tarallo pelo grande companheirismo e colaboração demonstrados.

Aos meus amigos professores colegas de trabalho da Faculdade de Informática de Presidente Prudente que sempre me incentivaram para a realização deste trabalho.

“Sonho que se sonha só,
É só um sonho que se sonha só,
Mas sonho que se sonha junto é realidade.”

Raul Seixas

“... Saber que se pode, querer que aconteça,
Esquecer os medos, jogá-los pra fora,
Pintar sua cara com a cor da esperança,
Tentar o futuro com o coração.

É melhor perde-se que nunca se achar,
Melhor tentar, que deixar de tentar,
Você vai ver seu sonho se realizar.

Sei que o impossível se pode mudar,
E que a tristeza um dia irá,
E assim será, a vida muda e mudará.”

Diego Torres

Resumo

ALMEIDA, L. L. **Construção automática de imagens de super-resolução a partir de mosaicos formados por sequências de imagens**. 2013. 180 f. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo. São Carlos, 2013.

As técnicas de super-resolução possibilitam combinar várias imagens de uma mesma cena para se obter uma imagem com resolução radiométrica e geométrica aumentada, denominada de imagem de super-resolução. Nessa imagem são realçadas características importantes possibilitando recuperar detalhes e informações. As aplicações envolvem diferentes áreas, tais como: na agricultura para identificar possíveis desmatamentos e controle de pragas, na área médica para a detecção de doenças em estágios iniciais, identificação facial de pessoas suspeitas em imagens de circuito fechado, reconstrução de filmes, identificação de placas de veículos, entre outras. No presente trabalho, é proposta uma metodologia para a geração de imagens de super-resolução a partir de uma região selecionada de um mosaico. Embora existam vários trabalhos publicados relacionados à geração de imagens de super-resolução, as metodologias não se aplicam para uma região específica do mosaico. E grande parte dos trabalhos utiliza uma imagem de referência, a partir da qual é gerada a imagem de super-resolução. Na metodologia proposta, inicialmente, é gerado um mosaico a partir de um conjunto de imagens baseando-se nos algoritmos SIFT ou SURF, BBF e RANSAC e é criada uma estrutura de dados, que organiza os pontos correlacionáveis das imagens com sobreposição, facilitando e simplificando o processo de fusão desses pontos para a obtenção da imagem de super-resolução. A ferramenta implementada a partir dessa metodologia, possibilita ao operador selecionar a região de interesse no mosaico, a partir da qual, é gerada a imagem de super-resolução utilizando as técnicas SIFT (ou SURF), interpolação Bicúbica e a fusão pelo valor mediano dos pontos da área com sobreposição das imagens da sequência. Para validar a metodologia, foram utilizados quatro conjuntos de imagens, que incluem imagens simuladas, obtidas com câmeras de baixa e alta resolução, imagens aéreas de áreas urbanas e rurais, coloridas e em escalas de cinza, e imagens contendo elementos textuais. Nas imagens simuladas foram adicionados ruídos e avaliada a imagem de super-resolução gerada por meio de duas métricas: raiz do erro médio quadrático (RMSE) e o índice de similaridade estrutural (SSIM). Os resultados mostraram que mesmo com valor de RMSE elevado, o SSIM foi acima de 70%, evidenciando o alto grau de similaridade. As imagens de super-resolução obtidas a partir de uma região dos mosaicos gerados foram comparadas com imagens superamostradas por meio de interpolações e avaliadas confrontando as imagens extrapoladas para verificação visual dos elementos da cena. Os resultados apresentados concluem que as imagens de super-resolução geradas, apresentam melhorias no que diz respeito à restauração das mesmas para futura análise de alvos de interesse, sem ter o retrabalho de adquirir novas imagens da cena, pois dependendo da cena analisada não será possível nova aquisição. O presente trabalho contribui com a geração de imagem de super-resolução, a partir de uma região do mosaico e com estruturas de dados e algoritmos que possibilitam a análise de regiões específicas do mosaico, sem que o mesmo tenha que ser processado integralmente.

Palavras-chave: Sequência de Imagens; Mosaico; Super-resolução; Imagens aéreas; SIFT.

Abstract

ALMEIDA, L. L. **Automatic construction of super-resolution images from mosaics formed by sequences of images**. 2013. 180 f. Dr. Thesis – Escola de Engenharia de São Carlos (EESC), Universidade de São Paulo. São Carlos, 2013.

Super-resolution techniques allow combining several images of the same scene in order to obtain an image with increased geometric and radiometric resolution, called super-resolution image. In this image are enhanced features allowing to recover important details and information. The applications involve different areas, such as: in the agriculture to identify possible deforestation and pest control, in the medical area to detect diseases in early stages, facial identification in images of suspects in closed loop, movies reconstruction, license plates recognition, among others. In this work, we propose a methodology for generating super-resolution images from a selected region of a mosaic. Although there are several published papers related to the generation of super-resolution images, the existing methodologies do not apply to a specific region of the mosaic. And the majority of studies use a reference image, from which is generated the super-resolution image. In the proposed methodology, initially, a mosaic is generated from a set of images based on the algorithms SIFT or SURF, BBF and RANSAC and creates a data structure that organizes the points correlate of the images with overlapping, facilitating and simplifying the fusion process of these points to obtain the super-resolution image. The tool implemented from this methodology allows to the operator to select the region of interest in the mosaic, from which is generated the image using super-resolution techniques SIFT (or SURF), Bicubic interpolation and fusion process by the median value of the points with overlapping area from the images of the sequence. In order to validate the methodology, we used four sets of images, including simulated images taken with cameras of low and high resolution, aerial images of urban and rural areas, color and grayscale images and images containing texts. In the simulated images were added noise and were evaluated the super-resolution image generated by two metrics: root mean square error (RMSE) and the structural similarity index (SSIM). The results showed that even with high RMSE value, the SSIM was above 70%, reflecting the high degree of similarity. The super-resolution images obtained from a region of the mosaics were compared with images generated by super-sampled interpolation and evaluated by comparing the images extrapolated to visual inspection of elements of the scene. From the results it can be concluded that the super-resolution images generated present improvements with regard to restoration of images for further analysis of targets of interest, without reworking to acquire new images of the scene, because depending on the analyzed scene it would not be possible a new acquisition. This work contributes to the generation of super-resolution image from a region of the mosaic and with data structures and algorithms which enable the analysis of specific regions of the mosaic without it has to be fully processed.

Key words: Image Sequence; Mosaic; Super-resolution; Aerial Images; SIFT.

Lista de Figuras

Figura 2.1: Projeção dos pontos correlacionáveis das imagens adquiridas sobre a imagem de super-resolução.....	32
Figura 2.2: Etapas para geração da imagem de super-resolução.....	33
Figura 2.3: Reamostragem: (a) Pixels na grade original; (b) Posição dos pixels transformados sobrepostos na grade original; (c) Ajustamento da grade transformada sobre a grade original. Fonte: (ATKINSON et al., 1996).....	36
Figura 2.4: Interpolação pelo Vizinho Mais Próximo.	37
Figura 2.5: Representação da sub-imagem.....	37
Figura 2.6: Interpolação Bilinear. Fonte: Adaptado de Richards (2013).	39
Figura 2.7: Interpolação Bicúbica. Fonte: Adaptado de Richards (2013).....	39
Figura 3.1: Representação do processo para construção das imagens resultantes da diferença de Gaussianas DoG em cada oitava no espaço escala. Fonte: Adaptado de Lowe (2004).	47
Figura 3.2: Exemplo do processo de construção de uma oitava.....	48
Figura 3.3: Pirâmide de imagens filtradas pelas Gaussianas mostrando o processo de construção das oitavas. Fonte: Adaptado de Heckel (2008).....	49
Figura 3.4: Exemplo do processo de detecção de extremos (máximos ou mínimos) das imagens resultantes da diferença Gaussiana. Fonte: Adaptado de Lowe (2004).	50
Figura 3.5: Resultado dos experimentos para determinar a escolha do número de amostras em escalas vizinhas por oitava para detectar pontos chave. Fonte: (LOWE, 2004).	51
Figura 3.6: Resultado do experimento para determinação do valor prévio de σ para cada oitava. Fonte: (LOWE, 2004).	52
Figura 3.7: Gradientes da imagem e descritor de pontos chave. Fonte: (LOWE, 2004).....	57
Figura 3.8: Exemplo de uma árvore k-d tree. Na esquerda o modo que o plano é subdividido e à direita a correspondente árvore binária. Fonte: (BERG et al., 2008).	61
Figura 3.9: Exemplo de busca BBF em uma árvore k-d tree com 8 pontos. Fonte: (BEIS; LOWE, 1997).	63
Figura 3.10: Estimação de uma reta ajustada a um conjunto de pontos 2D. Fonte: (HARTLEY; ZISSERMAN, 2003).	65
Figura 3.11: Pseudocódigo do algoritmo RANSAC para calcular um melhor ajuste na correspondência dos descritores SIFT. Fonte: (SILVA, 2012).	69
Figura 3.12: Imagem “box.png” detectada na imagem “box_in_cene.png” com 45 pontos com extração dos <i>inliers</i> pelo algoritmo RANSAC.	70
Figura 3.13: Imagem de desenho detectada com 39 pontos (<i>inliers</i>) extraídos pelo algoritmo RANSAC.....	71

Figura 3.14: <i>Matching</i> entre duas imagens da mesma cena.....	72
Figura 3.15: Determinação de pontos chave correspondentes com extração dos <i>inliers</i> pelo algoritmo RANSAC.....	72
Figura 3.16: Da esquerda para direita: Derivadas de segunda ordem de uma Gaussiana, discretizadas, e suas respectivas aproximações utilizadas. Fonte: (BAY et al., 2008)...	73
Figura 3.17: Integral da Imagem. Fonte: (BAY et al., 2008).....	75
Figura 3.18: <i>Haar Wavelets</i> d/dx e d/dy . Fonte: (BAY et al., 2008).....	76
Figura 4.1: Os quatro passos para geração de imagens de super-resolução. Fonte: (BAKER; KANADE, 1999).	82
Figura 4.2: Processo geral do algoritmo do fluxo ótico na determinação de imagens de super-resolução. Fonte: (BAKER; KANADE, 1999).....	83
Figura 4.3: Redução da dimensão dos descritores do SIFT.....	88
Figura 5.1: Visão geral das etapas de geração de uma imagem de super-resolução.....	103
Figura 5.2: Processo de retificação das imagens.....	104
Figura 5.3: Fusão entre os pixels das imagens que pertencem à cena.	105
Figura 5.4: Exemplo de processo para geração de imagens superamostradas através de Interpolação Bilinear.....	106
Figura 5.5: Visão geral da metodologia proposta.	110
Figura 5.6: Listas Dinâmicas Cruzadas para armazenamento dos pixels das imagens que pertencem ao mosaico.....	111
Figura 5.7: Exemplo da organização de um arquivo de dados.	114
Figura 5.8: Tela principal da ferramenta desenvolvida.....	119
Figura 5.9: Algoritmos de detecção e extração dos descritores disponibilizados pela biblioteca OpenCV versão 2.4.2.	119
Figura 5.10: Tela com a imagem de super-resolução gerada.....	120
Figura 6.1: Cena utilizada para adquirir uma sequência de imagens a partir de operações de translações, rotações e escalas na câmera.	122
Figura 6.2: Processo de comparação da imagem gerada (12 megapixels) com a imagem adquirida (12 megapixels).....	122
Figura 6.3: Processo geral, que envolve a simulação de uma sequência de imagens e validação da imagem de alta resolução gerada.	123
Figura 6.4: Processo de geração para comparação da imagem construída (12 megapixels) com a imagem adquirida (12 megapixels).	123
Figura 6.5: Conjunto de imagens de 3 megapixels adquiridas após aplicação de translação, rotação e escala à câmera.	124
Figura 6.6: Correspondências encontradas por meio do algoritmo SIFT entre a imagem de referência e a imagem obtida.	125
Figura 6.7: Correspondências encontradas por meio do algoritmo SURF entre a imagem de referência e a imagem obtida	126

Figura 6.8: Resultado obtido após processamento das imagens da Figura 6.5.	127
Figura 6.9: Visualização de parte da imagem de 12 megapixels adquirida e da imagem de super-resolução gerada (Figura 6.8).	128
Figura 6.10: Diferença entre a imagem de 12MP e a imagem gerada com uso do SIFT.	128
Figura 6.11: Diferença entre a imagem de 12MP e a imagem gerada com uso do SURF.	129
Figura 6.12: Diferença entre a imagem gerada com o uso do SIFT e a imagem gerada com uso do SURF.	129
Figura 6.13: Sequência de imagens originais adquiridas com a câmera digital Kodak DC-210.	131
Figura 6.14: Imagem de super-resolução a partir das imagens da Figura 6.13.	132
Figura 6.15: Diferença entre a imagem gerada com uso do SIFT e a imagem original com a raiz do erro médio quadrático e o índice de similaridade.	133
Figura 6.16: Diferença entre a imagem gerada com uso do SURF e a imagem original com a raiz do erro médio quadrático e o índice de similaridade.	134
Figura 6.17: Sequência de imagens originais adquiridas com a câmera digital Kodak DC-210 com aplicação de ruídos SHARP2.	135
Figura 6.18: Imagem resultante do registro e fusão da sequência evidenciada na Figura 6.17.	137
Figura 6.19: Diferença entre a imagem gerada (Figura 6.18), com uso do SIFT, e a imagem original (Figura 6.17(d)) com a raiz do erro médio quadrático e o índice de similaridade estrutural.	137
Figura 6.20: Diferença entre a imagem gerada, com uso do SURF, e a imagem original (Figura 6.17(d)) com a raiz do erro médio quadrático e o índice de similaridade estrutural.	138
Figura 6.21: Imagens que compõem uma cena com ruídos coloridos.	139
Figura 6.22: Imagem de super-resolução gerada, com uso do SIFT, a partir da sequência da Figura 6.21.	140
Figura 6.23: Diferença entre a imagem gerada, com uso do SIFT, e a imagem original com o erro médio quadrático e o índice de similaridade.	141
Figura 6.24: Diferença entre a imagem gerada, com uso do SURF, e a imagem original com o erro médio quadrático e o índice de similaridade.	142
Figura 6.25: Sequência de imagens originais adquiridas com a câmera digital Kodak DC-40.	143
Figura 6.26: Imagem resultante do registro da sequência evidenciada na Figura 6.25.	144
Figura 6.27: Diferença entre a imagem gerada, com uso do SIFT, e a imagem original com o erro médio quadrático e o índice de similaridade.	144
Figura 6.28: Diferença entre a imagem gerada com uso do SURF, e a imagem original com o erro médio quadrático e o índice de similaridade.	145
Figura 6.29: Sequência de imagens aéreas para geração do mosaico.	147
Figura 6.30: Imagem do mosaico gerado a partir do conjunto de imagens da Figura 6.29.	148

Figura 6.31: Imagem de super-resolução gerada a partir da seleção da região escolhida com uso do SIFT.	148
Figura 6.32: Imagem de super-resolução gerada a partir da seleção da região escolhida com uso do SURF.	149
Figura 6.33: Comparação da (a) região maximizada do mosaico (Figura 6.30) com (b) região maximizada da imagem de super-resolução obtida (Figura 6.31).	149
Figura 6.34: Sequência de imagens aéreas para geração do mosaico.	150
Figura 6.35: Mosaico obtido pelas imagens contidas na Figura 6.34.	151
Figura 6.36: Região selecionada na Figura 6.35 para geração da imagem de super-resolução.	151
Figura 6.37: Imagem de Super-resolução gerada a partir da seleção utilizando-se do SIFT.	152
Figura 6.38: Imagem de Super-resolução gerada a partir da seleção utilizando-se do SURF.	152
Figura 6.39: Comparação da (a) região maximizada do mosaico (Figura 6.35) com (b) região maximizada da imagem de super-resolução obtida (Figura 6.37).	153
Figura 6.40: Sequência de imagens aéreas em tons de cinza para geração do mosaico.	154
Figura 6.41: Mosaico obtido pelas imagens contidas na Figura 6.40.	155
Figura 6.42: Imagem de Super-resolução gerada a partir da seleção na Figura 6.41 utilizando-se do SIFT.	155
Figura 6.43: Imagem de Super-resolução gerada a partir da seleção na Figura 6.41 utilizando-se do SURF.	156
Figura 6.44: Comparação da (a) região maximizada do mosaico (Figura 6.41) com (b) região maximizada da imagem de super-resolução obtida (Figura 6.42).	156
Figura 6.45: Sequência de imagens com elementos textuais.	158
Figura 6.46: Resultado da seleção e fusão da área com elementos textuais.	159
Figura 6.47: Imagem adquirida com 12 megapixels com elementos textuais.	160
Figura 6.48: Parte da imagem de Super-resolução gerada a partir da seleção na Figura 6.46 utilizando-se do SIFT.	160
Figura 6.49: Parte da imagem de 12 megapixels adquirida, conforme Figura 6.47.	161
Figura 6.50: Regiões de imagens que contêm os caracteres avaliados na primeira parte do experimento: a) de alta resolução (HR), b) uma das 15 imagens de baixa resolução, usadas para obter uma imagem de super-resolução, c) imagem de super-resolução (SR) obtido com o método proposto.	161
Figura 6.51: Regiões de imagens que contêm os caracteres avaliados na segunda parte do experimento: a) de alta resolução (HR), b) uma das 15 imagens de baixa resolução, usadas para obter uma imagem de super-resolução, c) imagem de super-resolução (SR) obtido com o método proposto.	162

Lista de Tabelas

Tabela 1: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens em relação à imagem de referência Imagem5.jpg.	125
Tabela 2: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.13) em relação à imagem de referência Dcp01615.bmp.	132
Tabela 3: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.17) em relação à imagem de referência SHARP2IBSubDCP01615.bmp.	136
Tabela 4: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.20) em relação à imagem de referência Dcp01615.bmp.	140
Tabela 5: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.25) em relação à imagem de referência Imagem02.bmp.	143
Tabela 6: Resumo dos resultados obtidos pelas métricas de avaliação.	145
Tabela 7: Número de caracteres incorretamente identificadas nas 15 imagens de baixa resolução, super-resolução (SR) e imagens de alta resolução (HR).	162
Tabela 8: Número de caracteres incorretamente identificados nas 15 imagens de baixa resolução, super-resolução (SR) e imagens de alta resolução (HR).	163

Lista de Abreviaturas e Siglas

BBF:	<i>Best-Bin-First</i>
CCD:	<i>Charge Coupled Device</i>
DoG:	<i>Difference of Gaussian</i>
FBI:	<i>Federal Bureau of Investigation</i>
HR:	<i>High Resolution</i>
LR:	<i>Low Resolution</i>
MDI:	Múltiplas Instâncias de Janelas
OpenCV:	<i>Open Source Computer Vision Library</i>
PCA:	<i>Principal Component Analysis</i>
PCA-SIFT:	<i>Principal Component Analysis Scale Invariant Feature Transform</i>
PC Card:	<i>Picture Card</i>
PCMCIA:	<i>Personal Computer Memory Card International Association</i>
POCS:	Projeção em Conjuntos Convexos
RANSAC:	<i>RANdom SAmple Consensus</i>
RGB:	<i>Red, Green, Blue</i>
RMSE:	<i>Root Mean Square Error</i>
SCV:	Sinal Composto de Vídeo
SIFT:	<i>Scale Invariant Feature Transform</i>
SR:	Super-resolução
SSIM	<i>Structural SIMilarity</i>
SURF:	<i>Speeded Up Robust Features</i>
VISAR:	<i>Video Image Stabilization and Registration</i>

Sumário

Resumo	11
Abstract	13
Lista de Figuras	15
Lista de Tabelas	19
Lista de Abreviaturas e Siglas.....	21
1 Introdução.....	25
1.1 Considerações Iniciais	25
1.2 Justificativas	28
1.3 Objetivos	29
1.4 Contribuição	29
1.5 Estrutura do Trabalho	30
2 Super-resolução: Definições e Procedimentos Envolvidos.....	31
2.1 Considerações Iniciais	31
2.2 Técnicas de super-resolução.....	31
2.3 Processo de Reamostragem	34
2.4 Técnicas de interpolação	35
2.4.1 Interpolação pelo vizinho mais próximo.....	37
2.4.2 Interpolação utilizando a média	37
2.4.3 Interpolação bilinear.....	38
2.4.4 Interpolação bicúbica	39
2.5 Considerações Finais	40
3 Técnicas utilizadas para o registro de imagens automatizado	41
3.1 Considerações Iniciais	41
3.2 Introdução.....	41
3.3 Algoritmo SIFT	44
3.3.1 Detecção de extremos no espaço escala	45
3.3.2 Localização exata dos pontos chave.....	52
3.3.3 Determinação de Orientação dos Descritores	55

3.3.4	Descritor local.....	56
3.4	Correspondência dos descritores SIFT	58
3.4.1	Árvore <i>k-d tree</i>	60
3.4.2	Algoritmo de busca BBF	62
3.5	Extração dos <i>inliers</i>	64
3.5.1	Algoritmo RANSAC aplicado ao ajuste na correspondência dos descritores SIFT	68
3.5.2	Aplicação para detecção e reconhecimento de objetos.....	70
3.6	Algoritmo SURF.....	73
3.7	Considerações Finais	76
4	Trabalhos Relacionados	79
4.1	Considerações Iniciais	79
4.2	Trabalhos Relacionados à construção de imagens de melhor resolução	79
4.3	Trabalhos que realizam o registro de imagens usando o algoritmo SIFT para a reconstrução de imagens de melhor resolução	87
4.4	Trabalhos Relacionados à construção de mosaicos	90
4.5	Considerações Finais	96
5	Metodologia de geração de imagens de super-resolução.....	99
5.1	Considerações Iniciais	99
5.2	Descrição das imagens utilizadas.....	99
5.2.1	Conjunto A.....	100
5.2.2	Conjunto B.....	101
5.2.3	Conjunto C.....	101
5.2.4	Conjunto D.....	102
5.3	Metodologia geral aplicada ao trabalho	102
5.4	Etapas para a geração das imagens de super-resolução a partir de dados simulados.....	107
5.5	Etapas para a geração das imagens de super-resolução a partir de dados reais com elementos textuais	109
5.6	Etapas para a geração das imagens de super-resolução a partir de mosaicos.....	109
5.6.1	Implementação computacional	116
5.7	Ferramenta Desenvolvida: exemplo de uso	118
5.8	Considerações Finais	120
6	Resultados e Discussões	121

6.1	Avaliação dos Resultados.....	121
6.2	Experimentos com dados simulados – Conjunto B.....	122
6.3	Experimentos com dados reais	130
6.3.1	Experimento 1 – Sequência de imagens com dados reais sem adição de ruídos (Conjunto A2).....	130
6.3.2	Experimento 2 – Sequência de imagens com dados reais com aplicação de ruídos de contraste (Conjunto A2)	134
6.3.3	Experimento 3 – Sequência com dados reais e adição de ruídos aleatórios (Conjunto A2).....	138
6.3.4	Experimento 4 – Sequência com dados reais (Conjunto A1).....	142
6.3.5	Experimento 5 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de imagens aéreas urbanas (Conjunto C) ...	146
6.3.6	Experimento 6 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de imagens aéreas agrícolas (Conjunto C) .	150
6.3.7	Experimento 7 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de imagens aéreas em tons de cinza (Conjunto D)	153
6.3.8	Experimento 8 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de recortes de imagens com elementos textuais (Conjunto B)	156
6.4	Discussões	163
7	Conclusões e Sugestões	165
7.1	Sugestões de trabalhos futuros	168
7.2	Artigos Publicados e Submetidos.....	169
7.2.1	Artigos Publicados	169
7.2.2	Artigos Aceitos para Publicação	169
7.2.3	Participação como co-autor em Artigos Publicados	169
	Referências Bibliográficas	171

Capítulo 1

Introdução

1.1 Considerações Iniciais

O desenvolvimento de técnicas que permitem a geração de imagens de alta resolução a partir de uma sequência de imagens tem sido útil em várias aplicações, como para a identificação de placas de veículos, identificação facial de pessoas suspeitas (criminosos, por exemplo) em imagens de circuito fechado, na geração de imagens aéreas de alta resolução a partir de imagens de vídeo ou câmera fotográfica, na determinação de diagnósticos por meio de imagens médicas ou até mesmo a detecção de doenças ainda em estágios iniciais. Na astronomia pode revelar novos astros ou identificá-los melhor por meio de algoritmos aplicados a essas imagens, e também, na reconstrução de filmes antigos, onde se procura minimizar a quantidade de ruídos e borramentos que aparecem em cada *frame* projetado. E ainda na agricultura, possibilitando a verificação da área de interesse no que tange ao tipo de utilização da área, possíveis desmatamentos, controle de pragas entre outras aplicações.

A geração de imagens de alta resolução com base em uma sequência de imagens necessita primeiramente, que as imagens pertencentes à cena estejam retificadas, ou seja, o registro de imagens deve ter sido executado. Segundo Richards (2013), o registro de imagens compreende uma transformação geométrica que relaciona as coordenadas de uma imagem (linha, coluna) com coordenadas de um sistema de referência qualquer. Esse sistema de referência é, em última instância, o sistema de coordenadas planas de uma projeção cartográfica, ou ainda, uma outra imagem. E por isso, é considerado como a primeira e principal etapa nesse processo (LI; ZHOU, 2011). A partir disso, pode-se criar uma nova imagem com uma precisão maior e possível de ser utilizada para análises mais detalhadas.

Na realização do registro de imagens são determinados os pontos correspondentes da mesma para a geração de seus descritores, que são vetores de características de uma imagem ou de determinadas regiões de uma imagem e podem ser usados para se comparar regiões em imagens diferentes (SILVA, 2012).

Trabalhos recentes na literatura têm se concentrado em descritores invariáveis às transformações nas imagens, tais como translação, rotação e escala.

Exemplos de algoritmos mais recentes que incluem o processo de identificação de pontos de interesse e a geração de descritores e correspondência entre estes pontos, são os seguintes: o SIFT (Scale Invariant Feature Transform), proposto por Lowe (2004), o SURF (Speeded Up Robust Features), proposto por Bay et al. (2008) e suas respectivas variações, como o PCA-SIFT (Principal Component Analysis Scale Invariant Feature Transform) proposto por Ke e Sukthankar (2004), U-SURF (BAY et al., 2008) e o SURF-128 (BAY et al., 2008).

A seguir são descritos alguns trabalhos, apresentando um breve histórico a respeito da geração de imagens de melhor resolução.

Kim et al. (1990) por meio de uma aplicação em imagens de baixa resolução que apresentam ruídos, desenvolveram um algoritmo baseado na teoria de mínimos quadrados ponderados, para determinar o pixel ideal a ser utilizado na imagem de melhor resolução a ser gerada. O trabalho foi inicialmente motivado, naquela época, pela necessidade de processar imagens do satélite LANDSAT.

Peleg e Irani (1991), pioneiros pesquisadores nessa linha de pesquisa, sugeriram melhorias para o método apresentado por Peleg et al. (1987). Os autores tomaram o registro de imagens como base para uma boa determinação de uma imagem de alta resolução. Desenvolveram algoritmos iterativos para se obter um melhoramento na resolução das imagens com precisão de subpixel, tanto em imagens de baixa resolução em níveis de cinza como nas coloridas.

Ballard e Brown (1992) apresentaram uma metodologia para se realizar a restauração de imagens com base em sequências de imagens, que, a partir de uma imagem de referência buscam-se pontos nas imagens da sequência que melhor representa a coordenada verificada, obtendo-se assim, uma imagem de melhor resolução por meio de dados pertencentes à cena.

Hathaway e Meyer (1999), cientistas da *NASA's Marshall Space Flight (Huntsville, AL)*, desenvolveram um software denominado VISAR (*Vídeo Image Stabilization and Registration*), utilizado para melhorar a resolução de imagens. Foi idealizado com o intuito de se realizar análises de imagens científicas, vídeos da área médica, vídeos de circuitos fechados realizados por câmeras de segurança para identificação de intrusos, imagens que identificam motoristas infratores, aparelhos eletrônicos, além de uma série de outras situações que envolvem sequências de imagens.

O trabalho de Joyeux et al. (2001) aborda o processo de reconstrução de sequências de imagens degradadas com ênfase em restauração de filmes antigos em preto e branco. O

objetivo dos autores foi minimizar a quantidade de ruídos e riscos de linhas que aparecem em uma cena com base na sequência de imagens que compõe o filme.

Almeida e Tommaselli (2003) utilizam uma sequência de imagens da cena para melhorar a resolução de uma imagem. As imagens são registradas para então serem retificadas em relação a uma imagem de referência. Após a retificação, todas as imagens reamostradas da sequência estão em um mesmo sistema de coordenadas, então se executa o processo de fusão, gerando uma única imagem. Este processo foi realizado envolvendo as abordagens do método da média aritmética entre os níveis de brilho dos pixels e do método de minimização dos erros, a partir de uma imagem de alta resolução inicial aproximada. Com base nos resultados, observa-se que o método de minimização de erros gera uma sensível suavização nas regiões de alta frequência.

O algoritmo de super-resolução apresentado em Rubert et. al. (2004) baseia-se em amostras de treinamento para estimar detalhes em altas frequências nas imagens reamostradas. Este algoritmo é uma adaptação do método proposto por Freeman et. al. (2001) para operar com a componente de luminância Y do espaço de cores YUV e estimar as altas frequências perdidas no processo de reamostragem das imagens. Os resultados são comparados com a abordagem que utiliza o espaço de cores RGB. O algoritmo proposto apresentou vantagens devido à redução da dimensionalidade do conjunto de amostras de treinamento, consequentemente, diminuindo o tempo computacional da aplicação.

O trabalho de Tang et al. (2008) é destinado ao registro automático em imagens médicas de sequências microscópicas. Nesse trabalho, os autores propuseram um algoritmo que reduz a complexidade e as dimensões dos descritores SIFT. Assim, é possível a realização do *matching* entre as imagens da sequência com sobreposição para geração de imagens de melhor resolução.

Nemra e Aouf (2009) desenvolveram algoritmos eficientes e automatizados para geração de imagens de super-resolução vinculadas a mosaicos adquiridos por aeronave não tripulada.

O trabalho de Nasir et al. (2011) trata o problema de gerar imagens de alta resolução a partir de imagens de baixa resolução em três passos: registro da imagem, fusão da imagem baseada em SVD (*Singular Value Decomposition*) e interpolação.

Salari e Bao (2012) apresentam uma técnica de obtenção de uma imagem de super-resolução, com a melhoria do algoritmo de Papoulis-Gerchberg, que é aplicado no domínio da

frequência sobre imagens de baixa resolução. Os autores utilizam imagens simuladas, mas sugerem uma variedade de aplicações, incluindo sistemas de vigilância, imagens médicas e imagens de satélite.

Esses trabalhos evidenciam a grande aplicabilidade de imagens de alta resolução para uma análise detalhada de elementos contidos nas mesmas.

No presente trabalho é proposta uma metodologia para gerar uma imagem de super-resolução a partir de uma região selecionada de um mosaico. O mosaico é inicialmente construído a partir de imagens aéreas, adquiridas de forma a se ter região de sobreposição nas imagens. Uma região retangular desse mosaico pode ser selecionada pelo usuário. A região selecionada é então ampliada com base na sequência de imagens com sobreposição que formaram essa região. Essa ampliação é realizada pela fusão dos pontos reais que compõem a cena.

1.2 Justificativas

A necessidade de se recuperar elementos de uma região específica de uma cena composta por uma sequência de imagens é o que motivou a construção do presente trabalho, pois as imagens digitais, podem não apresentar resolução radiométrica e/ou geométrica satisfatórias para a análise desses elementos específicos (alvos de interesse), seja devido a fatores naturais como excesso de luz, ou até mesmo pelas características do mecanismo que está sendo utilizado para obter essas imagens.

Para o caso de mosaicos para a área agrícola é possível gerar imagens de alta-resolução para avaliar regiões que estejam possivelmente contaminadas com pragas, para a verificação de queimadas e mesmo o plantio de lavouras ilegais. A aplicação inclui outras áreas como a área de imagens médicas, imagens de circuitos de segurança e também imagens com elementos textuais.

A ferramenta implementada possibilita também o reaproveitamento de imagens de má qualidade, ou seja, tendo-se uma sequência de imagens com sobreposição, é possível que se obtenha outra com qualidade superior às que estão sendo analisadas buscando o elemento de melhor qualidade para aquela região de detalhe. O nível de brilho de cada ponto da imagem de super-resolução é determinado pela fusão dos pontos correspondentes nas outras imagens, todas transformadas (retificadas). Com isso, podem ser aproveitadas imagens que contenham cenários que não se repetirão, ou até mesmo eliminar ou minimizar problemas

físicos obtidos na construção de dispositivos, como por exemplo, deformação nas lentes, defeito do obturador, entre outros.

1.3 Objetivos

No presente trabalho é proposta uma metodologia para construir mosaicos a partir de imagens aéreas e possibilitar a seleção de uma região do mosaico, que é usada como referência para gerar uma imagem de super-resolução, ou seja, uma imagem ampliada geometricamente (tamanho) e radiometricamente (nitidez).

Como objetivos específicos pode-se citar:

- construção de mosaicos utilizando um algoritmo de extração de descritores de pontos chave (SIFT ou SURF);
- implementar uma ferramenta a partir da metodologia proposta que possibilite a seleção de algoritmos de detecção, de interpolação e de fusão que deverão ser utilizados durante o processamento;
- criar uma estrutura de dados que agilize o acesso aos pontos que devem ser utilizados para gerar a imagem de super-resolução;
- incluir na ferramenta recursos para se gerar uma imagem de super-resolução a partir de uma imagem de referência, escolhida de uma sequência de imagens;
- utilizar métricas para avaliar as imagens de super-resolução geradas.

1.4 Contribuição

Destaca-se como a principal contribuição deste trabalho a possibilidade de gerar imagens de super-resolução a partir de uma área selecionada em um mosaico (ou imagem panorâmica), construída a partir da alta correlação existente entre as imagens que compõem a cena geral. A imagem de super-resolução obtida é uma imagem gerada por meio de dados de múltiplas outras imagens, possuindo resolução radiométrica e geométrica aumentada, a partir dessa área destacada, sem influência e utilização de todo o mosaico construído. Esse processo é possível, em virtude do desenvolvimento de estruturas de dados e algoritmos que possibilitam a análise de regiões específicas do mosaico, sem que o mesmo tenha que ser processado integralmente, pois durante a primeira construção do mosaico, pontos

correlacionáveis entre as imagens da sequência, são organizados e armazenados para que o processamento de todos os elementos aconteça de forma otimizada.

Destacam-se então contribuições no que diz respeito ao desenvolvimento de novos elementos computacionais, tanto de processamento quanto de armazenamento de dados vinculados às imagens, que aprimoram a geração de imagens de super-resolução de forma automatizada e, principalmente, com desempenho mais robusto aos problemas de movimento. Isso possibilita também a utilização de imagens, mesmo que em resolução ruim, minimizando o custo e o retrabalho na reaquisição das mesmas.

Nos trabalhos da literatura relacionados à geração de imagens de super-resolução geralmente é escolhida uma imagem como referência, no conjunto de imagens utilizado, a partir da qual é gerada a imagem de super-resolução. Na ferramenta implementada neste trabalho esta opção também foi incluída. Portanto, é um trabalho que faz uso de técnicas existentes, mas propõe e implementa uma ideia de otimização de memória e processamento para obtenção de imagens de super-resolução.

1.5 Estrutura do Trabalho

No Capítulo 2 são apresentados conceitos básicos sobre super-resolução e o processo utilizado para a geração dessas imagens, que agrega a retificação e restauração das mesmas. No Capítulo 3, o foco principal está na determinação dos pontos de correspondência entre as imagens que compõem a cena, por isso, são detalhados os algoritmos mais utilizados nesse trabalho, o SIFT e o SURF, além dos procedimentos intermediários que possibilitam a realização do *matching* das mesmas. No Capítulo 4, são descritos vários trabalhos relacionados ao processo de construção de um mosaico de imagens e também da imagem de super-resolução. Já no Capítulo 5, são descritos os conjuntos de dados utilizados e a metodologia adotada para a elaboração desse trabalho. No Capítulo 6, encontram-se os experimentos mais relevantes, que geraram os resultados obtidos. E, finalmente, no Capítulo 7, são apresentadas as conclusões e o direcionamento para futuras pesquisas nessa linha.

Capítulo 2

Super-resolução: Definições e Procedimentos Envolvidos

2.1 Considerações Iniciais

Neste capítulo são descritas definições sobre técnicas de super-resolução e também sobre os procedimentos envolvidos para obtenção de imagens de super-resolução. Para isso, é necessário o entendimento sobre registro de imagens, para obtenção dos parâmetros utilizados na execução da correspondência dos pontos comuns entre as imagens que fazem parte do cenário analisado.

2.2 Técnicas de super-resolução

As técnicas de super-resolução proporcionam uma forma de obter imagens com resolução superior à fornecida pelo dispositivo de aquisição. Estas técnicas removem distorções causadas durante o processo de aquisição, realçam características importantes nas imagens, como as bordas dos objetos, e, ainda, recuperam detalhes pela utilização da informação contida em um grande conjunto de imagens capturadas. Enfim, combinam várias imagens de uma mesma cena com o intuito de obter uma ou várias imagens com resoluções radiométricas, que define o intervalo e o número aceitável de intensidades de brilho, necessários para representar uma imagem digital formada por pixels, e resolução geométrica aumentada, que representa o aumento de área física da imagem, conforme Figura 2.1. As imagens utilizadas possuem diferenças entre elas no que tange aos deslocamentos no nível de pixel ou subpixel, pequenas variações de resolução, diferenças em relação à escala e rotação no momento de aquisição das mesmas, e ainda, datas de aquisições diferenciadas (SALARI; BAO, 2012).

Os métodos aplicados, conforme evidenciados na Figura 2.2, se baseiam em algumas etapas, que são tratadas como a compensação do movimento de pixel ou subpixel, que

determina a correlação entre as imagens, a interpolação (Figura 2.1) desses pontos correlacionáveis para ajuste da grade da imagem de super-resolução para posterior fusão dos mesmos, e consequentemente, a correção de foco ou diminuição de ruídos que deverá compor a imagem de resolução superior.

Figura 2.1: Projeção dos pontos correlacionáveis das imagens adquiridas sobre a imagem de super-resolução.

Fonte: Adaptado de Salari e Bao (2012).

As técnicas de super-resolução diferenciam-se de técnicas simples de ampliação de imagens, como as de interpolação, principalmente porque estas não têm como objetivo remover distorções da aquisição. As de super-resolução têm uma formulação semelhante à das técnicas de restauração de imagens, porém são diferentes quanto aos objetivos. A restauração de imagem não busca aumentar a resolução, mas, apenas, remover as distorções que degradaram a imagem (THÉVENAZ et. al., 2000). A super-resolução pode ser vista como uma técnica mista de restauração e interpolação, pois proporciona ampliação e remoção de distorções, como detalhadas na Figura 2.2. Para a obtenção da imagem de super-resolução, Baker e Kanade (1999), propuseram que o registro das imagens é o ponto principal no processo, pois sem a correlação entre os pontos comuns das imagens que formam sequência torna-se impossível o reaproveitamento dos pontos para interpolação e posterior fusão das imagens correlacionáveis. E como a fusão ocorre por uma média aritmética, ocorre a minimização dos ruídos na imagem obtida.

Figura 2.2: Etapas para geração da imagem de super-resolução.

Fonte: Adaptado de Baker e Kanade (1999).

As aplicações que se beneficiam com a super-resolução podem ser classificadas quanto ao objetivo desejado. Assim, têm-se as aplicações com objetivo de obtenção de uma única imagem de resolução superior a partir do conjunto de imagens capturadas e as aplicações com objetivo de obtenção de uma sequência de imagens de resolução superior a partir da sequência de imagens adquiridas com sobreposição.

Na primeira situação, em que o objetivo é obter uma única imagem de super-resolução, podem-se citar os exemplos de uso de imagens estáticas a partir de vídeo (YANG et al., 2012); fotografias panorâmicas e mosaicos (LI; GENG, 2011); imagens aéreas e de satélite (MUAD; FOODY, 2010) e de diagnóstico médico (TANG et al., 2008) (MEITAV; RIBAK, 2011). Na segunda possibilidade, o objetivo é obter uma sequência de imagens de super-resolução, tais como: vídeo de alta definição; recuperação de vídeos antigos (JOYEUX et al., 2001); remoção de distorções causadas pela compressão de vídeo (YANG et al., 2012); sequência de imagens médicas para diagnósticos (NASIR et al., 2010), vigilância e sensoriamento remoto (MUAD; FOODY, 2010).

Hsu e Lin (2011) propõem um *framework* universal para melhorar o conjunto de imagens adquiridos por meio de uma imagem de super-resolução. O trabalho é baseado em pesquisas de sistemas de computação com poder de processamento e armazenamento limitados, ou seja, de baixo custo. Para este fim, o método proposto apresenta características importantes com múltiplas escalas e correções de orientações baseadas no algoritmo SIFT (extração dos descritores). Em seguida, usa os recursos extraídos para pesquisar no conjunto de imagens os melhores elementos para a geração da imagem de super-resolução. Uma vez que as correspondências são encontradas, pelos parâmetros de homografia, é possível gerar a imagem de super-resolução. Os resultados demonstram que o método proposto atinge

significativa melhora subjetiva e objetiva quando integrou várias imagens por meio de métodos que não aumentaram significativamente o custo.

Sezer e Altunbasak (2011) descrevem como técnicas de super-resolução podem ser aplicadas para melhorar tanto a resolução espacial como a espectral das imagens multicanal. O artigo é dedicado em formular soluções com imagens de super-resolução em espaço reduzido e de dimensão espectral, principalmente, para elementos hiperspectrais. Primeiramente, um conjunto de imagens é utilizado para definir o processo de aquisição do sensor aproximado em domínio espectral e espacial. Em seguida, um método de separação de origem é introduzido para decompor a informação nos domínios espectrais. A aplicação de técnicas de super-resolução para dados em multicanal mostra que o requisito e a concepção de equipamentos faz com que a captura do espectro de alta-resolução podem ser contornadas por um rápido pós-processamento de métodos simples e eficientes. O algoritmo, não só apresenta tempos de computação otimizados, em relação aos métodos baseados em pixel, mas também proporciona um significativo nível de robustez contra ruídos no momento da aquisição. É possível melhorar (reconstrução da imagem) e o desempenho do método apresentado pelo emprego de uma melhor mistura espectral. Os autores utilizam POCS (*Projection Onto Convex Sets*) e também o algoritmo SIFT para realizar as correspondências entre as imagens.

Este trabalho se enquadra nas aplicações pertencentes à obtenção de uma imagem de super-resolução, mas a ferramenta implementada possibilita experimentos com as imagens de super-resolução geradas, mas não totalmente automatizados.

2.3 Processo de Reamostragem

Reamostragem é tratada como a amostragem de um sinal discreto que originalmente foi amostrado, segundo Telles (2008).

Conforme Lillesand e Kiefer (2000), as técnicas de reamostragem são importantes em diversas operações de processamento digital, tais como: correção geométrica de imagens, registro de imagens de diferentes resoluções espaciais, aumento da resolução espacial de uma imagem, e também, em outras fontes de um Sistema de Informações Geográficas.

A função principal do processo de reamostragem envolve a atribuição de valores a novos pontos inseridos em um campo de valores já existentes, por exemplo, na necessidade do uso de imagens em uma grade maior ou menor. Conforme a aplicação pode-se chamar de reamostragem ou subamostragem da imagem. O resultado da reamostragem será uma imagem com valores reamostrados nas novas posições da grade criada. Essas operações são

necessárias quando se necessita, no domínio do tempo, de uma resolução maior do que aquela correspondente à taxa de amostragem, segundo Yaroslavsky (2002).

Os métodos de reamostragem mais utilizados são o Vizinheiro mais Próximo, Bilinear e Bicúbica. Estes métodos são populares devido à sua simplicidade computacional. Esta simplicidade, porém leva a uma baixa acurácia e a produção de artefatos (*aliasing*) que são sobreposições das bordas do sinal.

Wittman (2005) definiu 9 critérios básicos para um bom método de reamostragem e o último está relacionado ao cálculo computacional do mesmo. Os critérios são:

- Invariante à geometria: deve-se preservar a geometria e os tamanhos relativos dos objetos na imagem;
- Invariante ao contraste: deve preservar os valores de luminância dos objetos na imagem e o contraste da mesma como um todo;
- Método não deve acrescentar ruído ou outros artefatos na imagem;
- Deve-se preservar as bordas e realçá-las quando possível;
- *Aliasing*: o método não deve alterar as formas das bordas;
- *Preservar texturas*;
- *Over-smoothing*: não deve produzir artefatos indesejáveis com efeito de bloco;
- Conformidade com a aplicação: o método deve produzir resultados apropriados ao tipo de imagem;
- Sensibilidade a parâmetros: o método não deve ser muito sensível a parâmetros.

2.4 Técnicas de interpolação

Uma das principais características das técnicas de interpolação é assegurar que o valor reamostrado seja um dos valores originais, ou seja, os novos valores são gerados influenciados por elementos existentes na cena, e também, de apresentar bom desempenho na preservação do contraste.

Algumas transformações quando aplicadas em uma imagem, por exemplo, ampliação (zoom), translação, rotação e ou escala, podem causar efeitos indesejáveis na imagem, podendo assim, na reamostragem da mesma, apresentar baixa resolução radiométrica. Para isso é necessária a interpolação, a qual, dependendo da técnica utilizada, poderá causar um efeito de suavização na imagem. A interpolação também é utilizada quando se necessita

ajustar uma imagem sobre a outra, após a aplicação das funções de mapeamento. Isso acontece quando as grades (imagens) resultantes não são regulares, como mostra a Figura 2.3. A outra possibilidade de se utilizar uma técnica de interpolação é para a criação de uma imagem superamostrada, ou seja, uma imagem ampliada em relação à sua original.

As técnicas de interpolação mais utilizadas são: a técnica do Vizinho Mais Próximo, a Interpolação Bilinear e a Interpolação Bicúbica. Vários outros métodos estão disponíveis (tanto no domínio do espaço, quanto no domínio de frequências, por meio da utilização da Transformada de Fourier), porém, a relação custo computacional/qualidade dos resultados torna estes três métodos os mais usuais (RICHARDS, 2013).

Figura 2.3: Reamostragem: (a) Pixels na grade original; (b) Posição dos pixels transformados sobrepostos na grade original; (c) Ajustamento da grade transformada sobre a grade original.

Fonte: (ATKINSON et al., 1996).

Na reamostragem por vizinho mais próximo o nível de cinza de um ponto transformado é igual ao nível de cinza do pixel mais próximo da imagem digital. No caso da interpolação bilinear, o nível de cinza de um ponto transformado é obtido por interpolação dos quatro pixels mais próximos. Já a interpolação bicúbica utiliza os dezesseis pixels mais próximos de um ponto transformado para a interpolação do nível de cinza.

2.4.1 Interpolação pelo vizinho mais próximo

Por meio deste método, a cor de um pixel, durante a aplicação da função de transformação, é definida como sendo a cor do pixel definido mais próximo. A Figura 2.4 mostra a determinação da cor de um pixel por esse método. Esta interpolação trata-se do método mais simples de ser implementado, e também o mais rápido de ser executado, porém, dependendo da aplicação o resultado obtido poderá ser o menos preciso.

A interpolação ocorre com a verificação da menor distância entre os pixels vizinhos como é ilustrado na Figura 2.4.

Figura 2.4: Interpolação pelo Vizinho Mais Próximo.

2.4.2 Interpolação utilizando a média

Este método consiste em obter a cor do pixel através de um cálculo de média, aplicado entre as linhas e as colunas da imagem, conforme ilustrado pela Figura 2.5 e Equações (2.1) e (2.2). Trata-se de uma interpolação bilinear, mas somente para grades regulares. Como se observa na figura, os elementos da imagem original são representados pelas letras (a , c , g , i), enquanto os valores determinados pelas Equações (2.1) e (2.2), são representados pelas letras (b , d , e , f , h).

a	b	c
d	e	f
g	h	i

Figura 2.5: Representação da sub-imagem.

Assim,

$$b = \frac{a+c}{2}, \quad h = \frac{g+i}{2}, \quad d = \frac{a+g}{2}, \quad f = \frac{c+i}{2} \quad (2.1)$$

$$e = \frac{b+h+d+f}{4} = \frac{a+g+i+c}{4} \quad (2.2)$$

A sua implementação além de simples, apresenta bons resultados, mas só pode ser aplicada para a ampliação de imagens, ou para a retificação onde as malhas são regulares e com pequenos deslocamentos, conseqüentemente, para malhas não regulares, é inadequada para a retificação.

2.4.3 Interpolação bilinear

A interpolação bilinear utiliza-se de três interpolações sobre quatro pontos vizinhos, conforme a Figura 2.6. Sabe-se o nível de cinza dos pontos $g(i,j)$ e $g(i,j+1)$ e deseja-se encontrar o nível de cinza do ponto $g(i,j')$, fazendo-se, assim, a primeira interpolação. Na segunda interpolação usam-se os pontos $g(i+1,j)$ e $g(i+1,j+1)$. A terceira e última interpolação é calculada através dos pontos encontrados na primeira e segunda interpolação.

Segundo Richards (2013), o ponto encontrado na imagem corresponde ao apresentado na grade da Figura 2.6. Duas interpolações lineares são desenvolvidas ao longo das chamadas linhas de varredura ou linhas da imagem (*scan lines*) para encontrar o valor de $g(i,j')$ e $g(i+1,j')$ como é apresentado nessa figura. Estas interpolações são dadas por:

$$\begin{aligned} g(i, j') &= j' g(i, j+1) + (1-j') g(i, j) \\ g(i+1, j') &= j' g(i+1, j+1) + (1-j') g(i+1, j) \end{aligned} \quad (2.3)$$

onde: g é o valor de brilho e (i', j') é a posição para a qual a interpolação deve atribuir um valor de brilho (tom de cinza). A posição é medida com o respectivo (i,j) e assume na grade o espaçamento de uma unidade em ambas as direções. O passo final é interpolar linearmente sobre $g(i,j')$ e $g(i+1,j')$ e o valor final interpolado pode ser dado por:

$$\begin{aligned} g(i', j') &= (1-i') \{ j' g(i, j+1) + (1-j') g(i, j) \} \\ &\quad + i' \{ j' g(i+1, j+1) + (1-j') g(i+1, j) \} \end{aligned} \quad (2.4)$$

Figura 2.6: Interpolação Bilinear.
Fonte: Adaptado de Richards (2013).

2.4.4 Interpolação bicúbica

A interpolação bicúbica é outro método de interpolação, mas que exige um grande processamento computacional, por trabalhar com uma vizinhança de dezesseis pontos e cinco interpolações. A Figura 2.7 ilustra como seria a vizinhança ao usar este tipo de interpolação:

Figura 2.7: Interpolação Bicúbica.
Fonte: Adaptado de Richards (2013).

Esta interpolação utiliza os dezesseis pixels vizinhos. Quatro polinômios são calculados em quatro linhas de quatro pixels vizinhos ao ponto da imagem, como é mostrado na Figura 2.7. O quinto polinômio é então calculado para a interpolação de um valor de brilho para a coordenada de interesse.

O algoritmo a seguir (Equação 2.5) é usado para executar esta interpolação:

$$\begin{aligned}
g(i, j') = j' \{ j' [j' [g(i, j+3) - g(i, j+2) + g(i, j+1) - g(i, j)] \\
+ [g(i, j+2) - g(i, j+3) - 2g(i, j+1) + 2g(i, j)] \\
+ [g(i, j+2) - g(i, j)]] \} + g(i, j+1)
\end{aligned} \tag{2.5}$$

Esta expressão é avaliada por cada uma das quatro linhas descritas na Figura 2.7 para cálculo das quatro interpolações $g(i, j')$, $g(i+1, j')$, $g(i+2, j')$, $g(i+3, j')$. Estas são interpoladas verticalmente de acordo com:

$$\begin{aligned}
g(i', j') = i' \{ i' [i' [g(i+3, j') - g(i+2, j') + g(i+1, j') - g(i, j')] \\
+ [g(i+2, j') - g(i+3, j') - 2g(i+1, j') + 2g(i, j')] \\
+ [g(i+2, j') - g(i, j')]] \} + g(i+1, j')
\end{aligned} \tag{2.6}$$

Esta interpolação pode apresentar como produto, uma imagem com aspecto suavizado.

2.5 Considerações Finais

Neste capítulo foram descritos conceitos sobre super-resolução, reamostragem e métodos de interpolação, com o intuito de evidenciar o processo para geração de uma imagem de super-resolução. Todos os métodos descritos na seção 2.4 foram implementados na ferramenta desenvolvida e o que apresentou melhor resultado foi o da Interpolação Bicúbica. O próximo capítulo está pautado em técnicas utilizadas para obtenção do registro de imagens automatizado para que seja possível a construção do mosaico das imagens por meio da alta correlação existente entre as mesmas.

Capítulo 3

Técnicas utilizadas para o registro de imagens automatizado

3.1 Considerações Iniciais

Neste capítulo e em suas seções e subseções são descritos os algoritmos utilizados, existentes na literatura de Visão Computacional, para a realização do registro de imagens a partir de uma sequência obtida. Inicialmente é dada ênfase à explicação do algoritmo SIFT proposto por Lowe (2004) que é utilizado para a extração dos descritores dos pontos chave das imagens de interesse. Nesta atividade estão envolvidas a comparação dos descritores das imagens e a eliminação de pontos chave de ambas as imagens que não possuem correspondência entre si. No Capítulo 4 e em suas seções e subseções podem ser encontrados trabalhos correlatos às atividades de geração de imagens de melhor resolução envolvendo inclusive abordagens diferentes.

3.2 Introdução

Uma fase inicial para a geração de imagens, empregada neste trabalho, efetua a correspondência entre várias imagens; uma dessas imagens é definida como a imagem de referência a todo processo. Esta fase inicial é denominada de registro de imagens, que, segundo Richards (2013), compreende uma transformação geométrica que relaciona cada ponto de uma imagem com as coordenadas de um sistema de referência qualquer. Esse sistema de referência é, em última instância, o sistema de coordenadas planas de uma certa projeção cartográfica, ou ainda, uma outra imagem que fornecerá os pontos de controle para que as outras que fazem parte da cena, possam ser posicionadas em relação a esta que foi determinada como a de referência.

A correspondência de imagens (*image matching*) é um aspecto fundamental de muitos problemas de Visão Computacional, incluindo reconhecimento de objetos ou de cenas,

obtenção da estrutura 3D de múltiplas imagens, montagem automática de mosaicos, correspondência estéreo e acompanhamento de movimentos (LOWE, 2004). Uma abordagem para se trabalhar com correspondência de imagens é se usar descritores locais para se representar uma imagem. Descritores são vetores de características de uma imagem ou de determinadas regiões de uma imagem e podem ser usados para se comparar regiões em imagens diferentes. Esses descritores são utilizados em aplicações que se destinam a obter correspondência e reconhecimento de imagens.

Existem diversas técnicas para se descrever regiões locais em uma imagem. O mais simples descritor é um vetor com as intensidades dos pixels da imagem. A medida de correlação cruzada pode ser então usada para computar a similaridade entre duas regiões. Porém, a alta dimensionalidade de tal descritor aumenta a complexidade computacional da comparação. Então, esta técnica é principalmente usada para se encontrar correspondências ponto a ponto entre duas imagens. A vizinhança de um ponto também pode ser escalada de modo a reduzir sua dimensão. Outro descritor simples é a distribuição de intensidades de uma região representada por seu histograma.

Trabalhos recentes na literatura têm se concentrado em descritores invariáveis a transformações nas imagens. Algoritmos baseados em métodos que não detectam apenas pontos de interesse ou as características locais de regiões nas imagens, tem se mostrado bastante eficientes na resolução de problemas relacionadas à área de reconhecimento de padrões. Algoritmos como o SIFT (*Scale Invariant Feature Transform*), proposto por Lowe (2004), o SURF (*Speeded Up Robust Features*), proposto por Bay et al. (2008) e suas respectivas variações, como o PCA-SIFT (*Principal Component Analysis Scale Invariant Feature Transform*) proposto por Ke e Sukthankar (2004), U-SURF (BAY et al., 2008), SURF-128 (BAY et al., 2008), são algoritmos que englobam diversas metodologias relacionadas aos processos de identificação de pontos de interesse, geração de descritores e correspondência entre os pontos de interesse.

Alguns dos fatores que permitiram que esses algoritmos ganhassem popularidade devem-se ao fato dos mesmos agregarem características que os tornam invariantes a transformações como rotação, translação e escala, alterações inerentes às condições de iluminação do ambiente onde as imagens são adquiridas e à existência de ruídos gerados no processo de aquisição ou gravação das imagens (MACHADO et al., 2008).

Existe uma variedade de outros métodos disponíveis na literatura que possuem o mesmo objetivo que o SIFT e o SURF. No trabalho de Mikolajczyk e Schmid (2005) foi realizado um levantamento sobre outros detectores de pontos de interesse e descritores. Neste

trabalho foi constatado que o SIFT apresenta um desempenho superior em relação aos demais algoritmos mencionados em termos de invariância. Os autores do SURF, Bay et al. (2008), afirmam que o seu algoritmo é superior ao SIFT em termos de eficiência de tempo de processamento e que o mesmo apresenta um bom desempenho no processo de reconhecimento de objetos, considerando a qualidade dos pontos de características. O artigo de Bauer et al. (2007) também realiza um estudo comparativo entre os dois algoritmos e suas variações, descrevendo os aspectos positivos e negativos inerentes a cada um deles. Segundo Ihara et al. (2009), os descritores SIFT podem ser usados para reconhecimento de sinal de trânsito porque eles são robustos a mudanças na escala e rotação, mas o processamento em tempo real é difícil, pois o custo computacional da extração e correspondência SIFT é alto.

O algoritmo SIFT é um algoritmo que foi apresentado originalmente por Lowe (1999; 2004) para a extração de descritores. Ao contrário de muitos outros algoritmos com esse objetivo, o SIFT não busca segmentar regiões ou objetos para produzir um conjunto de descritores, mas sim, identificar pontos estáveis. Nesse caso em particular, pontos considerados estáveis são pontos na imagem que tendem a ser invariantes à escala e orientação, isto é, pontos que podem ser reconhecidos como similares independentemente do valor absoluto da escala e orientação que os descrevem. Além destas qualidades, o algoritmo SIFT foi escolhido para ser utilizado neste trabalho, devido à sua eficiência demonstrada em diversos trabalhos encontrados na bibliografia, como parte do processo para o reconhecimento de objetos em cenas com diversos tipos de elementos.

A grande variabilidade da aparência de sinal apresentada nas imagens obtidas em ambientes não controlados também justifica a escolha do SIFT. Ambiente não controlado, neste trabalho, refere-se ao local de obtenção das imagens. Nas imagens, os objetos de interesse, podem apresentar diferentes escalas, diferentes inclinações, diferentes tipos de iluminação (sombra, ao Sol), reflexo de luminosidade (reflexão da luz do Sol) e problemas de oclusão (árvores, nuvens, etc.). Este último problema, o da oclusão de objetos, que pode atrapalhar diretamente a detecção, não foi objeto de estudo deste trabalho de doutorado.

As Seções 3.3 e 3.4, e ainda, as Subseções de 3.3.1 a 3.3.4 seguintes descrevem a estrutura e o funcionamento do algoritmo SIFT para obtenção de pontos chave e extração dos seus descritores característicos. A Seção 3.4 e suas subseções descrevem atividades de correspondência dos descritores SIFT. A Seção 3.5 e suas subseções descrevem a extração dos *inliers* por meio do algoritmo RANSAC, que elimina possíveis pontos falsos correspondentes entre as imagens.

3.3 Algoritmo SIFT

SIFT (*Scale Invariant Feature Transform*) é um algoritmo de Visão Computacional desenvolvido e publicado por David G. Lowe em 1999 (LOWE, 1999). Lowe é professor do departamento de Ciência da Computação na *University of British Columbia*, em Vancouver, Canadá.

O algoritmo SIFT é dividido basicamente em duas partes, a do detector e a do descritor. O detector SIFT é baseado nos cálculos de diferença de Gaussianas e o descritor SIFT utiliza histogramas de gradientes orientados para descrever a vizinhança local dos pontos de interesse.

SIFT permite a detecção e extração de descritores que tem muitas propriedades que são desejáveis para correspondência (*matching*) de diferentes imagens de um objeto ou cena. Esses descritores são razoavelmente invariantes a mudanças de iluminação, ruído de imagem, rotação, escala e pequenas mudanças de ponto de vista.

Os descritores fornecidos pelo SIFT são bem localizados em ambos os domínios, o domínio do espaço e o da frequência, reduzindo assim a probabilidade de não haver correspondência dos descritores por oclusão ou ruído. Os descritores são altamente distintos, permitindo que uma determinada característica possa ser corretamente combinada (*matched*) com alta probabilidade em um banco de dados de descritores de diversas imagens, provendo uma base para o reconhecimento de um objeto ou cena. A obtenção de descritores de uma imagem pelo algoritmo SIFT é feita por meio de quatro estágios principais, onde os dois primeiros estágios descrevem a parte do detector e as duas seguintes a formação do descritor. Esses estágios são (LOWE, 2004):

- **Detecção de extremos no espaço escala:** Esse primeiro estágio da computação faz a busca em todas as escalas e locais de uma imagem. Isto é feito utilizando-se a diferença de filtros Gaussianos de modo a se identificar pontos de interesse invariáveis à escala e orientação. Este é o estágio mais custoso computacionalmente do algoritmo;
- **Localização de pontos chave:** Para cada local candidato, em que foi detectado um extremo (máximo ou mínimo), um modelo detalhado é ajustado para se determinar a localização e escala. Pontos chave, ou pontos de interesse, são então selecionados baseando-se em suas medidas de estabilidade;

- **Determinação de orientação:** Uma ou mais orientações são associadas para cada localização de ponto chave, baseadas em direções do gradiente local da imagem. Todas as operações seguintes são realizadas nos dados da imagem relativamente transformados em relação à orientação, escala e localização de cada ponto chave, assim provendo a invariância a estas transformações;
- **Descritor dos pontos chave:** Neste estágio da computação, os gradientes locais da imagem são medidos na escala selecionada, na região ao redor de cada ponto chave. Estas medidas são então transformadas para uma representação que permite observar níveis significantes de distorção de forma e mudança na iluminação. São então criados histogramas de orientação para compor o descritor.

Esses quatro estágios da computação do algoritmo SIFT são detalhados, respectivamente, nas Subseções de 3.3.1 a 3.3.4 na sequência.

3.3.1 Detecção de extremos no espaço escala

Este primeiro estágio do algoritmo SIFT consiste em identificar locais e escalas, que podem ser repetidamente associados ao mesmo objeto, sob diferentes pontos de vista. A detecção de locais que são invariantes às mudanças de escala da imagem pode ser realizada por busca de características estáveis nas diferentes escalas possíveis, utilizando uma função chamada de espaço escala (WITKIN, 1983). No SIFT é usada a função Gaussiana como sendo o núcleo (*kernel*) da função espaço escala, conforme mostram Koenderink (1984) e Lindeberg (1994) – esses autores também relatam que sob uma variedade razoável de suposições a função Gaussiana é o único núcleo possível para a função espaço escala. Resumidamente, são detectados extremos (máximos ou mínimos) em uma pirâmide da imagem convoluída com a função de diferença de filtros Gaussianos. Pontos chave correspondem a estes extremos para diferentes escalas.

O espaço escala de uma imagem é definido como uma função $L(x, y, \sigma)$, que é produzida a partir da convolução de uma Gaussiana de escala variável $G(x, y, \sigma)$, com uma imagem de entrada $I(x, y)$, conforme a Equação (3.1).

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3.1)$$

onde $*$ é a operação de convolução em x e y , e a função Gaussiana é dada por:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (3.2)$$

Neste filtro Gaussiano, o valor de σ representa o parâmetro que define o fator de suavização da imagem I .

Foi proposta em Lowe (1999) uma forma eficiente para detectar pontos chave estáveis usando os extremos do espaço escala com o uso de uma função de Diferença de Gaussianas (DoG – *Difference of Gaussian*). A diferença de Gaussianas, $D(x, y, \sigma)$, é computada pela diferença de duas imagens filtradas em escalas próximas, separadas por uma constante multiplicadora k . A função DoG é definida pela Equação (3.3).

$$\text{DoG} = G(x, y, k\sigma) - G(x, y, \sigma) \quad (3.3)$$

A convolução de uma imagem com a função DoG é definida por:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.4)$$

Esta convolução é a diferença entre imagens suavizadas por um filtro Gaussiano em escalas σ e $k\sigma$. A função DoG consegue detectar variações de intensidade na imagem, elimina detalhes indesejados e realça características fortes. Variando-se σ , é possível encontrar características em diferentes escalas.

Um modo eficiente para a construção da Diferença de Gaussianas (DoG) é feito por meio de uma pirâmide de imagens, mostrado na Figura 3.1. No algoritmo SIFT, inicialmente a imagem de entrada é redimensionada para o dobro de seu tamanho original, e isto é feito para se conseguir detectar mais pontos chave. Mudando-se esta situação, ou seja, utilizando-se a imagem de entrada no seu tamanho original, consegue-se um maior desempenho do algoritmo, porém são detectados menos pontos chave. Para algumas aplicações o ganho de desempenho pode ser mais interessante do que duplicar a imagem para se ter um maior número de pontos chave.

Para cada oitava no espaço escala, a imagem inicial sofre convoluções incrementais com filtros Gaussianos para produzir um conjunto de imagens separadas por um fator de escala constante k no espaço escala, gerando imagens suavizadas L , representadas na coluna da esquerda da Figura 3.1. É denominado de oitava o conjunto de imagens L e DoG. A Figura 3.1 ilustra duas oitavas, estando à esquerda o conjunto de imagens suavizadas pelos filtros Gaussianos e à direita o conjunto de imagens resultantes da Diferença de Gaussianas.

Figura 3.1: Representação do processo para construção das imagens resultantes da diferença de Gaussiana DoG em cada oitava no espaço escala.
 Fonte: Adaptado de Lowe (2004).

A imagem inicial é convoluída progressivamente com funções Gaussianas para produzir as outras $s + 2$ imagens utilizando um fator $k = 2^{1/s}$. São produzidas $s + 3$ imagens na pilha de imagens suavizadas com o filtro Gaussiano para cada oitava, de forma que a detecção de extremos cubra toda a oitava. A imagem é inicialmente filtrada por filtro Gaussiano com escala σ (sem o fator multiplicativo k). A partir de então, são geradas imagens que são progressivamente convoluídas, cada nova imagem é filtrada com escala k vezes a escala utilizada anteriormente. Imagens em escalas adjacentes são subtraídas para produzir imagens de diferença de Gaussiana, como mostrado na direita da Figura 3.1.

Para exemplificar o processo representado na Figura 3.1 de forma simples, deseja-se um intervalo $s = 2$, e como o fator utilizado é $k = 2^{1/s}$, então, tem-se $k = \sqrt{2}$. Neste caso são utilizadas cinco imagens na pilha das imagens filtradas com Gaussianas, pois $s + 3 = 5$. Estas imagens são:

- A primeira imagem filtrada por uma Gaussiana com escala σ ,
- Outras quatro imagens filtradas com escalas multiplicadas pelo fator k , elas são:
 - A segunda imagem usa $k\sigma$ como escala para o filtro Gaussiano;

- A terceira imagem usa a escala 2σ , obtida a partir de $k^2\sigma$, logo, $(\sqrt{2})^2\sigma = 2\sigma$;
- A quarta imagem usa a escala $2k\sigma$, obtida a partir de $k^3\sigma$, logo, $k^2k\sigma = (\sqrt{2})^2k\sigma = 2k\sigma$;
- A quinta imagem usa a escala $2k^2\sigma$, obtida a partir de $k^4\sigma$, logo, $k^2k^2\sigma = 2k^2\sigma$.

Desta maneira, tem-se no exemplo quatro imagens produzidas por DoG, sendo que a primeira imagem (Figura 3.2 (b)) é dada por $D(x, y, \sigma)$ e a última imagem da pilha de DoG é dada por $D(x, y, 2k\sigma)$. Para melhor entendimento, na Figura 3.2 (a), a primeira imagem à esquerda é $I(x, y, \sigma)$ e a última imagem é $L(x, y, 2k^2\sigma)$.

O processo apresentado na Figura 3.2 constrói apenas uma oitava, sendo que cada oitava representa um conjunto de imagens L e um conjunto de imagens D .

Figura 3.2: Exemplo do processo de construção de uma oitava.

Uma vez que uma oitava tiver sido processada, a imagem Gaussiana L que possuir duas vezes o valor inicial da escala σ (2σ corresponde à antepenúltima imagem da pilha superior) é redimensionada para a metade de seu tamanho, e esta será a primeira imagem da próxima oitava, assim como pode ser visto na Figura 3.1. Cada oitava produz o mesmo número de imagens no intervalo, e a computação é reduzida para cada nova oitava com imagens redimensionadas para a metade. A quantidade de oitavas depende da dimensão da imagem original e os redimensionamentos ocorrem até atingir uma dimensão mínima de

aproximadamente 4×4 *pixels*. No algoritmo SIFT, Lowe definiu a Equação (3.5), a partir das dimensões da imagem (*width* e *height*) para calcular a quantidade de oitavas a serem processadas.

$$octvs = \lfloor \log(\min(\text{width}, \text{height})) / \log(2) - 2 \rfloor \quad (3.5)$$

A pirâmide de imagens filtradas pelas Gaussianas mostrando o processo de construção das oitavas é apresentada na Figura 3.3, com três oitavas e cinco imagens L . Esta pirâmide representa uma forma mais intuitiva de entendimento dos exemplos de construção das imagens resultantes da diferença de Gaussianas DoG apresentada na Figura 3.1 e do processo de construção de uma oitava apresentado na Figura 3.2. Cada oitava é constituída por cinco imagens na pilha das imagens filtradas com Gaussianas, com diferentes escalas (σ , $k\sigma$, 2σ , $2k\sigma$, $2k^2\sigma$), ocasionando diferentes níveis de detalhes. Após o processamento de uma oitava, a oitava seguinte utiliza como base a imagem que possui escala 2σ redimensionada para a metade de seu tamanho.

Figura 3.3: Pirâmide de imagens filtradas pelas Gaussianas mostrando o processo de construção das oitavas.

Fonte: Adaptado de Heckel (2008).

Para detectar máximos ou mínimos locais de DoG, cada ponto é comparado com seus oito vizinhos na imagem corrente e nove vizinhos na escala superior e nove vizinhos na escala inferior, e somente é selecionado um ponto se ele for maior ou menor que todos. Os extremos são dados por valores máximo ou mínimo locais para cada $D(x, y, \sigma)$ que corresponda a um intervalo. As escalas superior e inferior são correspondentes às imagens vizinhas em uma mesma oitava para a pilha de imagens DoG, possuem a mesma dimensão. Quando se diz escala superior e inferior, está se fazendo referência à σ .

Os extremos (máximos ou mínimos locais) das imagens produzidas pela diferença Gaussiana são detectados pela comparação de cada pixel (marcado com “X”) e seus 26 vizinhos na região 3x3 da imagem corrente e escalas adjacentes (marcadas com círculos “O”). Esse processo de detecção de extremos apresenta um custo computacional de verificação relativamente baixo, devido ao fato que muitos pontos amostrados serão eliminados a partir das primeiras verificações. As três imagens DoG apresentadas correspondem à diferença entre imagens adjacentes da pirâmide Gaussiana em um pilha. A Figura 3.4 apresenta um exemplo do processo de detecção de máximos ou mínimos das imagens resultantes da diferença Gaussiana.

Figura 3.4: Exemplo do processo de detecção de extremos (máximos ou mínimos) das imagens resultantes da diferença Gaussiana.

Fonte: Adaptado de Lowe (2004).

A partir de experimentos baseados em tarefas de correspondências usando uma coleção de 32 imagens reais variadas, incluindo cenas ao ar livre, faces humanas, fotografias aéreas e imagens industriais, Lowe (2004) busca determinar o número confiável de amostras em escalas próximas por oitava para detectar os pontos chave estáveis usando os extremos do espaço escala.

Lowe conclui que a maior repetibilidade de pontos chave detectados na mesma posição é obtida quando são amostradas três escalas por oitava das imagens produzidas pela DoG. A Figura 3.5 apresenta os resultados dos experimentos de Lowe, e nela pode-se observar o número de escalas com a maior porcentagem de repetibilidade de pontos chave detectados no mesmo local.

Figura 3.5: Resultado dos experimentos para determinar a escolha do número de amostras em escalas vizinhas por oitava para detectar pontos chave.

Fonte: (LOWE, 2004).

No experimento de determinação do valor prévio de σ , Lowe (2004) investiga a determinação ideal desse valor, que é o valor mínimo de escala da função Gaussiana aplicado à primeira imagem de cada oitava, e para cada imagem de cada oitava aplica-se o fator multiplicativo k . A linha do topo do gráfico da Figura 3.6 apresenta a repetibilidade em porcentagem de pontos chave, que são detectados variando-se o valor prévio de σ . Lowe conclui que a repetibilidade de pontos chave detectados aumenta com o aumento de σ , entretanto, o custo computacional também aumenta com esse número, por isso, no algoritmo

SIFT, Lowe usa $\sigma = 1,6$, o qual provê uma boa taxa de repetibilidade com um custo computacional razoável.

Figura 3.6: Resultado do experimento para determinação do valor prévio de σ para cada oitava.

Fonte: (LOWE, 2004).

O próximo estágio da computação do algoritmo SIFT é definir a localização dos pontos chave e descartar os pontos de amostragem que não apresentam estabilidade.

3.3.2 Localização exata dos pontos chave

Todos os pontos detectados como extremos são possíveis candidatos a pontos chave, e o próximo passo é executar um ajuste detalhado aos pixels próximos em relação à escala, localização, e relação de curvaturas principais. Essa informação permite rejeitar pontos que tenham baixo contraste (são sensíveis ao ruído) ou localizados ao longo da borda. A localização dos pontos chave é especialmente importante para as últimas oitavas, porque o espaçamento amostral destas representa grandes distâncias em relação à imagem base. A

implementação inicial desse enfoque (LOWE, 1999) localizava pontos chave em locais e escala de pontos de amostragem centrais. Mais tarde Brown e Lowe (2002) desenvolveram um método que consiste em ajustar uma função quadrática 3D aos pontos de amostragem locais para determinar a localização interpolada do extremo, a fim de prover melhorias substanciais na correspondência e estabilidade (LOWE, 2004).

Para cada ponto analisado é utilizada uma expansão de Taylor da função no espaço escala $D(x, y, \sigma)$, deslocada de modo que a origem desta expansão esteja localizada no ponto de amostragem.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (3.6)$$

onde D e suas derivadas (primeira e segunda) são calculadas no ponto de amostragem analisado e $x = (x, y, \sigma)^T$ é o deslocamento (*offset*) desse ponto. Ou seja, D é o valor da função $D(x, y, \sigma)$ no ponto avaliado, x é o deslocamento em relação a este ponto e $D(x)$ é a aproximação do valor de $D(x, y, \sigma)$ interpolado para um ponto transladado com *offset* x .

Os coeficientes quadráticos são computados aproximando-se as derivadas através das diferenças entre pixels das imagens já filtradas. A localização do extremo, $\hat{x}$, é determinada ao se fazer a derivada de $D(x)$ (Equação 3.6) com relação à x , igualando o resultado a zero:

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (3.7)$$

Como sugerido por Brown e Lowe (2002), a matriz Hessiana e a derivada de D são aproximadas pelas diferenças de pontos amostrais da vizinhança. O resultado é um sistema linear 3x3 que pode ser resolvido com custo mínimo. Se o *offset* $\hat{x}$ for maior que 0.5 em alguma dimensão, significa que o extremo está mais próximo de outro ponto amostral. Nesse caso, o ponto amostrado é mudado e a interpolação é realizada para esse novo ponto. O *offset* $\hat{x}$ final é adicionado à localização do ponto analisado para se estimar a interpolação para a localização do extremo.

O valor da função no extremo, $D(\hat{x})$, é útil para a rejeição de extremos instáveis com baixo contraste, que seriam sensíveis a ruído. Substituindo-se a Equação (3.7) na Equação (3.6) obtém-se:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \quad (3.8)$$

É aconselhado por Lowe (2004) que se rejeitem valores de $|D(\hat{x})|$ inferiores a um determinado limiar (0,03), assumindo-se que os valores dos *pixels* em tons de cinza da imagem estejam entre [0, 1].

Além do procedimento apresentado para se descartar pontos, Lowe (2004) afirma que para a estabilidade, não é suficiente rejeitar pontos chave com baixo contraste. A função DoG tem uma resposta forte nas bordas, mesmo se a localização ao longo da borda não for bem determinada, isto faz com que estes pontos sejam instáveis até para valores pequenos de ruído.

Um pico não bem definido na função DoG terá uma curvatura principal grande pela extremidade das bordas, mas pequena na direção perpendicular, o que não é desejável. As curvaturas principais podem ser computadas por meio de uma matriz Hessiana 2x2 H , para a localização e escala dos pontos chave:

$$H(x, y) = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3.9)$$

onde D_{xy} é a derivada de $D(x, y, \sigma)$ na localização e escala em relação a x e y ; D_{xx} é a segunda derivada em relação a x ; e D_{yy} é a segunda derivada em relação a y .

As derivadas são estimadas levando em consideração as diferenças entre os pontos de amostra vizinhos à localização e escala definida, resultando nos autovalores (*eigenvalues*) de H , que são proporcionais às curvaturas principais de D (LOWE, 2004). As derivadas podem ser aproximadas por (BELO, 2006):

$$D_{xx} = D(x + 1, y, \sigma) - 2D(x, y, \sigma) + D(x - 1, y, \sigma) \quad (3.10)$$

$$D_{yy} = D(x, y + 1, \sigma) - 2D(x, y, \sigma) + D(x, y - 1, \sigma) \quad (3.11)$$

$$D_{xy} = \left(\begin{array}{l} D(x-1, y+1, \sigma) - D(x+1, y+1, \sigma) \\ + D(x+1, y-1, \sigma) - D(x-1, y-1, \sigma) \end{array} \right) / 4 \quad (3.12)$$

São computadas as somas dos autovalores de H e seus produtos do determinante (LOWE, 2004):

$$\begin{aligned} \text{Tr}(H) &= D_{xx} + D_{yy} = \alpha + \beta, \\ \text{Det}(H) &= D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned} \quad (3.13)$$

onde α é o autovalor com a maior magnitude e β o de menor magnitude.

Quando o determinante é negativo, as curvaturas têm sinais diferentes, então o ponto é descartado como não sendo um extremo. Considere r como sendo a razão entre o autovalor de maior magnitude e o de menor magnitude, assim $\alpha = r\beta$. Então,

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (3.14)$$

que só depende da relação dos autovalores, independente dos seus valores individuais. O valor $(r+1)^2/r$ é mínimo quando os dois autovalores forem iguais e aumenta com r . Para verificar se a razão das curvaturas principais está abaixo de um limiar (*threshold*), r , usa-se a Equação (3.15):

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r+1)^2}{r} \quad (3.15)$$

Lowe (2004) propõe o uso de $r = 10$. Assim, eliminam-se pontos chave que não são estáveis, apesar de estarem próximos de extremidades.

3.3.3 Determinação de Orientação dos Descritores

Até o estágio anterior da computação do algoritmo SIFT, foi vista a invariância dos pontos chave no espaço escala. Ao se atribuir uma orientação consistente para cada ponto chave baseada nas propriedades da imagem, pode-se representar os descritores em relação a esta orientação, conseguindo-se assim invariância quanto à rotação (LOWE, 2004).

A escala σ do ponto chave é utilizada para selecionar a imagem suavizada L com o filtro Gaussiano, com escala mais próxima ao ponto chave e oitava referente ao ponto avaliado. Dessa maneira, todas as computações passam a ser executadas com invariância à escala. Para cada imagem $L(x, y, \sigma)$, na mesma escala σ , a magnitude do gradiente $m(x, y)$ e a orientação $\theta(x, y)$ são computadas usando diferenças de pixels, conforme as equações (3.16) e (3.17).

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3.16)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (3.17)$$

Um histograma de orientações é construído com as orientações dos gradientes dos pontos amostrados que se encontram em uma região ao redor do ponto chave. A orientação do histograma tem 36 regiões (*bins*) cobrindo os 360 graus de orientações. Cada amostra na vizinhança do ponto chave, adicionada ao histograma, é ponderada pela sua magnitude do

gradiente e por uma janela Gaussiana circular com um σ que é 1.5 vezes o tamanho da escala σ do ponto chave.

Picos na orientação do histograma correspondem a direções dominantes para os gradientes locais. O maior pico no histograma é detectado e picos que correspondem a valores percentuais acima de 80% do maior pico são usados para se definir a orientação de cada ponto chave. Portanto, para locais com múltiplos picos similares em magnitude, haverá múltiplos pontos chave criados no mesmo local e escala, mas com diferentes orientações. Somente aproximadamente 15% dos pontos são atribuídos com orientações múltiplas, mas estes contribuem significativamente para a estabilidade da correspondência. Para se definir com maior precisão a orientação, uma parábola é ajustada entre os três valores do histograma mais próximos de cada pico, e então é interpolada a posição do pico para uma melhor precisão.

3.3.4 Descritor local

Nas operações realizadas até então, para cada oitava, foram atribuídos localização, escala e orientação para cada ponto chave, a fim de descrever uma região da imagem em um sistema de coordenadas 2D e prover invariância para esses parâmetros. Para isso, as magnitudes e orientações dos gradientes são amostradas ao redor da localização do ponto chave. O próximo passo consiste em computar o descritor que represente as regiões relativas aos pontos chave. Isto é feito para a região de imagem local que é altamente distintiva e possivelmente invariante para variações restantes, como mudança em iluminação ou ponto de vista 3D.

A computação do descritor de cada ponto chave é feita em relação à orientação definida na seção 3.3.3. Primeiramente é selecionada a imagem filtrada L por meio do filtro Gaussiano referente à escala σ e oitava relativas ao ponto chave. Para se obter invariância à orientação, as coordenadas do descritor e das orientações dos gradientes destes pontos são rotacionadas em relação ao ponto chave. Para maior eficiência, os gradientes são computados para todos os níveis da pirâmide das imagens convoluídas com a função de diferença de filtros Gaussianos, e esses são ilustrados com pequenas setas de cada local amostrado no lado esquerdo da Figura 3.7. Uma função Gaussiana é utilizada para dar peso à magnitude do

gradiente de cada ponto na vizinhança do ponto chave, σ é escolhido igual à metade da largura da janela em que será calculado o descritor.

O descritor de pontos chave é mostrado no lado direito da Figura 3.7, e para cada região, é construído um histograma de orientação de oito direções, com comprimento de cada seta correspondendo à magnitude de cada entrada do histograma.

Figura 3.7: Gradientes da imagem e descritor de pontos chave.

Fonte: (LOWE, 2004).

O peso referente à magnitude de cada pixel é atenuado pela função Gaussiana. A função Gaussiana não é aplicada de modo idêntico ao da seção anterior. É utilizado um peso α para interpolar a direção relativa no histograma.

O descritor é representado por um vetor contendo os valores de todas as orientações das entradas do histograma, correspondendo aos pesos das setas mostradas no lado direito da Figura 3.7. Essa figura ilustra uma matriz 2x2 de orientação de histogramas com 8 direções em cada posição, representando um vetor com 32 elementos descritores para cada ponto chave. Lowe (2004) afirma que os melhores resultados foram obtidos quando se usou uma matriz de 4x4 de histogramas, com 8 orientações cada, resultando em um vetor de 128 elementos descritores para cada ponto chave.

O vetor de elementos descritores é modificado para reduzir os efeitos na mudança de iluminação. O primeiro passo é normalizar o vetor, e em seguida fazer uma mudança no contraste da imagem, na qual cada valor de pixel é multiplicado por uma constante. Uma mudança no brilho na qual uma constante é adicionada para cada pixel da imagem não irá afetar os valores dos gradientes, pois eles são computados de diferentes *pixels*. Dessa forma, o descritor é invariante a pequenas mudanças de iluminação. Mudanças não lineares de iluminação podem ocorrer devido à saturação das câmeras ou devido a mudanças de iluminação que afetem superfícies 3D com diferentes orientações. Esses efeitos podem causar

uma grande mudança nas magnitudes relativas para alguns gradientes, mas afetam menos as orientações dos gradientes. Com isso consegue-se reduzir a influência das grandes magnitudes do gradiente com um valor de limiar. Após a normalização, todos os valores acima de um determinado limiar são ajustados para este limiar. Isto é feito para que direções com magnitude muito grande não dominem a representação do descritor. É sugerido usar um valor de limiar 0,2, e então o vetor é normalizado novamente para a unidade de tamanho. De acordo com Lowe (2004) o valor 0,2 foi obtido por meio de testes usando imagens com diferentes iluminações para um mesmo objeto 3D.

Para cada imagem, são construídos diversos descritores, cada um referente a um ponto chave. Quando se aplica o algoritmo SIFT em uma imagem, tem-se como resultado um conjunto de descritores, que são usados para se fazer a correspondência entre duas imagens.

3.4 Correspondência dos descritores SIFT

Um aspecto importante do algoritmo SIFT é a geração de um número grande de descritores que conseguem cobrir densamente uma imagem quanto à escalas e localização. A quantidade de pontos característicos é particularmente importante para o reconhecimento de objetos, onde a capacidade de se detectar pequenos objetos em ambientes desordenados requer ao menos três pontos característicos de cada objeto, que sejam corretamente combinados (*matched*) para uma identificação confiável.

Para a tarefa de comparação e reconhecimento de imagens, os descritores de pontos chave SIFT são primeiramente extraídos de diferentes imagens para então poderem ser comparados. Uma nova imagem é combinada por comparações individuais de cada descritor com os descritores das imagens *templates* previamente armazenadas em um banco de dados, a fim de encontrar descritores candidatos. Essa atividade é realizada usando a distância Euclidiana em todos os vetores de descritores dos pares de imagens, sendo uma a imagem de entrada e outra a imagem *template*.

Os descritores de pontos chave são altamente distintos, o que permite que um único descritor possa encontrar sua combinação com alta probabilidade em um grande banco de descritores. Entretanto, em uma imagem a ser buscada no banco de dados, muitos descritores de fundo certamente não possuem combinação, aumentando as falsas correspondências

(*matches*). A correta combinação pode ser filtrada do conjunto completo de pontos chave correspondentes pela identificação dos pontos chave que combinam com o objeto em sua localização, escala, e orientação na imagem.

Para se encontrar a correspondência (*match*) entre duas imagens, devem ser encontrados pontos em comum entre as duas. Os pontos de interesse (pontos chave) são detectados utilizando o algoritmo SIFT, onde se obtém para cada ponto, um vetor de descritores com 128 dimensões (subseção 3.3.4). Tendo-se descritores de duas imagens, a tarefa de se encontrar a correspondência de uma imagem em outra é resumida pela busca de se encontrar entre os descritores de uma imagem, os melhores candidatos a serem seus equivalentes na outra imagem. Segundo Lowe (2004), a melhor correspondência candidata para cada ponto chave é encontrada pela identificação de seu vizinho mais próximo em um banco de dados de pontos chave de imagens de treinamento ou *templates*. O vizinho mais próximo é definido como o ponto chave com menor distância Euclidiana para o vetor de descritores invariantes.

Para exemplificar a tarefa de encontrar a correspondência de forma simples entre duas imagens I_1 e I_2 , cujos descritores são definidos respectivamente por di_1 e dj_2 , sendo i, j os índices para cada um dos descritores de cada imagem, k o tamanho de cada descritor e m a magnitude de cada valor dos vetores, tem-se a seguinte representação:

$$\begin{aligned} di_1 &= (m_1 i_1, m_1 i_2, m_1 i_3, \dots, m_1 i_k) \\ dj_2 &= (m_2 j_1, m_2 j_2, m_2 j_3, \dots, m_2 j_k) \end{aligned} \quad (3.18)$$

A correspondência é realizada encontrando-se o melhor candidato dj_2 para determinado di_1 , e isto é feito procurando-se o vizinho mais próximo de di_1 entre todos os possíveis candidatos k em dj_2 . O melhor candidato é definido por dj_2 , que possui a menor distância Euclidiana em relação à di_1 , ou seja, deseja-se encontrar j para todo i que minimize a função:

$$|di_1 - dj_2| = \sqrt{(m_1 i_1 - m_2 j_1)^2 + (m_1 i_2 - m_2 j_2)^2 + \dots + (m_1 i_k - m_2 j_k)^2} \quad (3.19)$$

Dessa forma são encontrados todos os pares de descritores correspondentes nas duas imagens I_1 e I_2 . Essa busca exaustiva para encontrar a correspondência dos pontos chave baseada no vizinho mais próximo pode ser demorada, e para tal existem algumas maneiras para se acelerar essa busca.

Uma estrutura de dados usual para encontrar o vizinho mais próximo de descritores de imagens a partir de consultas é a *k-d tree* (BEIS; LOWE, 1997), que é uma forma de árvore de busca binária balanceada.

Uma árvore *k-d tree* (FRIEDMAN et al., 1977) é construída e nela é realizada a busca exaustiva para encontrar a correspondência dos pontos chave da imagem *template* e os pontos chave da imagem a ser trabalhada. O algoritmo denominado *Best-Bin-First* (BBF) (BEIS; LOWE, 1997) é utilizado para a busca, pois retorna o vizinho mais próximo com alta probabilidade. O algoritmo BBF usa uma ordenação de busca modificada para o algoritmo da árvore *k-d tree*, de forma que os descritores são procurados para suas distâncias mais próximas do local de consulta. Um bom resultado pode ser conseguido com baixo custo interrompendo a busca depois que um número específico de vizinhos tenha sido explorado.

3.4.1 Árvore *k-d tree*

Em Berg *et al.* (2008) é descrita a construção de uma árvore *k-d tree* a partir de um conjunto de pontos P bidimensionais (x, y) . Primeiramente o conjunto é dividido na coordenada x , depois na coordenada y , em seguida, novamente a coordenada x , e assim por diante. Na raiz o conjunto P é dividido com uma linha vertical ℓ em dois subconjuntos aproximadamente de tamanhos iguais. A linha de separação é armazenada na raiz. P_{left} , o subconjunto de pontos para a esquerda ou na linha de separação, é armazenado na subárvore esquerda, e P_{right} , o subconjunto à direita, é armazenado na subárvore direita.

No filho esquerdo da raiz, P_{left} é dividido em dois subconjuntos, com uma linha horizontal; os pontos abaixo são armazenados na subárvore esquerda do filho esquerdo, e os pontos acima são armazenados na subárvore direita. O filho esquerdo armazena a linha de divisão. Da mesma forma, o conjunto P_{right} é dividido com uma linha horizontal em dois subconjuntos, que são armazenadas na subárvore esquerda e direita do filho direito.

Nos netos da raiz, ocorre novamente uma divisão com uma linha vertical. Em geral, os nodos cuja profundidade é “par” são divididos com uma linha vertical, e os nodos cuja profundidade é “ímpar” são divididos com uma linha horizontal. A Figura 3.8 ilustra como é feita a divisão assim como a correspondente árvore binária. Uma árvore como essa é chamada de *k-d tree*, onde k representa a dimensão, portanto essa árvore em discussão seria uma *2-d tree*.

Figura 3.8: Exemplo de uma árvore k-d tree. Na esquerda o modo que o plano é subdividido e à direita a correspondente árvore binária.

Fonte: (BERG et al., 2008).

Pode-se construir a árvore *k-d tree* utilizando um procedimento recursivo, e nele são utilizados dois parâmetros: um conjunto de pontos P e um número inteiro (representando a profundidade utilizada nas chamadas recursivas). O parâmetro de profundidade é zero na primeira chamada. A profundidade é importante, pois determina se a divisão ocorre com uma linha vertical ou uma linha horizontal. O processo retorna a raiz da árvore *k-d tree*. A sequência de execução do algoritmo para construir uma árvore *k-d tree* é mostrada no Algoritmo 1 a seguir e foi obtido em Berg *et al.* (2008).

Algoritmo 1 - ConstruirKdTree(P , $depth$)

Entrada: Um conjunto de pontos P e a profundidade atual $depth$.

Saída: A raiz de uma árvore *k-d tree* armazenando P .

```

1: se  $P$  contém somente um ponto
2:   então retorne uma folha armazenando esse ponto
3:   senão se  $depth$  é par
4:     então divide  $P$  em dois subconjuntos com uma linha
       vertical  $\ell$  através da coordenada  $x$  mediana dos pontos
       em  $P$ . Considere  $P_1$  o conjunto de pontos para a folha
       esquerda de  $\ell$ , e  $P_2$  o conjunto de pontos para a folha
       direita de  $\ell$ .
5:     senão divide  $P$  em dois subconjuntos com uma
       linha horizontal  $\ell$  através da coordena  $y$  mediana dos
       pontos em  $P$ . Considere  $P_1$  o conjunto de pontos abaixo de
        $\ell$ , e  $P_2$  o conjunto de pontos acima de  $\ell$ .
6:      $v_{\text{left}} \leftarrow \text{ConstruirKdTree}(P_1, depth + 1)$ 
7:      $v_{\text{right}} \leftarrow \text{ConstruirKdTree}(P_2, depth + 1)$ 
8:     Cria um nodo  $v$  armazenando  $\ell$ , torne  $v_{\text{left}}$  o filho esquerdo
       de  $v$ , e  $v_{\text{right}}$  o filho direito de  $v$ .
9:     retorne  $v$ 

```

Os elementos armazenados na árvore *k-d tree* são vetores de alta dimensão em R^d . No primeiro nível da árvore (raiz), os dados são divididos em duas metades por um hiperplano ortogonal para uma dimensão escolhida, com um valor limiar. Geralmente, essa divisão é feita em média na dimensão com a maior variação no conjunto de dados. Ao comparar o vetor da consulta com o valor de particionamento, é fácil determinar a que metade os dados do vetor de consulta pertence. Cada uma das duas metades dos dados é então dividida recursivamente da mesma forma para criar uma árvore binária completamente balanceada. Na parte inferior da árvore, cada nodo da árvore corresponde a um único ponto do conjunto de dados, embora em algumas implementações, os nodos folha podem conter mais de um ponto. A altura da árvore será $\log_2 N$ onde N é o número de pontos no conjunto de dados (SILPA-ANAN; HARTLEY, 2008).

A árvore *k-d tree* foi introduzida em Bentley (1975) como uma generalização de uma árvore binária de dimensão elevada. Diversas variações na construção de uma árvore, incluindo a seleção aleatória do valor de particionamento, foram sugeridas. Mais tarde, uma árvore *k-d tree* otimizada, com um tempo logarítmico de busca, foi proposto em Friedman (1977). No entanto, essa busca em tempo logarítmico não se aplica a árvores de grande dimensão, onde o tempo de busca pode tornar-se quase linear.

Assim, embora as árvores *k-d trees* sejam eficazes e eficientes em baixas dimensões, a sua eficácia diminui para os dados com altas dimensões. Isto porque com dados de alta dimensão, uma árvore *k-d tree* geralmente leva muito tempo para encontrar a solução ideal. Busca-se encontrar o mínimo absoluto com um desempenho probabilístico dentro de limites razoáveis com o algoritmo *Best-Bin-First* (BBF).

3.4.2 Algoritmo de busca BBF

A busca pelo vizinho mais próximo é um componente importante em aplicações de Visão Computacional. Em uma aplicação típica, um grande número de descritores SIFT extraídos de uma ou muitas imagens são armazenadas em um banco de dados. Uma consulta geralmente envolve encontrar o vetor ou os vetores com descritores de melhor correspondência no banco de dados para um descritor SIFT (SILPA-ANAN; HARTLEY, 2008).

O algoritmo *Best-Bin-First* (BBF), é um algoritmo de pesquisa, que encontra o vizinho mais próximo para uma grande parte das consultas, e encontra um vizinho muito próximo, nos demais casos (BEIS; LOWE, 1997).

As folhas de uma árvore *k-d tree* formam uma partição completa do espaço de dados, com uma propriedade interessante, os *bins* (nodos da árvore) são menores em regiões de alta densidade e maiores em regiões de baixa densidade.

A busca do algoritmo BBF inicia percorrendo a árvore para encontrar o nodo que contém o ponto de consulta. Isso requer somente $d = \lceil \log_2 N \rceil$ (baseado em uma árvore construída de forma balanceada) comparações escalares, e em geral o ponto recuperado do nodo é uma boa aproximação para o vizinho mais próximo.

Toda ramificação da árvore pode ser podada se a região do espaço que ela representa está além do ponto de consulta q para o vizinho mais próximo, contendo uma distância superior a um raio de busca estabelecido. A busca termina quando todas as ramificações tenham sido podadas. Este processo pode ser muito eficiente em espaços com baixas dimensões, mas em altas dimensões há muito mais nodos adjacentes para o nodo central que deve ser examinado, e o desempenho degrada rapidamente. De forma interessante, uma grande porção dessa busca é gasta examinando nodos na qual somente uma pequena fração do seu volume poderia possivelmente prover o vizinho mais próximo.

A Figura 3.9 exemplifica uma busca BBF em uma árvore *k-d tree* com 8 pontos, onde $k = 2$. Na busca, os nodos próximos do ponto de consulta q são examinados. A Figura mostra um círculo centrado em q com um certo raio determinado, e o nodo adjacente mais próximo no espaço (abaixo de q).

Figura 3.9: Exemplo de busca BBF em uma árvore *k-d tree* com 8 pontos.
Fonte: (BEIS; LOWE, 1997).

3.5 Extração dos *inliers*

Para a detecção e reconhecimento de objetos em imagens, tomando como base a correspondência dos pontos chave extraídos pelo algoritmo SIFT (LOWE, 2004), faz-se necessário eliminar os pontos que não correspondem ou não se ajustam com o modelo do objeto desejado. Apenas são preservados aqueles pontos chave que realmente condizem com o objeto desejado. Neste processo o algoritmo RANSAC (*RANdom SAmple Consensus*) (FISCHLER; BOLLES, 1981) é um dos mais comumente encontrado na literatura para a obtenção de um resultado robusto. Além de ser amplamente empregado para o reconhecimento de objetos, o algoritmo RANSAC possibilita encontrar as correspondências características geometricamente consistentes usadas para resolver a homografia (matriz fundamental) entre pares de imagens. Esta atividade é realizada para a construção de panoramas de imagens utilizando descritores invariantes como é o caso dos extraídos das imagens com o algoritmo SIFT (BROWN; LOWE, 2007).

O algoritmo RANSAC proposto por Fischler e Bolles (1981) é um método de estimação robusto projetado para extração dos *inliers*¹ do conjunto total de dados de entrada. Esses dados de entrada seriam os pontos chave detectados pelo algoritmo SIFT (seção 3.3) aplicado às imagens. O RANSAC é geralmente usado em Visão Computacional, como por exemplo, para resolver simultaneamente o problema de correspondência entre pontos de duas imagens e estimar a matriz fundamental M relacionada ao par de imagens.

A principal vantagem no uso do RANSAC é que ele possui a propriedade de ser um estimador robusto. Estimadores de parâmetros clássicos, como mínimos quadrados, se baseiam em todos os dados que tomam como entrada. Desse modo, se grande parte dos dados vier de uma fonte em que haja erros inerentes, ou ruidosos, (isto é, algum tipo de *outlier*²) o modelo final estimado será prejudicado. O RANSAC, porém, foi desenvolvido com o intuito de evitar que *outliers* prejudiquem a computação do modelo final.

Como mencionado por Fischler e Bolles (1981), ao contrário das técnicas convencionais que usam muitos dados para obter uma solução inicial, e então, eliminar os *outliers*, o RANSAC utiliza apenas um conjunto com número mínimo e suficiente de pontos

¹ *inliers*: pontos de dados que se ajustam com um determinado modelo desejado dentro de uma certa tolerância de erro.

² *outliers*: pontos de dados que não se ajustam ao modelo correspondente ao objeto desejado, estão fora de uma certa tolerância de erro.

necessários para uma primeira estimativa, e procede aumentando o conjunto com pontos de dados consistentes. Uma amostra mínima depende do modelo a ser gerado. Geralmente, é o número de dados que será utilizado para inicializar um modelo. Por exemplo, no caso do modelo a ser estimado ser uma reta, uma amostra mínima consiste de apenas dois elementos.

A Figura 3.10 ilustra um exemplo apresentado por Hartley e Zisserman (2003) que pode ser facilmente visualizado. Nele é estimada uma reta ajustada a um conjunto de pontos bidimensionais. Isso pode ser feito estimando uma transformação unidimensional $x' = ax + b$ entre os pontos correspondentes em duas linhas.

O problema que é ilustrado na Figura 3.10 é encontrar a reta que minimiza a soma das distâncias a partir de um conjunto de pontos de dados 2D, sujeito à condição que nenhum dos pontos válidos diverge dessa reta por mais de t unidades. Deseja-se então a reta ajustada aos dados e a classificação de dados em *inliers* (pontos válidos) e *outliers*. O limiar t é fixo de acordo com a medida de ruído (por exemplo, $t = 3\sigma$). Esse exemplo mostra que o ajuste dos pontos é severamente afetado pelos *outliers*. Já com o uso do algoritmo RANSAC (Figura 3.10 (b)), dois pontos são selecionados aleatoriamente, e esses pontos definem uma reta. Essa seleção aleatória é repetida um número de vezes e a reta com maior número de pontos que ficam dentro de um limiar de distância é considerada como ajuste robusto. Os pontos dentro da distância do limiar são os *inliers* (e constitui o conjunto dos consensos).

Na Figura 3.10 (a) e (b) os pontos sólidos representam os *inliers*, e os pontos abertos representam os *outliers*. Na Figura 3.10 (b) as linhas pontilhadas indicam a distância do limiar, a linha passando pelos pontos a e b constitui a linha ajustada, e o ponto c é um *outlier*.

Figura 3.10: Estimação de uma reta ajustada a um conjunto de pontos 2D.

Fonte: (HARTLEY; ZISSERMAN, 2003).

A partir do conjunto mínimo necessário de pontos selecionados aleatoriamente, o RANSAC mede a precisão do modelo a ser estimado para todos os outros pontos, classificando-os em *inliers* ou *outliers*. Para tal, é definido um valor de limiar t , que é a

distância máxima do modelo que um dado pode estar para ser considerado um *inlier*. O conjunto de *inliers* associado a um modelo é o conjunto de consenso (*consensus set*) do modelo. Assim, com várias iterações, o algoritmo é capaz de encontrar um modelo preciso para o problema em questão. Uma vantagem do RANSAC é a sua habilidade de realizar a estimativa de parâmetros de um modelo de forma robusta, ou seja, ele pode estimar parâmetros com um alto grau de acerto mesmo quando um número significativo de *outliers* esteja presente nos dados analisados (FISCHLER; BOLLES, 1981).

O tamanho do conjunto de consenso é o que determina o quão bom um modelo é. Um parâmetro T define o número necessário que deve ser alcançado para um modelo ser considerado satisfatório. Durante a execução do RANSAC, quando for obtido mais de T *inliers*, o algoritmo finaliza, gerando novamente um modelo utilizando todos os T *inliers* obtidos, que será o modelo final encontrado pelo algoritmo (refinamento do modelo). É necessário definir um número máximo N de iterações que o algoritmo irá realizar, para evitar um número infinito de iterações. Se um determinado modelo não alcançar T *inliers*, o algoritmo gera um novo modelo a partir de uma nova seleção de amostra mínima dos dados de entrada. O algoritmo prossegue até conseguir obter um modelo satisfatório, ou alcançar o número máximo de iterações. No caso de a execução do algoritmo finalizar por alcançar o número máximo de iterações, pode-se retornar o conjunto com mais *inliers*, ou retornar que o algoritmo falhou.

Dado um modelo com parâmetros, deseja-se estimá-lo, para tal, os parâmetros podem ser estimados a partir de um número S de itens em um conjunto de dados conhecido. Os parâmetros que podem ser alterados de modo a determinar um bom resultado para o RANSAC são: t , que indica o limiar de distância que uma determinada entrada pode estar do modelo para ser considerado um *inlier*; T , que representa o número de *inliers* a ser alcançado para que um modelo possa ser considerado um bom modelo, sendo que isso depende de cada caso, normalmente determinado empiricamente; e N , que representa o número máximo de iterações que devem ser realizadas.

As etapas de execução são resumidas como segue no Algoritmo 2 e foram obtidas em Fischler e Bolles (1981) e Hartley e Zisserman (2003):

Algoritmo 2 – RANSAC

Objetivo: Ajuste robusto de um modelo para um conjunto S que contém *outliers*.

Passos:

- 1: Selecione aleatoriamente o número mínimo de pontos s requeridos de um conjunto S para determinar os parâmetros do modelo.
 - 2: Determine o conjunto de pontos de dados S_i que estão dentro de um limiar de distância t do modelo (matriz fundamental). O conjunto S_i é o conjunto de consensos da amostra e define os *inliers* de S .
 - 3: Se o tamanho de S_i (número de *inliers*) é maior que T (número necessário a ser alcançado), os parâmetros modelo são re-estimados usando todos os *inliers* identificados (S_i) e termina.
 - 4: Se o tamanho de S_i é menor que T , um novo subconjunto é selecionado e repete os passos anteriores.
 - 5: Após N tentativas o maior conjunto de consensos de S_i é selecionado, e o modelo é re-estimado usando todos os pontos no subconjunto S_i .
-

O número de iterações N é escolhido grande o suficiente para assegurar uma probabilidade p (normalmente estabelecido para 0.99), que pelo menos uma das amostras aleatórias do conjunto de s pontos não inclui um *outlier*. Considere w a representação da probabilidade que algum ponto de dado selecionado é um *inlier* e $\varepsilon = 1 - w$ a probabilidade de observar um *outlier*. Pelo menos N iterações (cada uma de s pontos) são requeridas, onde (HARTLEY; ZISSERMAN, 2003):

$$(1 - w^s)^N = 1 - p \quad (3.20)$$

e assim o número de iterações necessárias é dado por:

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \varepsilon)^s)} \quad (3.21)$$

O RANSAC é capaz de tratar com eficiência um conjunto de dados que contém uma porcentagem significativa de erros. Este paradigma é particularmente aplicável à análise de cena por causa dos detectores de características locais, que frequentemente cometem erros, que podem ser de dois tipos: erros de classificação e erros de medida. Erros de classificação ocorrem quando um detector de característica identifica uma porção de uma imagem incorretamente como uma ocorrência de uma característica. Erros de medida acontecem quando o detector de característica identifica a característica corretamente, mas comete erros de cálculo em um de seus parâmetros (por exemplo, seu local na imagem). Erros de medida geralmente seguem uma distribuição normal, e então a suposição de suavização é aplicável a eles. Porém, erros de classificação têm um efeito significativo maior que erros de medida (FISCHLER; BOLLES, 1981).

3.5.1 Algoritmo RANSAC aplicado ao ajuste na correspondência dos descritores SIFT

O pseudocódigo do algoritmo RANSAC, apresentado na Figura 3.11, calcula um melhor ajuste na correspondência dos descritores SIFT previamente extraídos de duas imagens. Para o problema específico de remoção de *outliers* na correspondência de duas imagens, a matriz fundamental M pode ser calculada resumidamente da seguinte forma:

- Selecionar aleatoriamente um subconjunto de quatro pontos correlacionados, retirados do conjunto total de pontos chave correlacionados entre duas imagens (pontos chave obtidos pelo algoritmo SIFT);
- Para cada matriz M computada, determinar o número de pontos com distância até a linha epipolar, ou residual, menor que um limiar;
- Selecionar a matriz M que apresenta o maior número de pontos com residual inferior ao máximo definido (*consensus*);
- Recalcular a matriz M considerando todos os *inliers*.

```

entrada:
1  features[] //vetor de descritores; somente os descritores com
   correspondência entre duas imagens são usados na computação da
   homografia. Esses elementos representam o conjunto de observações.
2  n //número de elementos no vetor de descritores.
3  m //número mínimo de correspondências necessárias para instanciar o
   modelo a ser computado, foi usado o valor 4.
4  p_badxform //probabilidade desejada que a transformação final
   retornada pelo RANSAC é corrompida pelos outliers (isto é, a
   probabilidade que em nenhuma amostra de todas os inliers são
   extraídos), foi usado o valor 0,001.
5  err_tol //correspondências dentro desta distância de um modelo
   computado são consideradas como inliers, foi usado o valor 3,0.

saída:
6  inliers[] //vetor de ponteiros para o conjunto final dos inliers.
7  n_in //número de elementos no conjunto final dos inliers.
8  M //matriz de transformação computada usando RANSAC ou NULO no
   erro ou se a transformação não pôde ser computada.
9  in_frac := 0.25 //estimativa pessimista de fração de inliers.
10 in_max := 0, k := 0, consensus[] := NULO, consensus_max[] := NULO
11 matched[] e nm := descritores com correspondência entre as duas
   imagens. nm é número de correspondências no vetor matched[].
12 se (nm < m) //número mínimo de correspondências insuficiente
13   fim
14 in_min := número mínimo de inliers calculados com uma função usando o
   número nm de supostas correspondências.
15 p := (1.0 - in_frac^m)^k //probabilidade que m pontos sejam inliers.
16 enquanto (p > p_badxform)
17   sample[] := amostra mínima de dados selecionados aleatoriamente, 4
   amostras.

```

```

18 pts[] e mpts[] := vetor dos pontos a partir de sample[] que
correspondem aos descritores das duas imagens.
19 M := matriz homográfica planar dos mínimos quadrados a partir dos
pontos de correspondência pts[] e mpts[].
20 se (M <> NULO)
21 consensus[] e in := um vetor de consensos a partir do conjunto
dos descritores com correspondência e da matriz homográfica M
(calculada anteriormente), para um modelo dado e função de
erro. in é o número de elementos do vetor consensus[].
22 se (in > in_max)
23 consensus_max[] := consensus[]
24 in_max := in
25 in_frac := in_max / nm
26 senão
27 consensus[] := NULO
28 sample[] := NULO
29 incrementa(k)
30 p := (1.0 - in_frac^m)^k //probabilidade que m pontos sejam
inliers.
31 fim enquanto
//calcula a transformação final baseada no melhor conjunto de
consensos
32 se (in_max >= in_min)
33 pts[] e mpts[] := vetor dos pontos a partir de consensus_max[] que
correspondem aos descritores das duas imagens.
34 M := matriz homográfica planar dos mínimos quadrados a partir dos
pontos de correspondência pts[] e mpts[].
35 consensus[] e in := um vetor de consensos a partir do conjunto dos
descritores com correspondência e da matriz homográfica M
(calculada anteriormente), para um modelo dado e função de erro. in
é o número de elementos do vetor consensus[].
36 pts[] e mpts[] := vetor dos pontos a partir de consensus[] que
correspondem aos descritores das duas imagens.
37 M := matriz homográfica planar dos mínimos quadrados a partir dos
pontos de correspondência pts[] e mpts[].
38 se (inliers[] <> NULO)
39 inliers[] := consensus[]
40 consensus[] := NULO
41 n_in := in
42 senão se (consensus_max <> NULO)
43 se (inliers[] <> NULO)
44 inliers[] := NULO
45 n_in := 0
46 retorna inliers[], n_in, M

```

Figura 3.11: Pseudocódigo do algoritmo RANSAC para calcular um melhor ajuste na correspondência dos descritores SIFT.

Fonte: (SILVA, 2012).

3.5.2 Aplicação para detecção e reconhecimento de objetos

Nesta subseção são apresentados alguns exemplos de detecção e reconhecimento de objetos, cujo ajuste na correspondência dos descritores SIFT foi realizado com o uso do algoritmo RANSAC. Assim como já elucidado anteriormente, o algoritmo RANSAC é um bom algoritmo para ser usado no reconhecimento de objetos, pois minimiza a influência de *outliers*, eliminando-os do conjunto de pontos chave correspondentes entre duas imagens. São preservados apenas os pontos chave que realmente condizem ao objeto buscado.

A Figura 3.12 apresenta um resultado muito bom de detecção e reconhecimento de um objeto (imagem “box.png”) em uma imagem (imagem “box_in_cene.png”), onde 45 pontos chave apresentaram correspondência entre as duas imagens. As imagens utilizadas neste experimento são imagens de teste disponibilizadas pela biblioteca de Visão Computacional OpenCV (OPENCV, 2010). Na esquerda na Figura 3.12 são mostrados os pontos chave que correspondem aos pontos chave do objeto desejado, e na direita, são mostrados os *inliers* extraídos com o RANSAC.

Figura 3.12: Imagem “box.png” detectada na imagem “box_in_cene.png” com 45 pontos com extração dos *inliers* pelo algoritmo RANSAC.

A Figura 3.13 apresenta um segundo resultado bom com extração dos *inliers* pelo algoritmo RANSAC. Neste exemplo de detecção e reconhecimento a imagem de desenho foi detectada com 39 pontos chave.

Figura 3.13: Imagem de desenho detectada com 39 pontos (*inliers*) extraídos pelo algoritmo RANSAC.

Nos exemplos da Figura 3.12 e da Figura 3.13 são exibidas as localizações exatas dos objetos detectados nas imagens de teste. Isto foi possível, pois o algoritmo RANSAC possibilita encontrar as correspondências caracteristicamente consistentes usadas para resolver a homografia (matriz fundamental) entre pares de imagens, atividade muito usada na construção de panoramas de imagens (BROWN; LOWE, 2007).

Diversos experimentos de detecção foram realizados com sequências de imagens usando o algoritmo RANSAC para eliminar os *outliers* e, preservar com isso apenas os pontos chave realmente correspondentes (*inliers*) entre as imagens que pertencem a mesma cena.

A Figura 3.14 e a Figura 3.15 apresentam dois exemplos de experimentos realizados para detecção de elementos em imagens com sobreposição, ou seja, sequência de imagens. Estes exemplos proporcionaram bons resultados com uma quantidade razoável de *inliers* extraídos. A Figura 3.14 apresenta o resultado de correspondência de pontos de interesse entre as duas imagens, ou seja, apresenta o *matching* entre as imagens.

Figura 3.14: *Matching* entre duas imagens da mesma cena.

A Figura 3.15 apresenta o resultado da determinação de elementos de interesse entre as duas imagens usando o algoritmo RANSAC para a extração dos pontos chave realmente correspondentes entre as duas imagens (*inliers*).

Figura 3.15: Determinação de pontos chave correspondentes com extração dos *inliers* pelo algoritmo RANSAC.

No algoritmo RANSAC, quando uma quantidade razoável de pontos chave for classificada como *inliers* relativos ao modelo estimado, tem-se bons resultados no reconhecimento de objetos. Porém, em muitos casos em que a quantidade de pontos chave correspondentes entre as imagens que estão em conformidade com o modelo não for suficiente, tem-se a possibilidade de resultados enganosos ou o que é mais provável, o não reconhecimento do objeto.

Por isso, Silva (2012) ressalta que alguns experimentos podem não apresentar resultados bons, devido ao fato de poucos pontos chave terem sido estimados ao modelo (pontos chave das imagens dos *templates*) ou estarem muito próximos. Com a extração de

inliers não suficientes para o algoritmo RANSAC calcular a homografia (matriz fundamental) entre as duas imagens, as tentativas de reconhecimento podem causar falhas.

3.6 Algoritmo SURF

Proposto por Bay et al. (2008), o algoritmo SURF (*Speeded-Up Robust Features*) é um método que descreve imagens por meio de elementos denominados de pontos de interesse, pontos chave ou descritores encontrados pelo cálculo dos determinantes H das Matrizes Hessianas da imagem. Seja $f(x,y)$ uma função de duas variáveis, o determinante $H(f(x,y))$ é definido como:

$$H(f(x,y)) = \begin{vmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x \partial y} \\ \frac{\partial^2 f(x,y)}{\partial y \partial x} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{vmatrix} \quad (3.22)$$

Os pontos de interesse são definidos onde o determinante é máximo local. Esses conceitos são transportados para o algoritmo ao se considerar a imagem analisada como uma função $I(x,y)$ da intensidade dos pixels e são aplicados pela convolução da imagem com uma máscara apropriada conforme a Figura 3.16.

Figura 3.16: Da esquerda para direita: Derivadas de segunda ordem de uma Gaussiana, discretizadas, e suas respectivas aproximações utilizadas.

Fonte: (BAY et al., 2008).

A convolução entre uma imagem I e uma máscara M pode ser definida como:

$$(I * M)(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} M(m, n) I(x + m, y + n) \quad (3.23)$$

Ou seja, a convolução no pixel (x,y) será dada pela soma dos pixels de sua vizinhança, com os pesos dados a partir da máscara utilizada, sendo o pixel central da máscara o peso do pixel (x,y) da imagem.

Para que se evitem os problemas que poderiam ser causados por ruídos na imagem, não se faz a convolução com máscaras que simplesmente calculem as derivadas de segunda ordem. Antes disso, é necessário suavizar a imagem por meio da convolução com uma Gaussiana. Desse modo, as máscaras aplicadas são aproximações das derivadas de segunda ordem da função de suavização.

Para se analisar imagens em diferentes níveis de resolução, a estratégia usual é construir uma Pirâmide de Imagens a partir da imagem analisada. Uma Pirâmide de Imagens é uma sequência na qual cada imagem subsequente tem um nível de resolução menor do que a anterior. Uma maneira de gerar os novos níveis é construir uma Pirâmide, na qual o nível imediatamente acima é gerado a partir da convolução com uma máscara que reflita uma distribuição Gaussiana. Porém, nota-se que é um processo de grande custo computacional.

A estratégia utilizada para se reduzir o custo foi fazer uso da representação da imagem introduzida por Viola e Jones (2001), chamada Integral de Imagem (Equação 3.24).

$$II(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (3.24)$$

A Figura 3.17 mostra como, depois de calculada a Integral da Imagem, é possível achar a área de qualquer região retangular à partir de três operações. Além disso, mostra que em uma única passada pode-se calcular a Integral da Imagem $II(x,y)$ a partir das seguintes recorrências (Equação 3.25):

$$\begin{aligned} s(x, y) &= s(x, y-1) + I(x, y) \\ II(x, y) &= II(x-1, y) + s(x, y) \end{aligned} \quad (3.25)$$

Figura 3.17: Integral da Imagem.

Fonte: (BAY et al., 2008).

Devido ao cálculo da integral da imagem, não é mais necessário que se passe pelos diferentes níveis subsequentes da Pirâmide de Imagens, pois o cálculo envolve apenas alguns pixels em cada nível. Não sendo necessário passar por níveis cada vez menores na escala da pirâmide, a estratégia do algoritmo para fazer a análise multiescala é ir aumentando o tamanho das máscaras aplicadas conforme Figura 3.16, o que é feito em tempo constante. Desse modo, reduz-se o custo computacional drasticamente, o que leva à eficiência em se localizar os pontos de interesse.

A descrição dos Pontos de Interesse é feita, primeiramente, determinando-se uma orientação em torno do ponto a ser descrito. A orientação é determinada pela região de sua vizinhança que congregue, em torno de um ângulo de 30° , o maior número de pontos descritores vizinhos.

Posteriormente, determina-se uma região quadrada na vizinhança do Ponto de Interesse, subdividida em 16 subregiões, com tamanho proporcional à variância da Gaussiana utilizada. Cada subregião, contribui com 4 descritores (Equação 3.26), Dx , Dy , $|Dx|$ e $|Dy|$, totalizando 64 Descritores:

$$Dx = \sum_{i=1}^{25} dx_i, \quad Dy = \sum_{i=1}^{25} dy_i, \quad |Dx| = \sum_{i=1}^{25} |dx_i| \quad e \quad |Dy| = \sum_{i=1}^{25} |dy_i| \quad (3.26)$$

Onde dx_i e dy_i são cada uma das 25 derivadas em x e y , extraídas de 25 pontos igualmente distribuídos dentro de cada subregião, a partir da convolução com máscaras chamadas de *Haar Wavelets* (Figura 3.18).

Figura 3.18: *Haar Wavelets* d/dx e d/dy .
Fonte: (BAY et al., 2008).

Os Descritores são vetores de 64 dimensões. Para se achar as correspondências, calcula-se a soma do quadrado das diferenças entre cada uma das subregiões correspondentes entre cada vetor da imagem-objeto e da imagem de busca (Equação 3.27).

$$Distância = \sum_{i=1}^{16} (Dx_i^{obj} - Dx_i^{img})^2 + (Dy_i^{obj} - Dy_i^{img})^2 + (|Dx_i^{obj}| - |Dx_i^{img}|)^2 + (|Dy_i^{obj}| - |Dy_i^{img}|)^2 \quad (3.27)$$

Dessa forma, as correspondências são feitas a partir das menores distâncias. A correspondência entre descritores da imagem-objeto e da imagem de busca é analisada de modo a se afirmar se há ou não casamento (*matching*) entre as imagens.

3.7 Considerações Finais

Neste capítulo, foram descritos os algoritmos utilizados para compor a metodologia proposta no presente trabalho, para realizar o registro das imagens, com o objetivo de gerar uma imagem de super-resolução. O algoritmo SIFT proposto por Lowe (2004) foi utilizado para a extração dos descritores dos pontos chave das imagens utilizadas (imagens de entrada). Foi visto que, para a detecção de pontos de correlação, em uma fase inicial, é efetuada a correspondência (*matching*) entre essas imagens, utilizando para isso, o algoritmo *Best-Bin-First* (BBF) (BEIS; LOWE, 1997). O algoritmo SURF proposto por Bay et al. (2006, 2008) também foi descrito e é uma boa alternativa para a detecção dos pontos de correspondência

entre as imagens de referência e as demais que compõem a cena analisada. Esses algoritmos buscam encontrar pontos em comum entre várias imagens, ou seja, buscam encontrar entre o conjunto de descritores de uma imagem, os melhores candidatos a serem seus equivalentes, na outra imagem.

Constata-se também que o algoritmo RANSAC (FISCHLER; BOLLES, 1981), é um dos mais presentes nos trabalhos encontrados na literatura para identificar os pontos chave que se ajustam com o modelo do elemento desejado (identificar os *inliers*), por isso o uso desse algoritmo torna o processo de seleção de pontos de interesse adequado ao trabalho.

Capítulo 4

Trabalhos Relacionados

4.1 Considerações Iniciais

Neste capítulo, resumem-se os resultados das investigações realizadas para obter conhecimento de trabalhos relacionados ao apresentado nesta tese, envolvendo inclusive abordagens diferentes. Esta busca objetivou a aquisição de conhecimento das técnicas, algoritmos e metodologias para a construção de imagens de melhor resolução a partir de sequências de imagens digitais ou imagens individuais, por meio de interpolações e ainda, sequências de vídeo, com processamento *off-line* ou em tempo real. Com base na investigação realizada, são apresentados neste capítulo, trabalhos que utilizam algoritmos que retificam e restauram imagens que estão mais próximos da metodologia apresentada neste trabalho de doutorado.

4.2 Trabalhos Relacionados à construção de imagens de melhor resolução

Devido à crescente demanda por sistemas inteligentes de auxílio a usuários, reconhecimento de elementos em imagens tornou-se um tema de grande interesse. Com a possibilidade da detecção e do reconhecimento de elementos específicos, muitas aplicações para diferentes situações em diversas áreas tornam-se viáveis. Em virtude disso, o desenvolvimento de técnicas que permitem a geração de imagens de alta resolução a partir de uma sequência de imagens, tem sido útil em várias aplicações. Tais aplicações são destinadas, por exemplo, para a identificação de placas de veículos (HATHAWAY; MEYER, 1999), para a identificação facial de pessoas suspeitas (criminosos) em imagens de circuito fechado (SEZER, 2003), na geração de imagens aéreas de alta resolução a partir de imagens de vídeo ou câmera fotográfica (NEMRA; AOUF, 2009) (NASIR; STANKOVIC; MARSHALL,

2011), para a determinação de diagnósticos por meio de imagens médicas (TANG et al., 2008) (KENNEDY et al., 2006), e também, na reconstrução de filmes antigos, onde se procura minimizar a quantidade de ruídos e borramentos que aparecem em cada *frame* projetado (IRANI; ANANDAN; HSU, 1995) (SALARI; BAO, 2012).

A geração de imagens de alta resolução com base em uma sequência necessita, primeiramente, que as imagens pertencentes à cena estejam retificadas, ou seja, o registro de imagens deve ter sido executado, e por isso, é considerado como a primeira etapa nesse processo. A partir disso, pode-se criar uma nova imagem com uma precisão maior e possível de ser utilizada para análises mais precisas (ALMEIDA, 2001). Existem diversas técnicas ou métodos capazes de construir imagens de alta resolução, que são discutidos na sequência.

Tsai e Huang (1984) foram os pioneiros a desenvolver pesquisas sobre o problema da reconstrução de imagem de alta resolução a partir de uma sequência de *frames* de imagens de baixa resolução. O modelo proposto pelos autores baseia-se na translação de movimentos para resolver o problema do registro e da restauração, mas não foram considerados por eles, os efeitos da degradação do sinal e do ruído. O método explora a relação entre transformada rápida de cosseno e transformada direta de Fourier dos frames subamostrados. Já Stark e Oskui (1989) foram os primeiros a propor a reconstrução de imagens de alta resolução usando a formulação da projeção sobre conjuntos convexos (POCS), a partir de um conjunto de imagens. Segundo Stark (1988), o método de projeção sobre conjuntos convexos, também conhecidos como método de projeção convexa, é um algoritmo de restauração de imagens que usa informações prévias já conhecidas das imagens, sendo essas informações os parâmetros de correlação entre as imagens.

Peleg et al. (1987) destacam que, a partir de imagens de baixa resolução, é possível produzir uma imagem com resolução melhorada, usando os deslocamentos conhecidos entre as imagens. Essas imagens que pertencem à sequência a ser analisada foram subamostradas utilizando-se a média ponderada de um bloco (4x4) de 16 pixels. Neste caso, os pixels centrais foram os mais valorizados em relação às extremidades.

Kim et al. (1990) estenderam o método de Tsai e Huang (1984) considerando a aplicação em imagens de baixa resolução que apresentam ruído e borramento, e desenvolveram um algoritmo baseado na teoria de mínimos quadrados ponderados. O

trabalho foi inicialmente motivado, naquela época, pela necessidade de processar imagens do satélite LANDSAT, que fora lançado à órbita para capturar imagens multiespectrais para processamento posterior, a fim de identificar os diferentes fenômenos ambientais, como a distribuição de algum tipo geral de vegetação, estruturas geológicas regionais e extensão da área de água na superfície. Posteriormente o método foi melhorado por Kim e Su (1993) que consideraram o borramento em cada uma das imagens de baixa resolução.

Em Peleg e Irani (1991) são sugeridas melhorias para o método apresentado por Peleg et al. (1987). Os autores tomaram o registro de imagens como base para uma boa determinação de uma imagem de alta resolução. Desenvolveram algoritmos iterativos para se obter um melhoramento na resolução das imagens com precisão de subpixel, tanto em imagens de baixa resolução em níveis de cinza como nas coloridas.

Gross e Ur (1992) desenvolveram uma abordagem para se gerar uma imagem de melhor resolução, através do conhecimento a priori, dos deslocamentos existentes nas imagens de baixa resolução que pertençam à cena. Apesar de abordar um processo que se utiliza das frequências dos sinais gerados pela imagem, esse processo possui a mesma ideia do processo desenvolvido por Peleg e Irani (1991).

Wiman (1992) trabalhou com o aumento da resolução de uma imagem digital (tons de cinza) por meio de super amostragem. Desenvolveu um método para obtenção de uma imagem de alta resolução por meio de imagens de baixa resolução digitalizadas em scanners.

Ballard e Brown (1992) apresentam e descrevem cada etapa do processo para se realizar a restauração de imagens com base em sequências de imagens. Os autores abordam o efeito das transformações de rotação, projetiva e das que utilizam as equações de colinearidade, discutindo as possíveis deformações que as imagens originais deverão sofrer para serem registradas e retificadas, e posteriormente, ser executado o processo de fusão das mesmas.

Hathaway e Meyer (1999), cientistas da *NASA's Marshall Space Flight (Huntsville, AL)*, desenvolveram um software denominado VISAR (*Vídeo Image Stabilization and Registration*), utilizado para melhorar a resolução de imagens. Foi idealizado com o intuito de se realizar análises de imagens científicas, vídeos da área médica, vídeos de circuitos fechados

realizados por câmeras de segurança para identificação de intrusos, imagens que identificam motoristas infratores, aparelhos eletrônicos, além de uma série de outras situações que envolvem sequências de imagens. Inicialmente, o software era de uso restrito aos laboratórios do FBI, sendo posteriormente comercializado. Ele serve de apoio para um sistema que captura a cena (sequência de imagens) e que possibilita a determinação de novos detalhes envolvidos na mesma, levando em consideração a intensidade de luz existente e a velocidade (evitar o efeito de arrasto – borramentos), no momento da aquisição. Ele gera imagens de alta resolução removendo os defeitos ocasionados pelos deslocamentos e efeitos de zoom ocorridos durante a aquisição com base nos frames ou imagens que compõem a cena, resultando na minimização de ruídos existentes nas imagens.

No trabalho de Baker e Kanade (1999) é apresentada uma solução para a geração de imagens de super-resolução, a partir de sequências de imagens não-planares, não-rígidas e sujeitas a auto-oclusões, já que, a grande maioria dos trabalhos desenvolvidos, nessa linha, utiliza imagens planares e rígidas. Os autores buscaram melhorar a resolução das imagens, ou partes das mesmas, que envolviam expressões faciais, pois há grande dificuldade na construção do modelo 3-D da face humana. Com base no algoritmo do fluxo ótico (algoritmo para a determinação dos parâmetros de registro e retificação entre imagens), consegue-se gerar, simultaneamente, os parâmetros para retificação e registro das imagens, e a de super-resolução, conforme pode ser evidenciados na Figura 4.1 e Figura 4.2.

O processo geral para se obter uma imagem de super-resolução pode ser observado na Figura 4.1.

Figura 4.1: Os quatro passos para geração de imagens de super-resolução.

Fonte: (BAKER; KANADE, 1999).

A Figura 4.2, ilustra de forma geral, como este algoritmo converte, simultaneamente, uma sequência de imagens em uma versão da imagem de super-resolução, calculando, o fluxo ótico, que é o responsável por identificar os pontos de correlação que estão dispostos nas imagens pertencentes a sequência de imagens que compõem o mosaico.

Figura 4.2: Processo geral do algoritmo do fluxo ótico na determinação de imagens de super-resolução.

Fonte: (BAKER; KANADE, 1999).

Na delineação de trabalhos a serem realizados, os autores, sugerem a utilização de um algoritmo que executa o registro e a fusão das imagens simultaneamente, denominado de Maximum a Posteriori (MAP) descrito em Hardie et al. (1997). Eles ainda recomendam investigar o relacionamento entre o algoritmo de super-resolução do fluxo ótico e imagens representadas pela “pirâmide” Laplaciana.

Kang et al. (2000) desenvolveram um software que visa gerar imagens de alta resolução a partir de imagens de baixa resolução. Para isso, buscou-se determinar o número mínimo de imagens que deveriam ser adquiridas levando-se em consideração a minimização de custos em relação à aquisição das imagens e equipamentos utilizados, o tempo de processamento, o mínimo de ruídos gerados.

No trabalho de Nguyen e Milanfar (2000) é apresentado um algoritmo para geração de imagens de super-resolução baseado na transformada *wavelet*. O algoritmo é a combinação do processo de interpolação e restauração. Os autores exploraram a estrutura de

entrelaçamento da grade de amostragem em imagens de super-resolução e derivaram um método computacionalmente eficiente de interpolação utilizando a transformada *wavelet* em duas dimensões.

O trabalho de Joyeux et al. (2001) aborda o processo de reconstrução de sequências de imagens degradadas com ênfase em restauração de filmes antigos em preto e branco. O objetivo dos autores é minimizar a quantidade de ruídos e riscos de linhas que aparecem em uma cena com base na sequência de imagens que compõe o filme.

No trabalho de Freeman et al. (2001) são apresentados os algoritmos desenvolvidos por eles para geração de imagens de super-resolução utilizando métodos Bayesianos e redes de Markov. Seus métodos utilizam um conjunto de imagens de treinamento composto por imagens com baixas, médias e altas frequências.

Shrestha e Arai (2003), utilizaram técnicas de super-resolução para melhorar a resolução dos sinais do radar de penetração do solo GPR (*Ground Penetration Radar*). Segundo os autores, não é fácil obter alta resolução como sinais de GPR, pois são muito fracos e envolvidos por ruído. O algoritmo MUSIC (*Multiple Signal Classification*), que é conhecido por sua capacidade de super-resolução, foi implementado para o processamento de sinal e de imagem do GPR. A técnica de estimação espectral convencional, FFT (*Fast Fourier Transform*), também foi implementada para uma precisão maior do nível do sinal recebido. Os autores propuseram um método de processamento, que combina tempo de resposta do algoritmo MUSIC e IFFT (*Inverse Fast Fourier Transform*) convencional para se obter um nível de sinal de super-resolução e de alta precisão. Para comprovação da proposta, os autores realizaram uma simulação detalhada analisando o SNR (*signal-to-noise ratio*), concluindo ser robusta do ponto de vista do ruído. O método proposto poder ser implementado para praticamente detectar tubulações de água enterradas, tubulações de gás, cabos, e até minas terrestres.

Rubert (2004) realizou pesquisas a partir do desenvolvimento de uma aplicação do método proposto por Freeman et al. (2001) sobre as aplicações de imagens de super-resolução em imagens de sensores remotos. No mesmo ano, Capel (2004) efetuou uma vasta pesquisa sobre diferentes métodos probabilísticos, por meio da utilização de imagens forenses, reconhecimento de faces e imagens obtidas pela sonda Mars Pathfinder.

Willett et al. (2004) implementaram um método híbrido que utiliza *wavelets* e métodos baseados na FFT (*Fast Fourier Transform*) para restaurar imagens de telescópios espaciais. Os autores buscam a geração de imagens de alta resolução astronômicas reconstruídas a partir de imagens borradas e ruidosas (ruído Gaussiano ou ruído de Poisson) de baixa resolução.

No trabalho de Molina et al. (2005), os autores implementaram a construção de imagens de super-resolução das bandas multiespectrais e da banca pancromática do sensor ETM+ do satélite Landsat 7 utilizando-se de métodos Bayesianos.

O trabalho de Sezer (2003) utilizou o método de projeção em conjuntos convexos (POCS – *Projection Onto Convex Sets*) e a análise em componentes principais (PCA – *Principal Component Analysis*) no problema do reconhecimento de faces em imagens de vídeo, onde ele utiliza os subespaços da PCA no problema da super-resolução e reconhecimento de faces. O objetivo do autor é aumentar o desempenho do reconhecimento de faces de sistemas existentes que usam imagens estáticas por meio do melhoramento da resolução das imagens, incorporando informações provenientes de múltiplos frames de vídeo. Segundo os autores, o algoritmo de reconstrução baseado em POCS no domínio do subespaço formulado para super-resolução melhora o desempenho de reconhecimento facial em 6,25% no banco de dados de rosto VPA SR Face Database (SEZER, 2003) comparado com super-resolução no domínio do pixel.

Stelzer (2005) também se utilizou de métodos de projeção de conjuntos convexos (POCS) para a restauração de imagens de microscopia por computador, que apresenta borramento. Além do borramento apresentado pelos microscópios utilizados na pesquisa, também apresenta um nível considerável de ruído Poisson. Para atenuar o ruído Poisson, o autor utilizou a Transformada de Anscombe (AT) (ANSCOMBE, 1948) e o filtro de Wiener pontual (ANDREWS; HUNT, 1977). Segundo os autores, os métodos POCS apresentaram uma boa melhoria na qualidade da imagem, embora o método iterativo possuir um custo computacional alto, o que torna viável a utilização de algoritmos elaborados de forma paralela.

Kennedy et al. (2006) desenvolveram um método de super-resolução para melhorar a resolução de imagens médicas obtidas com o uso de scanners clínicos de tomografia por emissão de pósitrons (PET – *positron emission tomography*). A super-resolução de imagens foi obtida pela combinação de quatro conjuntos de dados com deslocamentos espaciais entre aquisições consecutivas aplicando um algoritmo iterativo. Segundo os autores, resolução PET é limitada a parâmetros físicos, tais como dispersão, contagem estatísticas, faixa de pósitrons e movimentação do paciente, geometria da matriz do detector, e o protocolo de aquisição implementado. Eles demonstraram como um processo iterativo de super-resolução pode ser aplicado para melhorar a resolução de PET usando deslocamentos e rotações no plano transaxial, bem como ao longo da direção axial.

Telles Jr e Rosa (2007) apresentam a super-resolução de imagem do satélite China Brazil Earth Resources Satellite – CBERS 2. O método de super-resolução utilizado é baseado no método das projeções em conjuntos convexos (POCS), modificado para utilizar um interpolador *sinc* (YAROSLAVSKY, 2002), ao invés dos interpoladores tradicionalmente utilizados, como: vizinho mais próximo, bilinear e convolução cúbica e deslocamentos subpixel. Como resultado, os autores aumentaram a resolução espacial das imagens providas pelo CBERS-2 que, originalmente é de 20 metros, para imagens de super-resolução com resolução espacial de 10 metros.

Papa et al. (2008) apresentam uma implementação de métodos de projeção de conjuntos convexos (POCS) na restauração de imagens do sensor CCD do satélite CBERS-2. As imagens usadas nos experimentos, segundo os autores, foram obtidas a partir da imagem observada através de um operador pré-determinado, o filtro modificado inverso (*Modified Inverse Filter*). Segundo os autores, eles foram os primeiros na aplicação de técnicas POCS para restauração imagens de sensoriamento remoto e do uso do Índice de Qualidade Universal (WANG; BOVIK, 2002) nessa área de conhecimento.

Salari e Bao (2012) apresentam uma técnica de obtenção de uma imagem de super-resolução, com a melhoria do algoritmo de Papoulis-Gerchberg, que é aplicado no domínio da frequência sobre imagens de baixa resolução. Com isso, elimina as baixas frequências a partir de um ajuste dinâmico durante o processo de reconstrução da imagem de maior resolução. Os

resultados apresentados são computacionalmente mais rápidos e oferece uma melhor qualidade da imagem obtida em comparação ao algoritmo original de Papoulis-Gerchberg.

Apesar de o trabalho utilizar imagens simuladas, os autores sugerem uma variedade de aplicações, incluindo sistemas de vigilância, imagens médicas e imagens de satélite.

4.3 Trabalhos que realizam o registro de imagens usando o algoritmo SIFT para a reconstrução de imagens de melhor resolução

Diferentemente do processo que já foi adotado há alguns anos para realizar o registro de imagens de forma manual (PELEG et al., 1987) (GROSS; UR, 1992) ou de forma controlada (PELEG; IRANI, 1991) (WIMAN, 1992) (ALMEIDA, 2001), nos tempos atuais são adotadas diversas técnicas para a realização desse processo de forma automática. A busca por uma melhor qualidade no processo de registro de imagens, e com isso alcançar melhores resultados na reconstrução de imagens, fez surgir diversos trabalhos nesse sentido. Alguns desses trabalhos utilizam o algoritmo SIFT (LOWE, 2004) como etapa na localização dos pontos chave que são combinados entre as imagens. Alguns trabalhos mais recentes da literatura são descritos nesta seção.

O trabalho de Tang et al. (2008) é destinado ao registro automático em imagens médicas de sequências microscópicas. Nesse trabalho, os autores propuseram um algoritmo que reduz a complexidade e as dimensões dos descritores SIFT (LOWE, 2004). Um algoritmo de correspondência bidirecional foi adotado para eliminar pontos correspondentes repetidos. O algoritmo RANSAC (FISCHLER; BOLLES, 1981) foi aplicado para remover correspondências de pontos chave errados.

Um novo método de extração de descritores em círculos concêntricos é proposto, pelos autores, para reduzir a dimensão dos descritores do SIFT, que originalmente, como proposto por Lowe (2004) é de 128 valores. Para cada ponto chave, oito círculos concêntricos são extraídos, cujos raios são 1, 2, 3, 4, 5, 6, 7, e 8 pixels respectivamente. O gradiente é calculado em pontos de amostragem na imagem em cada círculo concêntrico em torno da localização do ponto chave, como mostrado na Figura 4.3.

Figura 4.3: Redução da dimensão dos descritores do SIFT.
Fonte: (BAKER; KANADE, 1999).

A sub-região formada pelos círculos concêntricos é expressa por dados de 8 dimensões, ao redor da localização de cada ponto chave, que pode ser expressa por dados de 64 dimensões, o que representa o descritor de cada ponto chave. Segundo os autores, comparado com os algoritmos tradicionais de registro, os resultados mostram que o algoritmo SIFT aperfeiçoado tem melhora do processo de registro de imagem tanto em economia de tempo quanto em redução de complexidade.

Em Hossain et al. (2010) os autores identificaram um problema associado com a forma como os histogramas de orientação são construídos no algoritmo Symmetric-SIFT (CHEN; TIAN, 2009). Foi proposto um histograma de orientação modificado que, segundo os autores, é mais eficaz na distinção entre conteúdos visuais de diferentes imagens. As estratégias propostas pelos autores para incrementar os *bins* do histograma de orientação, são:

- 1) Incremento de 1. Neste caso, o histograma vai representar o número de ocorrências de gradientes em diferentes direções;
- 2) Incremento pela média da diferença de quadrados (ASD – *Average of Squared Difference*) das magnitudes.

Como o Symmetric-SIFT é basicamente uma variante do SIFT, segundo os autores, a abordagem proposta é também aplicável ao SIFT (LOWE, 2004) e outras técnicas que seguem o mesmo princípio na construção do histograma de orientação. Nos experimentos, os autores usaram imagens monomodais e multimodais. Ainda segundo os autores, os resultados experimentais mostram que, em média, a precisão do conjunto de correspondências aumenta em 16,94% quando o ASD é utilizado para construir os histogramas de orientação, ao invés da utilização de magnitudes na forma tradicional do algoritmo SIFT. Por outro lado, em média, a precisão aumenta em 5,47%, se simplesmente for contado o número de ocorrências nos *bins* do histograma e descartar qualquer magnitude de ponderação.

No trabalho de Li e Zhou (2011) é apresentado um algoritmo rápido para realizar a correspondência dos descritores SIFT entre duas imagens. Os autores propuseram um método para otimizar a prioridade do algoritmo de busca *k-d tree* pela escolha do número apropriado de nodos folhas examinados (E_{max}) em uma árvore *k-d tree* comum. Segundo os autores, um valor de E_{max} pode ser selecionado para alcançar uma precisão aproximada em um tempo rápido. A busca com prioridade (ARYA, 1995) é um algoritmo eficiente, ele usa uma lista ordenada para armazenar subárvores (ramos da árvore) que estão na ordem de distância crescente entre o elemento de consulta e a subárvore. A subárvore com distância mais próxima é examinada antes. A diferença entre busca padrão e busca com prioridade é apenas o processo de retrocesso. A busca com prioridade somente examina as subárvores armazenadas na lista ordenada.

O trabalho de Nasir et al. (2011) trata o problema de gerar imagens de alta resolução a partir de imagens de baixa resolução em três passos: registro da imagem, fusão da imagem baseada em SVD (*Singular Value Decomposition*) e interpolação. Primeiramente, os autores personalizaram uma técnica para registro de imagem usando o algoritmo SIFT (LOWE, 2004), Belief Propagation (FELZENSZWALB; HUTTENLOCHER, 2004) e RANSAC (FISCHLER; BOLLES, 1981). A matriz de transformação é estimada após as correspondências corretas dos pontos terem sido estabelecidas e a imagem é reamostrada utilizando o modelo de transformação ótimo, e então, a imagem registrada I_r é utilizada na fase de fusão. Convencionalmente, após o registro da imagem, o processo de interpolação é realizado na imagem registrada seguido pelo processo de restauração. Nesse trabalho, os autores realizaram a fusão na imagem para integrar as informações das imagens de baixa resolução antes do passo de interpolação, com o objetivo de manter informações importantes das imagens de baixa resolução, que pode não ocorrer se a interpolação for realizada diretamente na imagem registrada. A fusão baseada em SVD foi utilizada para integrar as características importantes das imagens de baixa resolução. Para a fase de interpolação, os autores utilizaram uma interpolação bicúbica. Segundo os autores, os resultados mantêm as características importantes e melhora significativamente os resultados de super-resolução, devido ao fato do método de fusão proposto, baseado em SVD, ser capaz de reter as informações importante nas imagens registradas de baixa resolução.

No trabalho Zhang et al. (2011) é apresentado um algoritmo rápido e robusto, de forma automática, para registro de imagens. Os autores usaram a Transformada de Fourier em

combinação com o algoritmo SIFT (LOWE, 2004) para reduzir o tempo de extração dos descritores de pontos chave, e com isso obter uma região eficaz de sobreposição. Com isso, são extraídos os pontos chave efetivos na região de sobreposição entre duas imagens. Segundo os autores, a Transformada de Fourier é usada no registro de imagem para calcular o relacionamento de translação, e obter a posição relativa entre as imagens de entrada. Para cada duas imagens com sobreposição, descritores de pontos chave são detectados usando o algoritmo SIFT. Com as coordenadas dos pontos chave e as posições de relação entre duas imagens, são calculados os pontos característicos da região de sobreposição. Posteriormente, a relação geométrica é utilizada para reduzir as falsas correspondências de alguns pontos chave que tem correspondências semelhantes, e assegurar com isso o registro robusto da imagem. Segundo os autores, este método pode detectar os pontos chave efetivos reduzindo a faixa de detecção dos pontos chave, sem intervenção humana, e reduzir o tempo de execução, garantindo a acurácia na correspondência dos pontos chave. Nos seus experimentos, os autores utilizaram imagens de tamanhos 300 x 384 e 512 x 384 pixels, e afirma obter uma melhoria na eficiência de tempo de execução de registro de imagens em 30% a 40%.

4.4 Trabalhos Relacionados à construção de mosaicos

Liu, Chen e Guo (2008) desenvolveram um algoritmo para criar um mosaico com imagens aéreas de campo. O projeto usa um aeromodelo com uma câmera acoplada para filmar por onde passa. Para o vídeo obtido, o SIFT é usado para extrair pares de pontos característicos entre dois *frames* sequenciais. Na sequência, o RANSAC é utilizado para excluir possíveis falsas correspondências e para calcular a matriz homográfica entre todos os *frames* adquiridos. Feito isso, um *frame* vai ser considerado a imagem base e após, faz uso da cascata de matriz homográfica para calcular as matrizes homográficas de todos *frames* relativos a imagem base. Com isso é possível transladar todo *frame* ao plano da imagem, em que o *frame* base se localiza. Quando o aeromodelo mantém uma rota constante, os *frames* não tem rotação e a variação entre *frames* sequenciais é pequena. Quando o aeromodelo retorna não há rotações notáveis entre *frames* sequenciais, contribuindo para uma boa junção das imagens. A proposta se mostra estável e de alta qualidade, uma vez que as imagens de mosaicos são suaves e completas e, com a boa obtenção das imagens feita pelo aeromodelo,

os *frames* podem ser juntados sem qualquer ruptura durante a aquisição, de acordo com os autores.

Goltz e Shiguemori (2008) também utilizam o algoritmo SIFT para determinar correspondências entre as imagens aéreas sequenciais, obtidas por um VANT (veículo aéreo não tripulado) que pertencem a uma mesma cena. A busca é feita por meio de pontos-chave correspondentes, utilizando seus descritores. A distância euclidiana é utilizada em três abordagens diferentes. Os algoritmos de cada abordagem são: DistEuclidConvencional (1), que faz uma busca diretamente, sem tratar seus resultados; DistEuclidRedundante (2), que chama a função de busca duas vezes e, a segunda chamada é feita invertendo-se os parâmetros da função; somente as correspondências que ocorram em ambas são guardadas e DistEuclidEsc (3), que observa a continuidade de escala entre segmentos de reta traçados entre os pontos pertencentes e às correspondências geradas por esta função. Os autores concluem que a utilização do cálculo da distância euclidiana entre vetores que representam descritores de pontos gerados pelo SIFT mostra-se bastante eficiente. A abordagem DistEuclidRedundante foi a mais precisa na avaliação realizada, no entanto, não foi a que apresentou o menor custo computacional.

O trabalho de Zhan-Long e Bao-Long (2008) trata do desenvolvimento de um software para mosaico automático de imagens utilizando o SIFT para extração de características da imagem seguido do algoritmo RANSAC para eliminação dos *outliers* e estimação da matriz homográfica (LOWE, 2004). O método foi aplicado e testado em um grande número de pares de imagens que sofreram alterações de rotação, translação, escala e adição de ruídos. Os autores afirmam que os testes realizados comprovam a robustez do algoritmo proposto para essas alterações e mencionam que os algoritmos convencionais lidam bem com pares de imagens com fator de escala até 4, mas que o algoritmo proposto, comporta-se bem com fator de escala até 10.

Bei e Haizhen (2009) propõem um algoritmo de mosaicos de imagens baseado também na utilização do SIFT, combinando vários métodos de correspondência de características para obter alta precisão na construção dos mosaicos. O método é aplicado sobre imagens que possuem informações similares de estrutura e textura. Após obtenção das características SIFT, um método, denominado *Cosine Similarity* foi adotado para corresponder os pontos característicos. Este método usa valor de correspondência de 0,99

(limiar) e com isso, foi possível verificar que informações similares de textura e estrutura trazem algumas dificuldades para a correspondência das características, gerando resultados instáveis, podendo produzir falsas correspondências. Então foi utilizado o conceito de grau de *Definitude* e distância euclidiana para obter pontos correspondentes com alta precisão. O valor de *Definitude* representa a discriminação entre a intensidade máxima de pontos correspondentes e a próxima intensidade abaixo do máximo. Os autores concluem que o emprego do algoritmo SIFT juntamente, com filtros para minimizar a quantidade de pontos correspondentes, torna o algoritmo mais robusto e eficiente para os resultados.

Cho, Bae e Durand (2009) desenvolveram um trabalho para reconstruir ambientes 3D a partir de imagens e vídeos capturados em determinados pontos fixos de uma cidade para permitir propagação automática de conhecimento urbano entre mapas e imagens digitais. Este trabalho foi patrocinado pelo Departamento da Força Aérea dos EUA. O mosaico de imagens digitais e vídeos são georregistradas por meio da correspondência de algumas de suas características 2D com as suas contrapartes 3D em imagens Ladar aéreas. Uma vez que os planos das imagens estão alinhados com os mapas gerados pelo Ladar, o conhecimento urbano abstraído pode se propagar de um para o outro. Foram coletadas imagens digitais e vídeos em três locais diferentes do campus do Instituto de Tecnologia de Massachusetts (MIT), onde as câmeras foram programadas para operações de rotação e escala, mas não foram programadas para translação. Os testes mostram que o conhecimento abstrato pode ser rapidamente projetado a partir do espaço 3D em mosaicos estáticos e subsequentemente em sequências de vídeos dinâmicos. Inicialmente os autores estudaram técnicas para construir mosaicos 3D a partir de várias imagens 2D de entrada. A técnica para extrair características foi definida com o uso do SIFT e, em seguida, o RANSAC também foi usado para minimizar os falsos pares correspondentes.

Nemra e Aouf (2009) desenvolveram algoritmos eficientes, robustos e automatizados que se utilizam de imagens de mosaicos para a geração de imagens de super-resolução. A pesquisa é baseada no uso do SIFT para realizar o registro das imagens que compõem o mosaico. As imagens foram adquiridas por aeronaves não tripuláveis e os resultados foram validados em comparação às técnicas clássicas e os softwares Photo Stitch da CANON e o REALVIZ Stitcher.

Li e Geng (2010) desenvolveram um software para a construção de mosaicos de imagens utilizando SIFT para a extração de características e a técnica BBF para acelerar a busca de correspondência de pontos característicos. Na sequência, o algoritmo RANSAC foi usado para remover falsas correspondências entre pares de imagens. Imagens capturas por uma câmera de vigilância foram usadas como entrada para testar o software. Os autores concluem que o algoritmo usando o SIFT é capaz de gerar bons mosaicos de imagens sequenciais, mesmo usando câmeras de vigilância, mas que é custoso computacionalmente, calcular os descritores para as várias imagens.

Wang et. al (2010) constroem automaticamente mosaicos de imagens de fundo de olho por meio dos descritores SIFT e pela BBF que determina a correspondência de pontos de controle. A escolha do SIFT foi definida em virtude da capacidade desse algoritmo em reconhecer características de baixo contraste e pela pouca distorção da geometria entre diferentes campos de visão deste tipo de imagem. Os *outliers* são descartados usando o algoritmo MLESAC (*Maximum Likelihood Estimation Sample Consensus*) ao invés do RANSAC. O processo inicia-se por um pré-processamento aplicando um filtro homomórfico no domínio da frequência para comprimir a faixa dinâmica da componente de iluminação e melhora o contraste da componente de reflexão. Na sequência é aplicado o SIFT para gerar os descritores de características e o algoritmo MLESAC para filtrar os falsos pontos correspondentes. O registro de imagens é completado com a correção das imagens distorcidas utilizando uma transformação espacial e com isso as imagens registradas alinham-se com as imagens de referência, podendo assim ser realizado o mosaico. O registro de imagens foi analisado por alguns algoritmos de transformação do espaço da imagem para mosaicos automáticos; estes foram testados e comparados pelos autores. As transformações são as seguintes: afim, perspectiva, polinomial quadrático e polinomial cúbico. Para estas transformações foram analisados o tempo de processamento e a precisão de coincidência de pontos característicos do registro de imagens. O erro de registro de imagem é definido como a média da distância euclidiana entre cada ponto característico na imagem de referência. Estas várias transformações resultam em regiões de sobreposição. Usando a transformação em perspectiva e polinomial quadrática é possível observar que os capilares estão bem alinhados. Embora a transformação polinomial cúbica tenha um erro menor, os capilares não estão localizados corretamente. De acordo com estas considerações, os autores optaram por adotar o método polinomial quadrático. Os autores concluem que o algoritmo proposto melhora os mosaicos de imagens de fundo de olho, utilizando filtragem homomórfica para minimizar o

baixo contraste das características. Quando se obtém pares de pontos correspondentes, o MLESAC é usado para rejeitar *outliers*, e na sequência o melhor modelo de registro de imagem é determinado e, finalmente, é construído um mosaico automático com imagens de fundo de olho.

Xing e Huang (2010) desenvolveram um método também a partir do SIFT para construir mosaicos com imagens sequenciais de alta resolução obtidas por um veículo aéreo não tripulado (UAV - *Unmanned Aerial Vehicle*). No entanto, o propósito dos UAV's não pode ser alcançado com precisão, o que leva à diminuição da precisão e eficiência da correspondência automática. Existem algumas características do SIFT para este tipo de aplicação, tais como a invariância de escala, rotação e luminosidade que são muito úteis para aumentar a precisão da correspondência das imagens UAV. A detecção dos descritores pelo SIFT correspondeu às características entre imagens e a transformação em perspectiva juntamente com o RANSAC, para filtrar possíveis pontos falsos (*outliers*). A principal novidade empregada é o uso do algoritmo L-M para otimizar os parâmetros que levam a construção do mosaico de imagens. A otimização é feita por um método não linear nos oito parâmetros da matriz de transformação, por conta dos erros na etapa de correspondência das características e a instabilidade da projeção em perspectiva, que é linear. Os autores concluem que mesmo com o SIFT despendendo um tempo excessivo para extrair características é conveniente usá-lo, pois obtém pares correspondentes estáveis, necessários para a criação de um mosaico preciso.

Os pesquisadores Nasir, Stankovic e Marsahall (2010) utilizaram o SIFT para realizar vários testes com derivações desse algoritmo. Realizaram comparações dos algoritmos SIFT, SIFT-BP (Belief Propagation), SIFT-RANSAC e SIFT-BP-RANSAC. Neste trabalho, a proposta foi de utilizar SIFT-BP-RANSAC baseado em registro de imagens para obter uma imagem de super-resolução. A técnica foi aplicada em imagens simuladas e reais e os resultados foram considerados muito bons, especialmente quando comparado com o SIFT tradicional. A vantagem da proposta reside na sua capacidade de eliminar os *outliers* introduzidos no método SIFT-BP e, portanto, obtém corretamente a matriz de transformação. As imagens de super-resolução resultantes apresentaram melhor qualidade visual em comparação com o caso SIFT, SIFT-BP ou SIFT-RANSAC.

Suzuki, Amano e Hashizume (2010) mostram como gerar mosaicos de imagens utilizando uma aeronave não tripulada (UAV) para realizar sensoriamento remoto efetivo. Limites de carga é a fonte da maioria dos erros neste tipo de UAV para a geração de mosaicos precisos. Para superar essas desvantagens é integrado um GPS e sensores inerciais, além da própria câmera digital para a geração de um mosaico grande. A posição e altitude do UAV são estimados usando os sensores e correlação de imagem a partir das características SIFT entre imagens sobrepostas. Um algoritmo de otimização não linear é utilizado para otimizar a posição e atitude do UAV. A área da imagem projetada na DEM (Modelo de Elevação Digital) é calculada usando os dados de posição e atitude. A taxa de sobreposição é calculada usando a área da imagem projetada e avaliação de imagens sobrepostas combinadas. Os pontos chave nas imagens são identificados pela Transformada SIFT e correspondidos para formar imagens sobrepostas. Dados de posição e atitude são estimados utilizando correlação de imagens para minimizar o somatório do erro quadrado do custo da função. O mosaico é gerado usando estimação de posição e atitude baseada na imagem projetada na DEM.

O trabalho de Hiraga e Silva (2011) propõe uma metodologia para construção de imagens panorâmicas utilizando-se do SIFT, algoritmo BBF e algoritmo RANSAC, mas ainda havendo uma preocupação de minimizar posições de distorções nas junções, utilizando para isso retificação cilíndrica. E em Silva et. al (2012) é apresentado um novo algoritmo para a construção de imagens panorâmicas (incluindo panoramas em 360°), que tem como principal característica evitar a distorção que ocorre pela junção de várias imagens sucessivas. Foram utilizados os algoritmos SIFT e RANSAC para encontrar áreas de sobreposição entre os pares de imagens, bem como um algoritmo de Blend para suavizar as junções. O algoritmo proposto não causa distorções durante a junção das imagens, assim, uma correção posterior não é necessária, o que contribui para um melhor desempenho. Os resultados dos experimentos usando softwares comerciais e também o algoritmo proposto foram comparados através de uma análise quantitativa, utilizando medidas numéricas calculadas sobre uma imagem panorâmica, gerada a partir de uma sequência de vinte imagens de uma região mapeada e georreferenciada pelo *Google Earth* e com isso ficou comprovado que a nova metodologia desenvolvida apresentou menores distorções do que os outros softwares comerciais testados, comparados com o *Google Earth*.

Tarallo (2012) propõe a montagem de um mosaico automático de imagens digitais aéreas agrícolas por meio do algoritmo SIFT com aplicação de processamento paralelo em

virtude do elevado tempo de processamento para a geração dos mesmos. As imagens foram obtidas a partir de uma câmera de alta resolução acoplada a um suporte externo de um avião. As imagens contêm áreas de pastagens, matas e área urbana. Para que o mosaico possa ser construído, necessita-se de uma área de sobreposição entre as imagens sequenciais e devido a isso, algumas normas foram estabelecidas para a aquisição das imagens. O uso do algoritmo SIFT, que consegue fornecer os pontos de interesse nas imagens, com uma descrição satisfatória, e que favorece a correspondência entre os pontos de interesse de duas imagens, é um dos diferenciais dessa metodologia, contribuindo decisivamente para a qualidade dos resultados. Esses resultados também apresentam contribuição evidenciando a qualidade dos mosaicos construídos, mas o tempo despendido para a geração desses mosaicos são elevados para a base de imagens sem redimensionamento mesmo com a implementação paralela (*OpenMP*) em *threads*.

Ali (2012) compara a Transformada SIFT com SURF para área de construção de imagens panorâmicas. As imagens deste artigo variam em alguns fatores: iluminação, ponto de vista, rotação, escala, borramento e ruído e estes fatores afetam na detecção das características e na fusão das imagens. Os autores afirmam que o algoritmo SIFT é indicado para aplicações que exigem alta precisão e fidelidade durante a fusão das imagens que apresentam sobreposição. O SURF é colocado como um algoritmo que detecta mais pontos que o SIFT, mas a precisão é menor.

4.5 Considerações Finais

Dentre as diferentes abordagens descritas nesta seção, pode-se verificar que o problema comum encontrado para a geração de imagens de alta resolução por meio de sequências de imagens, está na determinação dos parâmetros para o registro e retificação. Os trabalhos investigados mostram que, o SIFT é considerado um forte descritor de características locais em termos de sensibilidade e precisão segundo a avaliação provida por Mikolajczyk e Schmid (2005). O SIFT é considerado por Farag e Abdel-Hakim (2004) um bom algoritmo de descrição de pontos chave, apresentando vantagens para o processo de correspondência. O SURF também se apresenta adequado para utilização, mas não tem a

mesma produtividade do que o SIFT em relação à extração dos descritores de interesse que fazem parte das imagens analisadas.

A revisão bibliográfica evidenciou a possibilidade de se utilizar o algoritmo SIFT e/ou o SURF para a detecção de pontos de interesse no processo de geração de imagem de melhor resolução (TANG et al., 2008) (HOSSAIN et al., 2010) (LI; ZHOU, 2011) (NASIR et al., 2011) (ZHANG et al., 2011) (BAY et al., 2006), pelo fato dos descritores SIFT e SURF serem robustos. Vários autores citados optaram pela aquisição controlada dos dados, e a partir disso, ter os parâmetros conhecidos a priori (PELEG; IRANI, 1991) (GROSS; UR, 1992) (JOYEUX et al., 2001) (ALMEIDA; TOMMASELLI, 2003).

Outra possibilidade observada é a abordagem utilizando-se do domínio da frequência e com a utilização de filtros para a geração de uma nova imagem de melhor qualidade (TSAI; HUANG, 1984) (SHRESTHA; ARAI, 2003) (WILLETT et al., 2004) (WANG et al., 2010) (SALARI; BAO, 2012).

Após o registro de imagens, é necessária a realização da fusão das imagens para a geração do mosaico a ser analisado. Para o entendimento do processo, são evidenciados vários artigos que relatam as etapas e os algoritmos utilizados para a criação do mosaico. É verificado que a maioria dos autores utilizou o SIFT e o RANSAC para a determinação correta dos parâmetros de correlação entre as imagens.

Capítulo 5

Metodologia de geração de imagens de super-resolução

5.1 Considerações Iniciais

Neste capítulo são descritos a metodologia e os materiais utilizados no presente trabalho de doutorado, para se realizar automaticamente a geração de imagens de super-resolução a partir de sequências de imagens que formam um mosaico ou que possibilitam a formação de imagens panorâmicas.

5.2 Descrição das imagens utilizadas

No presente trabalho foram utilizados os seguintes conjuntos de imagens:

- Conjunto A: imagens coloridas obtidas com câmeras digitais de baixa resolução para validação da ferramenta;
- Conjunto B: imagens coloridas com elementos textuais adquiridas com câmeras de alta resolução para aplicação em ferramentas de reconhecimento de caracteres;
- Conjunto C: imagens aéreas coloridas adquiridas com câmeras de alta resolução para composição dos mosaicos;
- Conjunto D: imagens aéreas em níveis de cinza para composição dos mosaicos com essas características.

A seguir será detalhado cada conjunto de imagens.

5.2.1 Conjunto A

O conjunto A é composto por imagens adquiridas com câmeras de baixa resolução, obtidas do trabalho de Almeida e Tommaselli (2003) e são utilizadas no presente trabalho para validar o algoritmo desenvolvido. Optou-se por essas imagens, pois foram adquiridas de maneira controlada em relação à translação, rotação e escala mínimas, tendo sido utilizados os seguintes equipamentos (ALMEIDA; TOMMASELLI, 2003):

- Conjunto A1: obtido com a câmera KODAK DC-40, que possui lentes com distância focal fixa de 46,981 mm, diâmetro das lentes de 7,7 mm e distância hiperfocal de 1,2 m. A câmera utiliza o processo digital para capturar, gravar e processar as imagens. O sistema de lentes possui 5 elementos ópticos, que transmitem a luz para o CCD. As dimensões do CCD são de 768 x 512 pixels (6,9 mm x 4,6 mm). Um buffer armazena temporariamente as imagens digitais, que posteriormente são transferidas para o computador, através de um cabo interligando o dispositivo de saída da câmera à porta serial do computador;
- Conjunto A2: obtido com a câmera KODAK DC-210, que possui lentes com distância focal variável de 29 ou 58 mm. Também utiliza o processo digital para capturar, gravar e processar as imagens, sendo que estas permanecem armazenadas em um PC Card (Picture Card), cartão de imagens, que é equivalente ao cartão PCMCIA. Possui uma resolução 1152 x 864 pixels, possuindo também, um monitor LCD para visualização das imagens antes de serem transferidas para o computador, o que é feito através de um cabo interligando a câmera à porta serial do computador.

As imagens foram coletadas pelas câmeras digitais descritas, introduzindo pequenos deslocamentos no mesmo plano das imagens para minimizar mudanças de escala, e assim, obter sequências de imagens.

Com a câmera Kodak DC-40, as imagens do Conjunto A1 foram obtidas a uma distância de 2m, obtendo-se uma escala equivalente a 1:43, com tamanho do pixel na cena de 6mm. Enquanto que o Conjunto A2, adquirido com a câmera Kodak DC-210, foi obtido a uma distância de aproximadamente 1m, e no segundo experimento, a uma distância de 2,13m, obtendo-se o pixel de tamanho igual a 2,2mm (no espaço objeto) e escala equivalente a 1:73.

Esta coleta pode ser realizada também, de forma contínua por meio de câmeras de vídeo, mas seria necessária, a utilização de software específico para a separação dos *frames* de interesse.

5.2.2 Conjunto B

As imagens contendo elementos textuais para testes foram adquiridas com a câmera Nikon modelo P500 de 12 megapixels. As características específicas da câmera são:

- velocidade do obturador: ISO-280;
- escala focal: f/3,4;
- distância focal: 4 mm;
- abertura máxima: 3,5 mm;
- tempo de exposição: 1/30 s.

5.2.3 Conjunto C

As imagens aéreas utilizadas no presente trabalho foram fornecidas pela EMBRAPA (Empresa Brasileira de Pesquisas Agropecuárias) Instrumentação, situada em São Carlos – SP. As imagens contêm áreas de pastagens, lavouras e áreas urbanas, adquiridas com o auxílio de um avião tripulado de pequeno porte, contendo uma câmera acoplada a um suporte externo. As imagens possuem 24 bits por pixel, com 72 dpi (pontos por polegada) no formato JPG e foram obtidas em 8 megapixels. A câmera utilizada é do modelo CANON EOS 30D com as seguintes características:

- possui 12 megapixels de resolução;
- velocidade do obturador e tempo de exposição: 1/5000 s;
- abertura da lente e abertura do diafragma: f/3,5;
- comprimento focal: 18 mm.

As imagens foram obtidas seguindo um padrão de aquisição, que inclui os seguintes itens:

- a câmera foi acoplada ao suporte externo do avião;
- o avião percorreu sempre a mesma distância, em trajetórias horizontais em uma área demarcada;

- após obter uma imagem, a câmera foi configurada para aguardar um tempo fixo (3s) antes de obter a próxima.

5.2.4 Conjunto D

As imagens aéreas utilizadas no presente trabalho foram fornecidas pela Faculdade de Ciências e Tecnologia (FCT) – UNESP Campus de Presidente Prudente – Departamento de Engenharia Cartográfica. As imagens foram adquiridas por câmera aérea acoplada a uma aeronave e são imagens em tons de cinza. Não há especificação sobre as características da câmera utilizada. As imagens possuem 8 bits por pixel, no formato BMP e de dimensão 500 x 500 pixels.

5.3 Metodologia geral aplicada ao trabalho

Uma visão geral da metodologia utilizada para o desenvolvimento do trabalho pode ser vista na Figura 5.1 que mostra as etapas para a geração das imagens de super-resolução na forma de diagrama em blocos.

Inicialmente, é adquirida uma sequência de imagens por uma câmera, em seguida é realizado o registro de imagens, com o objetivo de se determinar os parâmetros de correlação existentes entre as imagens utilizando os algoritmos SIFT ou SURF seguido pelo RANSAC, que elimina as falsas correspondências e estima a matriz homográfica. A etapa seguinte consiste na construção das imagens de super-resolução a partir de uma imagem de referência definida pelo operador.

Figura 5.1: Visão geral das etapas de geração de uma imagem de super-resolução.

Como já mencionado, para se obter o registro das imagens, pode-se utilizar o algoritmo SIFT ou SURF durante a determinação dos pontos de correspondência entre as imagens que pertencem à mesma cena. Para o conjunto A essas coordenadas (em pixels) são transferidas para o referencial fotogramétrico (em milímetros) por meio de cálculos realizados com as equações descritas em 5.1.

$$\begin{aligned} x_{fot} &= (x_i - NCol / 2) * TamPixel \\ y_{fot} &= -(y_i - NLin / 2) * TamPixel \end{aligned} \quad (5.1)$$

Onde:

x_{fot}, y_{fot} : coordenada no referencial fotogramétrico;

x_i, y_i : coordenada imagem (coluna, linha);

$NCol$: número de colunas da imagem adquirida;

$Nlin$: número de linhas da imagem adquirida;

$TamPixel$: tamanho do pixel em milímetros na imagem a ser gerada.

Após os parâmetros terem sido determinados, faz-se simultaneamente, a retificação e reamostragem das imagens que pertencem à sequência em relação à imagem de referência. A Figura 5.2 mostra, de forma geral, o processo de retificação implementado.

Figura 5.2: Processo de retificação das imagens.

Com base nos parâmetros que possibilitaram o registro das imagens (detectados e extraídos pelo SIFT ou pelo SURF), determina-se a correspondência dos pixels nas imagens de baixa resolução. Como as coordenadas projetadas na imagem de baixa resolução estão em frações do pixel, há a necessidade de uma interpolação entre os vizinhos para a determinação do nível de brilho relativo ao que se deseja determinar na imagem de super-resolução. O método de Interpolação Bilinear e a Bicúbica foram os mais utilizados por apresentarem melhores resultados, levando-se em consideração a influência de cada pixel em relação ao que deveria ser gerado.

Para se realizar a fusão e consequentemente a geração da imagem de super-resolução pode-se selecionar um dos seguintes procedimentos:

- Média entre os valores de brilho dos pixels: é realizada a partir dos pixels correspondentes determinados pelos processos de registro e retificação, para a determinação do nível de brilho do pixel na imagem resultado, ou seja, é realizada a média aritmética dos valores de brilho, que leva em consideração, a

somatória dos níveis de brilho de cada ponto e a quantidade de imagens que estão sendo analisadas, conforme mostra a Figura 5.3. Isso acarretou maior suavização na imagem obtida em relação aos outros algoritmos implementados, tais como, uso da mediana e minimização dos erros.

Figura 5.3: Fusão entre os pixels das imagens que pertencem à cena.
Fonte: (ALMEIDA; TOMMASELLI, 2003)

- Processo com minimização de erros: inicialmente realiza-se uma primeira estimação da imagem de alta resolução, por exemplo, a imagem de referência superamostrada pela interpolação bilinear ou bicúbica, conforme processo ilustrado na Figura 5.4. A partir da imagem de alta resolução aproximada, projeta-se cada pixel nas imagens de baixa resolução, dessa forma, o valor com menor diferença é utilizado;
- Mediana entre os valores de brilho dos pixels: é definido o elemento mediano entre os elementos que estão vinculados a posição da imagem, isso faz com que exista menos suavização na imagem a ser obtida;
- Média entre três valores medianos: após se determinar o elemento mediano, faz-se a média aritmética entre ele, o seu antecessor e o seu sucessor. Os resultados obtidos foram similares à utilização do elemento mediano apenas.

O método utilizado durante a fusão das imagens que apresentou os melhores resultados foi o vinculado a Mediana entre os três elementos centrais definidos. Assim

diminui-se o efeito de suavização, conforme evidenciados nos testes realizados com uso dos demais métodos.

Figura 5.4: Exemplo de processo para geração de imagens superamostradas através de Interpolação Bilinear.

Fonte: (ALMEIDA; TOMMASELLI, 2003)

Para o desenvolvimento do trabalho foi utilizado o ambiente de programação Microsoft Visual Studio 2010, a linguagem de programação C++ e a biblioteca OpenCV 2.4.2.

A biblioteca OpenCV 2.4.2 disponibiliza os algoritmos detectores de descritores (*features*), que são: SIFT, SURF, BREAK e ORB. E ainda os algoritmos de extração de descritores (*features*): SIFT, FAST, STAR, SURF, ORB e MUSER. Todos esses algoritmos foram incluídos na ferramenta como opção para o operador. No entanto, os melhores resultados foram utilizados com os algoritmos SIFT – *Scale Invariant Feature Transform* (LOWE, 2004) e SURF – *Speeded Up Robust Features* (BAY et al., 2008), e devido a isso todos os resultados se baseiam nesses dois algoritmos.

Como já mencionado no Capítulo 3, o algoritmo SIFT é praticamente invariante à mudança de iluminação, ruído, rotação, escala e pequenas mudanças no ponto de vista. Além disso, ele detecta e extrai a maior quantidade de pontos de correlação entre as imagens do que o algoritmo SURF, apresentando por essa razão maior custo computacional.

Nas seções seguintes são descritas as etapas para a geração de imagens de super-resolução a partir dos conjuntos A, B C e D.

5.4 Etapas para a geração das imagens de super-resolução a partir de dados simulados

Para validar a metodologia da Seção 5.3 foi utilizado o conjunto A alterando a posição da câmera para se obter as imagens adquiridas com pequenos valores de translações, rotações e escalas. Devido a isso, os parâmetros para o registro e retificação das imagens são conhecidos antecipadamente para posterior fusão dos pontos de correlação entre as imagens, gerando-se assim a imagem de super-resolução por meio de elementos reais da cena e não simplesmente calculados por métodos de interpolação.

É importante ressaltar que no processo de geração das imagens de super-resolução com o conjunto A é escolhida uma imagem como referência e as informações necessárias para compor a matriz de super-resolução são buscadas nas demais imagens do conjunto A, conforme ilustra a Figura 5.3.

Com o mesmo conjunto A, no qual já se aplicou as operações de translação, escala e rotação, foram inseridos ruídos aleatórios em todas as imagens da sequência, com o intuito de gerar uma imagem degradada, sendo usado para isso o software gratuito de processamento de imagens GIMP 2.8.0, com 20% dos pontos em relação ao total nas imagens de cada componente de cor (R,G,B). Em seguida, as imagens são subamostradas, processadas para obtenção dos parâmetros para realização do registro de imagens, usando como referência uma das imagens. Finalmente, é gerada a imagem de super-resolução que é comparada à imagem original, tendo ambas, as mesmas resoluções.

Para avaliar a imagem de super-resolução gerada nos dois casos (com e sem ruídos) é feita a sua comparação com a imagem original e são utilizadas duas métricas de avaliação: a Raiz do Erro Médio Quadrático e o Índice de Similaridade Estrutural.

O índice de similaridade possibilita uma avaliação qualitativa, enquanto que o erro médio quadrático possibilita uma avaliação quantitativa. A definição de cada métrica é apresentada a seguir:

- Raiz do Erro Médio Quadrático: a raiz do erro médio quadrático (RMSE, do inglês *Root Mean Square Error*) corresponde à raiz quadrada da soma do quadrado das diferenças de cada ponto das imagens de baixa resolução (LR) e a imagem de mais alta resolução (HR). Essa medida indica o grau de similaridade entre as imagens LR e HR, tendo valor igual a zero quando as duas imagens são idênticas. A medida é dada pela Equação (5.2) em que M e N são as dimensões das imagens.

$$RMSE = \sqrt{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [LR(x, y) - HR(x, y)]^2} \quad (5.2)$$

- Índice de Similaridade Estrutural: o índice de similaridade estrutural (SSIM, do inglês *Structural SIMilarity*) foi proposto em Wang *et al.* (2004). Este índice também é utilizado para mensurar a similaridade entre duas imagens. A qualidade é medida comparando-se as correlações locais em luminância, contraste e estrutura entre a imagem de referência e a imagem a ser avaliada. O SSIM é dado pela Equação (5.3) em que as constantes C_1 , C_2 e C_3 são utilizadas para estabilizar a equação quando $(\mu_x^2 + \mu_y^2)$ e $(\sigma_x^2 + \sigma_y^2)$ são próximos de zero. Como sugerido por Wang *et al.* (2004), utilizou-se $C_1 = 0,01$ e $C_2 = 0,03$.

O SSIM é dado por

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{(\sigma_x^2 + \sigma_y^2 + C_2)} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (5.3)$$

onde os termos:

- $l(x, y)$ é a luminância (μ) local da imagem;
- $c(x, y)$ é o contraste (σ); e
- $s(x, y)$ é a correlação estrutural.

Wang *et al.* (2004) simplificaram a equação 5.3, fazendo $C_3 = C_2 / 2$, que resulta em

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5.4)$$

sendo que:

$$\mu_x = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad e \quad \sigma_x = \left(\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (5.5)$$

5.5 Etapas para a geração das imagens de super-resolução a partir de dados reais com elementos textuais

Nesta etapa é utilizado o conjunto B contendo imagens que vincula elementos textuais e que foram obtidas com câmera de alta resolução. Estas imagens foram adquiridas em 3 megapixels e 12 megapixels, para que haja a possibilidade de comparação de imagens adquiridas com alta resolução e imagens de super-resolução geradas a partir das imagens de menor resolução.

Durante a aquisição foram gerados deslocamentos na câmera para aquisição de imagens de diferentes pontos de vista, obtendo-se o conjunto B. A partir da fusão das imagens da sequência é possível ao operador selecionar a área de interesse que será utilizada para a construção da imagem de super-resolução. Essa área de interesse é equivalente a seleção de uma imagem de referência conforme descrito na seção 5.4.

O objetivo de se utilizar elementos textuais nas imagens é o de prover o reconhecimento de caracteres de maneira automática a partir de elementos contidos nas imagens que não são possíveis de serem detectados visualmente.

5.6 Etapas para a geração das imagens de super-resolução a partir de mosaicos

Para a geração das imagens de super-resolução a partir de mosaicos é adquirida uma sequência de imagens aéreas por uma câmera. Em seguida, é realizado o registro de imagens, com o objetivo de se determinar os parâmetros de correlação existentes entre as imagens utilizando os algoritmos SIFT ou SURF seguido pelo RANSAC, que elimina as falsas correspondências e estima a matriz homográfica. A etapa seguinte consiste na geração do mosaico e armazenamento dos pontos correlacionáveis em listas dinâmicas cruzadas que serão utilizados na construção das imagens de super-resolução a partir de uma área definida pelo operador. Com isso, o acesso aos pontos em comum às várias imagens é facilitado, otimizando muito o algoritmo construído.

O processo completo é evidenciado na Figura 5.5.

Figura 5.5: Visão geral da metodologia proposta.

Durante a etapa de construção do mosaico é necessário rotular cada ponto à imagem original a qual o mesmo pertence, para que seja possível a geração de uma imagem de maior resolução com base na sequência original que levou à formação do mosaico.

A estrutura utilizada na implementação desta etapa é detalhada nesta seção, iniciando-se pelas estruturas de dados que possibilita o armazenamento em memória de elementos utilizados durante a retificação de cada pixel, e organizando-os em listas dinâmicas cruzadas.

A implementação realizada permite que estas listas sejam expandidas em qualquer sentido, deslocando os dados antigos e seus índices conforme a inserção necessitar. Por exemplo: inserir 10 colunas do lado esquerdo da lista implicará que se o antigo índice (0,0) existir, ele será movido para o índice (0,10). Os elementos das listas podem ser verificados na Figura 5.6.

Figura 5.6: Listas Dinâmicas Cruzadas para armazenamento dos pixels das imagens que pertencem ao mosaico.

Foram utilizados diversos métodos e os seguintes registros que definem as estruturas de dados (Listas Dinâmicas Cruzadas) para armazenamento dos pixels das imagens pertencentes ao mosaico:

```
struct Celula {
    int linha, coluna;
    LISTA_INDICE *lista;
    Celula *proxima_linha,*proxima_coluna;
};

struct MatLinCol {
    int linha, coluna;
    MatLinCol *proxima_linha,*proxima_coluna;
    Celula *proxima_celula;
};

struct Descritor {
    MatLinCol *primeira_linha,*primeira_coluna;
};
```

A seguir há a descrição dos campos das estruturas de dados utilizadas para melhor entendimento:

a) Célula:

- Este registro existe para representar uma posição (X,Y) na lista encadeada e armazenar os índices das imagens que possuem pontos correlacionados a uma determinada área da imagem mosaico gerado. Para isso, é necessária a existência dos seguintes campos:

Linha e Coluna:

- Este campo armazena a posição (X,Y) na lista encadeada que é a mesma posição (X,Y) no mosaico obtido.

Lista

- Este campo armazena o início de uma lista dinâmica simplesmente encadeada onde cada elemento da lista representa um índice de uma imagem que contém um pixel que é correlacionado com as demais imagens.

Proxima_Linha e Proxima_Coluna

- Este campo armazena o ponteiro para a próxima célula da linha ou para a próxima célula da coluna das listas encadeadas.

b) **MatLinCol**

- Esta estrutura é utilizada para armazenar o começo da linha ou coluna indicada na lista encadeada;

Linha e Coluna

- Este campo armazena a coluna ou a linha na lista encadeada.
- Possui os seguintes campos:

Proxima_Linha e Proxima_Coluna:

- Este campo armazena o ponteiro para a próxima coluna ou linha dos cabeçalhos da linha ou coluna da matriz.

Proxima_Celula

- Este campo armazena o ponteiro para a próxima célula na lista encadeada.

c) **Descritor**

- Possui os seguintes campos:
`*primeira_linha, *primeira_coluna;`
- Armazenam elementos com capacidade de acessar qualquer elemento que faz parte do mosaico gerado.

A construção das listas cruzadas é feita durante o processo de geração do mosaico. As listas crescem conforme cada iteração do processo, e isso pode acontecer em qualquer direção, mantendo os dados anteriormente armazenados, sem a necessidade de reorganização dos mesmos para a manutenção de seus índices. Este processo é feito inserindo colunas ou linhas negativas (inserir para esquerda ou para cima) ou positivas (inserir para a direita ou para baixo). Se forem inseridas colunas negativas, toda a estrutura é então corrigida, somando cada índice da Célula e da estrutura que forma a grade (Matlincol) com o valor de linhas inseridas para linha ou coluna.

Por serem listas cruzadas, somente serão armazenadas as posições correspondentes no mosaico que possuem, no mínimo, um pixel que foi obtido de alguma imagem considerada para o processo de construção do mosaico.

Com listas cruzadas em memória, o processo de gravação e recuperação binária do disco ocorre da seguinte forma:

- a) a estrutura `HEADER_MIMAGE` é gravada no início do arquivo, contendo os tamanhos das listas (altura e largura), a quantidade de imagens que foi utilizada para a criação do mosaico e um vetor de caracteres com o texto `DRI`;
- b) todos os endereços das imagens utilizadas no processo são gravados sequencialmente, onde sua posição significa seu índice que será utilizado para os índices possíveis das células;
- c) nesta etapa, são percorridas todas as posições possíveis das listas. Durante este processo, toma-se uma das seguintes decisões até que todas elas tenham sido percorridas:
 - se a posição atual não existe na lista, é gravado um `CUSTOM_PIXEL` com um índice inválido (255) e com o campo 'nova' verdadeiro.
 - se a posição atual existir é gravado um `CUSTOM_PIXEL` com o índice da primeira posição da lista de índices e com o valor de 'nova' em verdadeiro. Se a lista possuir mais índices, cada um deles também é gravado no arquivo e um novo `CUSTOM_PIXEL` também é gravado, mas todos estes adjacentes possuirão no campo 'nova' o valor falso.

O processo inverso é possível: a construção da lista com os valores dos índices. Este processo segue os seguintes passos:

- a) é lida do arquivo uma estrutura `HEADER_MIMAGE`, e se o campo 'info' possui o valor `DRI`, então o algoritmo continua, pois é um arquivo no formato correto. Esta

- estrutura informa a quantidade de NOME_IMAGEM que deverá ser lida e o tamanho total em altura e largura das listas;
- b) conforme a quantidade obtida pelo HEADER, é lida a quantidade de registros NOME_IMAGEM, armazenando estes endereços em um outro vetor de endereços para o processamento posterior, se necessário;
 - c) neste passo, é preparada uma forma de incrementar linha e coluna conforme os passos i), ii) e iii) e é lido do arquivo a mesma estrutura, CUSTOM_PIXEL, até o fim do arquivo e para cada leitura é tomada uma das seguintes decisões:
 - i. Se o valor do campo 'indiceImagem' for igual a 255, não é criada uma Célula, mas é incrementado a coluna pois esta leitura significou que as listas não devem possuir um valor nesta posição de linha e coluna. Nesta situação não faz diferença o valor do campo 'nova', mas ele deve ser verdadeiro;
 - ii. Se o valor do campo 'indiceImagem' possui um valor diferente de 255 e o campo 'nova' for verdadeiro, é criado uma Célula na matriz na posição de linha e coluna atual com o índice lido;
 - iii. Se o valor do campo 'nova' for falso, o índice lido será inserido ordenadamente e sem repetições na lista da última Célula inserida. Este processo pode se repetir.

Conforme incrementa o índice da coluna e este alcançar o tamanho de largura descrito no HEADER, seu valor é zerado e o índice da linha é incrementado.

O arquivo de dados (binário) possui cabeçalho e organização conforme Figura 5.7.

Cabeçalho do arquivo	Nome das imagens	N1	N2	N...	Índices dos pixels	I1	I2	I...
----------------------	------------------	----	----	------	--------------------	----	----	------

Figura 5.7: Exemplo da organização de um arquivo de dados.

Após o processo de criação do mosaico e das listas com os índices em memória, estas são gravadas em disco para que seja possível a recuperação dessas informações para a geração de novas imagens de super-resolução sem a necessidade de determinar novamente o mosaico.

A estrutura do arquivo utiliza-se dos registros a seguir:

```

struct HEADER_MIMAGE {
    char info[4];
    int altura, largura;
    short quantidadeImagens;
};

struct NOME_IMAGEM {
    char nome[130];
};

struct CUSTOM_PIXEL {
    byte indiceImagem;
    bool nova;
};

struct LISTA_INDICE {
    byte indiceImagem;
    LISTA_INDICE* proxima;
};

```

HEADER_MIMAGE: este registro é o cabeçalho do arquivo. Ele deve sempre ser o primeiro a ser gravado e possui os campos:

- Info: atributo para descrever o tipo deste arquivo. Arquivos que não possuírem DRI neste campo não serão abertos;
- Altura e largura: altura e largura da estrutura em memória que deve ser criada. Ela possui o mesmo tamanho da imagem panorâmica gerada;
- Quantidade imagens: este campo determina a quantidade de imagens que foram analisadas. Ela determina a quantidade de Nomes de imagem que estão gravados logo após este cabeçalho.

NOME_IMAGEM: esta estrutura armazena o endereço de uma das imagens referenciadas e possui somente o campo nome.

- Nome: o endereço da variável.

CUSTOM_PIXEL

Este registro armazena um índice da lista de índices de uma Célula que estava em memória. Sua gravação é sequencial, e para cada índice na Lista de índices da célula, será inserido um registro, sendo o primeiro com o campo 'Nova' verdadeiro e os seguintes com este campo em falso. Ele possui os seguintes campos:

- Índice Imagem: este campo armazena um índice da imagem;
- Nova: este campo armazena verdadeiro ou falso. Sua funcionalidade é a de determinar se durante a abertura deste arquivo, deve-se continuar associando cada

índice para a mesma posição X,Y em memória, ou deve ser criada uma nova posição em memória para armazenar o índice armazenado em `indiceImagem`.

Após esse processo de organização dos pontos das imagens, o processo de geração da imagem de super-resolução acontece após a seleção da região de interesse do operador no mosaico.

5.6.1 Implementação computacional

A implementação aborda as etapas definidas na Seção 5.3, além de outros algoritmos desenvolvidos com o intuito de se obter experimentos de forma controlada em imagens de tons de cinza ou colorida.

Utilizou-se do conceito de matrizes dinâmicas que é disponibilizado pela linguagem C++, ou seja, as matrizes para armazenamento das imagens são criadas com tamanhos definidos no momento da execução da ferramenta, evitando o desperdício de memória e possibilitando o aumento de desempenho, no que se refere a processamento das imagens. Da mesma forma, o ambiente de desenvolvimento Visual Studio 2010, possibilita o uso de formulários (objetos tipo janela MDI – Múltiplas Instâncias de Janelas) que podem ser criados à medida que o usuário necessite e, por isso, o limite de imagens que se pode utilizar é definido pela quantidade de memória que o hardware possui. Com isso, a navegação entre as funções, visualização das imagens e resultados foram implementados em C#. Parte do processamento das imagens foi desenvolvido em C++ em conjunto com a biblioteca gráfica OpenCV (*Open Source Computer Vision Library*) versão 2.4.2.

A utilização do OpenCV permitiu o uso do SIFT (*Scale-Invariant Feature Transform*) e do SURF (*Speeded-Up Robust Features*), e de outros algoritmos que realizam a detecção e extração de descritores locais, que são utilizados para realizar a retificação entre as imagens da sequência.

A ferramenta desenvolvida disponibiliza a abertura de arquivos de imagens com extensão bmp (imagem de BitMap), jpg (JPEG) e png (PNG), e também, a gravação dos mesmos, além de possibilitar a escolha das imagens que se deseja analisar com base no conteúdo descrito na Seção 5.3. São vários os processos disponibilizados, alguns utilizados apenas para testes, tais como:

- A conversão de imagens coloridas (padrão RGB) para imagens em tons de cinza, onde cada pixel é decomposto em suas componentes, calculando-se o nível de intensidade de cinza conforme Equação (5.6);

$$\text{Valor de brilho resultante} = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \quad (5.6)$$

Onde:

R: Componente *Red* (Vermelha);
 G: Componente *Green* (Verde);
 B: Componente *Blue* (Azul).

- Diferença entre imagens: gera uma imagem diferença (níveis de brilho), a raiz do erro médio quadrático (RMSE) existente entre as mesmas e a similaridade estrutural (SSIM) existente entre as imagens, conforme descritos na Seção 5.4;
- Função para execução de algoritmos de extração de detectores, disponibilizando a ligação ou não dos pontos correlacionáveis entre as imagens (SIFT, SURF, ORB, FAST, MUSER, STAR, FREAK). Esses algoritmos foram utilizados vinculando a biblioteca OpenCV;
- Obter a transformação homográfica vinculando os pontos equivalentes nas imagens com sobreposição;
- Possibilidade de duplicação da imagem para testes;
- Opções para interpolação permitindo a escolha dos métodos: bilinear, vizinho mais próximo, bicúbica e média por área;
- Funções para a geração de deslocamentos em imagens (translação, rotação e escala), gerando-se uma sequência simulada;
- Superamostragem das imagens utilizando os métodos da média por área, do vizinho mais próximo, da interpolação bilinear e da interpolação bicúbica. Esses métodos foram implementados e depois substituídos por funções vinculadas a biblioteca OpenCV;
- Processos de retificação que são executados com pares de imagens, ou seja, aplicado a uma imagem em relação à outra (referência), gerando uma imagem retificada e superamostrada;
- Processos para a fusão das imagens (coloridas ou em tons de cinza) retificadas e superamostradas, ou seja, o operador pode retificar as imagens individualmente, e

fundir apenas as de interesse. A fusão que apresentou o melhor resultado foi com o uso da Mediana;

- Processos de geração do mosaico de imagens, e ainda, para construção da imagem de super-resolução com base nas metodologias descritas nas Seções 5.5 e 5.6, onde o usuário determina os algoritmos a serem utilizados para a determinação dos parâmetros ideais e dispara o processo que resultará na imagem de super-resolução;
- Durante o processo para geração do mosaico, são armazenados em memória os pontos que pertencem às imagens que fazem parte da área selecionada no mosaico, para geração da imagem de super-resolução, por meio de estruturas de dados dinâmicas, como listas encadeadas cruzadas. Após a realização desse processo, esses pontos são armazenados em arquivos binários para agilizar o processo de recuperação da imagem de super-resolução sem a necessidade de reconstrução do mosaico.

5.7 Ferramenta Desenvolvida: exemplo de uso

A ferramenta desenvolvida, conforme descrita na seção anterior necessita de interação junto ao operador, pois é necessária a escolha dos algoritmos a serem utilizados para detecção e extração dos descritores e também para a utilização de métodos de interpolação, definidos e utilizados com o auxílio da biblioteca OpenCV versão 2.4.2. Todos os outros algoritmos foram desenvolvidos durante a construção da ferramenta, algoritmos esses que são mostrados na Figura 5.8 por meio dos menus da ferramenta.

Para mostrar a ferramenta em uso, foi definido um pequeno exemplo para a geração de imagens de super-resolução selecionando várias imagens para posterior interpolação e fusão entre as imagens.

Figura 5.8: Tela principal da ferramenta desenvolvida.

A partir do menu Arquivo, pode-se acessar a opção ‘Gerar Mosaico (planar)’ que solicita ao seu usuário alguns parâmetros que são os algoritmos a serem utilizados, conforme mostra a Figura 5.9 na definição dos algoritmos de detecção e extração dos descritores.

Figura 5.9: Algoritmos de detecção e extração dos descritores disponibilizados pela biblioteca OpenCV versão 2.4.2.

A partir dessa escolha, são obtidas as matrizes homográficas que compõem os parâmetros de correlação entre as imagens com sobreposição. E em seguida, o processo executa a fusão das imagens que compõem a cena, disponibilizando a imagem de super-resolução gerada de maneira direta, conforme Figura 5.10.

Figura 5.10: Tela com a imagem de super-resolução gerada.

A ferramenta possibilita muitas outras funções, envolvendo vários algoritmos que foram desenvolvidos durante o processo de implementação da mesma, conforme seções 5.6 e 5.6.1.

5.8 Considerações Finais

A metodologia da ferramenta descrita neste capítulo possibilita a geração de imagens de super-resolução com a construção prévia do mosaico ou apenas por meio da fusão das imagens que possuem pontos de sobreposição entre as imagens que pertencem à sequência.

Optou-se por imagens com as características descritas para realizar testes diversificados de aplicabilidade do trabalho.

Capítulo 6

Resultados e Discussões

Este capítulo apresenta os experimentos realizados e validados pela ferramenta com base na metodologia adotada no Capítulo 5. A ferramenta desenvolvida foi testada com dados simulados e com dados reais. Foram realizados vários experimentos com imagens coloridas e tons de cinza no decorrer do trabalho, para verificar a melhor sequência durante a aplicação dos métodos e estabelecer a imagem de super-resolução a partir de sequências de imagens que formam o mosaico. Após a determinação de uma área de interesse do operador, a ferramenta possibilita extrair informações detalhadas de uma região específica do mosaico.

A partir da Seção 6.1 são apresentados os experimentos com maior relevância.

6.1 Avaliação dos Resultados

O processo de avaliação da qualidade da imagem de super-resolução (SR) obtida em relação à imagem de referência, considerando que esta última é uma imagem de menor resolução (LR), não é uma tarefa simples, uma vez que essas imagens possuem resoluções diferentes. Para a utilização de métricas de avaliação de imagens é necessário ter imagens como padrão de qualidade a ser comparado com as imagens obtidas, pois esse processo é realizado apenas em conjuntos de imagens com dados simulados. O outro processo de avaliação é realizado confrontando as imagens visualmente, com o intuito de mostrar a quantidade de informações recuperadas após a aplicação da metodologia descrita que não eram perceptíveis anteriormente.

Antes de serem apresentados os resultados referentes às metodologias descritas para a geração de imagens de super-resolução das seções 5.4, 5.5 e 5.6, será apresentado na seção 6.2 a imagem de super-resolução obtida a partir de um conjunto de imagens simuladas de menor resolução com o intuito de ilustrar, inicialmente, o funcionamento da ferramenta.

6.2 Experimentos com dados simulados – Conjunto B

Para o teste inicial da ferramenta foram utilizadas imagens simuladas, ou seja, o conjunto de imagens foi adquirido simulando uma sequência que compõe uma determinada cena (resolução 2048x1536 – 3 megapixels) pela câmera Nikon P500. Essa sequência foi obtida por meio da aplicação de movimentos de translações, rotações e escalas na câmera.

A cena utilizada é mostrada na Figura 6.1.

Figura 6.1: Cena utilizada para adquirir uma sequência de imagens a partir de operações de translações, rotações e escalas na câmera.

O processo para gerar a imagem de super-resolução é realizado por meio das sequência das imagens de 3 megapixels, obtendo uma imagem com resolução de 12 megapixels, e assim, são realizadas comparações entre a imagem original de 12 megapixels e a imagem gerada, que ficou com a mesma resolução, conforme ilustrado na Figura 6.2.

Figura 6.2: Processo de comparação da imagem gerada (12 megapixels) com a imagem adquirida (12 megapixels).

Figura 6.3: Processo geral, que envolve a simulação de uma sequência de imagens e validação da imagem de alta resolução gerada.

O processo geral para a geração dessa imagem de maior resolução é ilustrado na Figura 6.3. Foi realizado por um processamento pixel a pixel, com a utilização da interpolação bicúbica para determinar o nível de brilho do pixel correspondente em cada imagem de menor resolução, obtendo-se, após isso, a fusão das imagens por meio da mediana dos valores de cada posição em todas as que pertencem à sequência (Figura 6.4). A imagem gerada pode então ser comparada visualmente com a imagem adquirida em 12 megapixels, conforme ilustra a Figura 6.2.

Figura 6.4: Processo de geração para comparação da imagem construída (12 megapixels) com a imagem adquirida (12 megapixels).

Na Figura 6.5, observa-se a sequência de imagens (2048 x 1536) obtidas após os deslocamentos, rotações e escalas aplicados à câmera Nikon P500 no momento da aquisição das imagens de 3 megapixels.

Figura 6.5: Conjunto de imagens de 3 megapixels adquiridas após aplicação de translação, rotação e escala à câmera.

Conforme a Figura 6.5, as imagens foram adquiridas com menor resolução (3 megapixels) e submetidas ao algoritmo SIFT para a extração dos descritores e, em seguida, ao algoritmo RANSAC que determina os melhores pontos para o registro das imagens. A etapa final consiste na fusão das imagens por meio do valor mediano das imagens com sobreposição.

A Tabela 1 mostra os parâmetros estabelecidos pelo SIFT e SURF, gerando assim a matriz homográfica em relação à translação, rotação e escala. Nesta mesma tabela é mostrada também a quantidade de pontos chave detectados e os descritores extraídos para a execução do *matching* entre as imagens que compõem a cena utilizada.

Tabela 1: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens em relação à imagem de referência Imagem5.jpg.

		Escala em x	Escala em y	Rotação em x	Rotação em y	Translação em x	Translação em y	Pontos chave	Descritores
Imagem0	SIFT	1.1532	1.1072	-0.0125	0.0183	78.7251	-20.5248	1480	1523
	SURF	1.1090	1.0854	-0.0319	0.0209	95.6830	-0.8244	3002	3955
Imagem1	SIFT	0.1506	0.1067	1.1318	-1.1344	1667.4453	-341.0164	1494	1523
	SURF	-1.2200	-1.2254	-1.0365	-1.4424	1217.5079	1034.3651	3349	3955
Imagem2	SIFT	0.6878	0.6578	-0.9575	0.9485	-218.8255	1186.3836	1681	1523
	SURF	-0.4789	-1.2146	-0.1613	-3.6073	1588.7877	534.95228	2511	3955
Imagem3	SIFT	-1.0808	-1.1253	0.2082	-0.2224	2198.4185	1304.3538	1596	1523
	SURF	-2.2288	1.5571	-1.7429	1.9912	534.2811	417.8151	3520	3955
Imagem4	SIFT	0.8889	0.8306	-0.8391	0.8213	-319.8285	1027.2006	1792	1523
	SURF	0.1178	-1.7637	0.1162	-1.7879	1120.6936	1105.5696	3219	3955
Imagem5	SIFT	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1523	1523
	SURF	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	3955	3955

Por meio dos parâmetros obtidos pelo algoritmo SIFT na Tabela 1, são visualizadas na Figura 6.6 as correspondências existentes entre a imagem de referência (Imagem5.jpg) e a imagem de super-resolução gerada.

Figura 6.6: Correspondências encontradas por meio do algoritmo SIFT entre a imagem de referência e a imagem obtida.

A Figura 6.6 ilustra os pontos da imagem de referência com correspondências na imagem obtida, com base nos parâmetros definidos pelo SIFT que são mostrados na Tabela 1.

A Figura 6.7 mostra os pontos correspondentes, com base nos parâmetros definidos pelo algoritmo SURF, entre as imagens de referência e a gerada.

Confrontando a Tabela 1, a Figura 6.6 e a Figura 6.7, pode-se validar o procedimento realizado, ratificando que o algoritmo SURF estabelece quantidade superior de pontos chave em relação ao SIFT. Mas é possível verificar também, que o SIFT apresenta pontos de correlação melhores, pois define com maior qualidade quais os pontos que melhor se correlacionam, enquanto que o algoritmo SURF utiliza pontos que apresentam correspondências concentradas na região central da imagem, além de não definir uma simetria envolvendo toda a imagem, como acontece com o algoritmo SIFT.

Figura 6.7: Correspondências encontradas por meio do algoritmo SURF entre a imagem de referência e a imagem obtida .

A Figura 6.8 mostra o resultado obtido após a fusão das imagens da Figura 6.5 utilizando a “Imagem5.jpg” como a imagem de referência para todo o processo. Ou seja, a Imagem5.jpg é superamostrada em 2X por meio das informações obtidas nas demais imagens por meio do algoritmo SIFT. E o processo final de fusão é definido com a mediana dos valores correlacionados.

Figura 6.8: Resultado obtido após processamento das imagens da Figura 6.5.

Realizando a comparação da imagem adquirida com 12 megapixels (Figura 6.9(a)) e as imagens obtidas (Figura 6.9(b) e Figura 6.9(c)) a partir de uma sequência de imagens de 3 megapixels, é possível verificar que a imagem apresenta diminuição das altas frequências. Isso ocorre porque o método utilizado para a fusão dos pixels que apresentaram correlação foi o da mediana. Devido a isso o efeito *aliasing* é minimizado, conforme ilustra a Figura 6.9(a).

A diferença existente entre a Figura 6.9(b) e Figura 6.9(c) está no algoritmo utilizado para a realização do registro das imagens da sequência, pois na primeira foi utilizado o SIFT e na segunda o SURF. Esse resultado ocorreu em virtude do *matching* falho obtido pelo SURF.

Confrontando os resultados vinculados à Figura 6.10 e a Figura 6.11, é possível verificar pela diferença em níveis de pixels que a Figura 6.10 está mais próxima da imagem original, enquanto que a Figura 6.11 apresenta maiores variações e por isso a diferença no valor gerado pela RMSE. A métrica SSIM mostra que as imagens geradas mantêm, praticamente, a mesma similaridade em relação à estrutura da imagem obtida.

Figura 6.9: Visualização de parte da imagem de 12 megapixels adquirida e da imagem de super-resolução gerada (Figura 6.8).

Figura 6.10: Diferença entre a imagem de 12MP e a imagem gerada com uso do SIFT.

Figura 6.11: Diferença entre a imagem de 12MP e a imagem gerada com uso do SURF.

A Figura 6.12 ilustra a diferença entre as imagens geradas com o uso do SIFT e a imagem gerada com o uso do SURF. As áreas mais escuras apresentam regiões de mínima ou nenhuma diferença entre as mesmas, além disso, o índice estrutural mostra a grande similaridade entre as imagens.

Figura 6.12: Diferença entre a imagem gerada com o uso do SIFT e a imagem gerada com uso do SURF.

6.3 Experimentos com dados reais

Nesta seção são apresentados os resultados e discussões referentes às Seções 5.4, 5.5 e 5.6 da metodologia.

6.3.1 Experimento 1 – Sequência de imagens com dados reais sem adição de ruídos (Conjunto A2)

Neste experimento são utilizadas as imagens do Conjunto A2. A Figura 6.13 apresenta as imagens obtidas originalmente.

Para a realização dos experimentos, as imagens adquiridas pela câmera digital Kodak DC210, foram subamostradas, ou seja, reduzidas à metade das dimensões em relação à imagem original, para que se possa observar a eficiência do método utilizado, pois a imagem gerada terá a mesma resolução geométrica da imagem original, possibilitando a verificação da raiz do erro médio quadrático (RMSE) e o índice de similaridade (SSIM) entre ambas em relação à resolução radiométrica. A imagem referência é mostrada na Figura 6.13(d).

Figura 6.13: Sequência de imagens originais adquiridas com a câmera digital Kodak DC-210.
Fonte: Almeida e Tommaselli (2003).

Na Tabela 2 encontram-se os parâmetros vinculados a matriz homográfica de cada imagem da sequência em relação à imagem de referência, que neste caso é a Figura 6.13 (d). A Tabela 2 mostra também, a quantidade de pontos de controle obtidos pelo SIFT e SURF.

Tabela 2: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.13) em relação à imagem de referência Dcp01615.bmp.

		Escala em x	Escala em y	Rotação em x	Rotação em y	Translação em x	Translação em y	Pontos chave	Descritores
Dcp01612.bmp	SIFT	1.0372	1.0185	0.0169	-0.0032	-32.7878	-5.4988	115	121
	SURF	1.0298	1.0191	0.0137	0.0044	-32.5290	-5.0042	319	317
Dcp01613.bmp	SIFT	1.0427	1.0249	0.0102	0.0092	-30.2203	-5.5346	107	121
	SURF	1.0376	1.0234	0.0076	0.0091	-29.6572	-5.1647	317	317
Dcp01614.bmp	SIFT	0.9954	0.9998	-0.0010	0.0023	-21.4263	-0.0835	103	121
	SURF	0.9902	0.9962	-0.0029	0.0023	-20.9232	0.3617	301	317
Dcp01615.bmp	SIFT	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	121	121
	SURF	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	317	317
Dcp01616.bmp	SIFT	1.0085	1.0025	0.0027	-0.0003	-4.9143	-1.7149	118	121
	SURF	1.0043	1.0023	0.0002	0.0009	-4.3784	-1.2048	337	317
Dcp01617.bmp	SIFT	1.0123	1.0049	0.0035	0.0001	-18.9347	-1.9474	112	121
	SURF	1.0102	1.0062	0.0024	0.0015	-18.6807	-1.9396	332	317
Dcp01618.bmp	SIFT	1.0001	1.0011	-0.0010	0.0059	-24.0846	-0.0272	107	121
	SURF	0.9961	0.9981	-0.0009	0.0051	-23.4797	0.0230	308	317

A Figura 6.14 mostra a imagem de super-resolução gerada a partir das imagens subamostradas da Figura 6.13 com o uso do algoritmo SIFT.

Figura 6.14: Imagem de super-resolução a partir das imagens da Figura 6.13.

O resultado da diferença entre a imagem de super-resolução, com uso do SIFT, e a imagem original é mostrado na Figura 6.15, juntamente com a RMSE e o SSIM. O SSIM é bastante elevado indicando a alta similaridade entre as imagens, e o RMSE é o menor valor em comparação com os experimentos com adição de ruídos das Seções 6.3.2 e 6.3.3.

Figura 6.15: Diferença entre a imagem gerada com uso do SIFT e a imagem original com a raiz do erro médio quadrático e o índice de similaridade.

A Figura 6.16 mostra o resultado da diferença da imagem gerada com o uso do SURF e a imagem original. Pode-se observar que os valores da RMSE e do SSIM são muito próximos aos obtidos com o SIFT, conforme os exemplos da viagem. Nesse experimento que utiliza, durante a fusão das imagens com sobreposição, o elemento de valor mediano entre as imagens que apresentam sobreposição.

Figura 6.16: Diferença entre a imagem gerada com uso do SURF e a imagem original com a raiz do erro médio quadrático e o índice de similaridade.

6.3.2 Experimento 2 – Sequência de imagens com dados reais com aplicação de ruídos de contraste (Conjunto A2)

Neste experimento são utilizadas as imagens do Conjunto A2. A Figura 6.17 apresenta as imagens com adição de ruídos de contraste SHARP2, aplicados pelo software GIMP 2.8.0.

A sequência de imagens adquiridas pela câmera digital Kodak DC210, foi subamostrada, ou seja, cada imagem foi reduzida à metade de suas dimensões em relação à imagem original, para que se possa observar a eficiência do método utilizado, pois a imagem gerada terá a mesma resolução geométrica da imagem original, possibilitando a verificação da raiz do erro médio quadrático (RMSE) e a relação de similaridade (SSIM) entre ambas em relação à resolução radiométrica. A imagem de referência é mostrada na Figura 6.17 (d).

Figura 6.17: Sequência de imagens originais adquiridas com a câmera digital Kodak DC-210 com aplicação de ruídos SHARP2.

Fonte: Adaptado de Almeida e Tommaselli (2003).

A Tabela 3 apresenta os parâmetros reais, extraídos pelo algoritmo SIFT e SURF, e ainda, selecionados pelo RANSAC, com o intuito de realizar o registro de cada imagem da sequência em relação à imagem de referência que faz parte da cena.

Tabela 3: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.17) em relação à imagem de referência SHARP2IBSubDCP01615.bmp.

		Escala em x	Escala em y	Rotação em x	Rotação em y	Translação em x	Translação em y	Pontos chave	Descritores
SHARP2IBSub Dcp01612.bmp	SIFT	1,0346	1,0191	0,0161	-0,0026	-32,2815	-5,5565	141	136
	SURF	1,0373	1,0220	0,0180	-0,0015	-32,8052	-6,0578	1023	983
SHARP2IBSub Dcp01613.bmp	SIFT	1,0390	1,0276	0,0066	0,0098	-29,7635	-5,2753	139	136
	SURF	1,0422	1,0265	0,0118	0,0074	-29,9557	-6,1226	991	983
SHARP2IBSub Dcp01614.bmp	SIFT	0,9903	0,9957	-0,0029	0,0017	-20,6107	0,5632	152	136
	SURF	0,9926	0,9993	-0,0017	0,0034	-21,2821	-0,0357	1025	983
SHARP2IBSub Dcp01615.bmp	SIFT	1,0000	1,0000	0,0000	0,0000	0,0000	0,0000	136	136
	SURF	1,0000	1,0000	0,0000	0,0000	0,0000	0,0000	983	983
SHARP2IBSub Dcp01616.bmp	SIFT	1,0100	1,0054	0,0020	-0,0015	-4,7895	-1,7550	146	136
	SURF	1,0103	1,0058	0,0040	0,0004	-5,1799	-2,3999	1041	983
SHARP2IBSub Dcp01617.bmp	SIFT	1,0162	1,0093	0,0038	0,0026	-19,4512	-2,0581	146	136
	SURF	1,0126	1,0071	0,0032	0,0027	-19,2941	-2,0182	1011	983
SHARP2IBSub Dcp01618.bmp	SIFT	0,9955	1,0005	-0,0016	0,0065	-23,3771	-0,1565	141	136
	SURF	0,9974	0,9977	0,0002	0,0018	-23,3489	-0,3483	988	983

A Figura 6.18 mostra a imagem de super-resolução gerada, com uso do SIFT, onde visualmente observa-se o realce das altas frequências. A Figura 6.19 mostra o resultado da subtração entre esta imagem e a original, juntamente com os valores do RMSE e SSIM. O valor do RMSE é mais elevado do que do experimento 3 (Seção 6.3.1), sem ruídos adicionados. Apesar do RMSE elevado, o SSIM é de 83%, mostrando que há uma boa similaridade entre as duas imagens.

Na Figura 6.20, é ilustrada a diferença entre a imagem de super-resolução gerada, com o uso do SURF para obtenção dos pontos de correlação entre as mesmas. Os valores retornados pelas métricas de avaliação são muito próximos dos valores vinculados à Figura 6.19, que foi utilizado o algoritmo SIFT.

Figura 6.18: Imagem resultante do registro e fusão da sequência evidenciada na Figura 6.17.

Figura 6.19: Diferença entre a imagem gerada (Figura 6.18), com uso do SIFT, e a imagem original (Figura 6.17(d)) com a raiz do erro médio quadrático e o índice de similaridade estrutural.

Figura 6.20: Diferença entre a imagem gerada, com uso do SURF, e a imagem original (Figura 6.17(d)) com a raiz do erro médio quadrático e o índice de similaridade estrutural.

6.3.3 Experimento 3 – Sequência com dados reais e adição de ruídos aleatórios (Conjunto A2)

Neste experimento, foi utilizado o Conjunto A2 (Seção 5.2). Os parâmetros para o registro e retificação das imagens foram determinados com o algoritmo SIFT. Nas imagens desta sequência, foram aplicados ruídos, fazendo com que ocorresse efeito de degradação nas imagens, simulando dessa forma, imagens adquiridas por dispositivos de baixa qualidade, como por exemplo, câmeras de vídeo de circuito fechado.

A Figura 6.21 mostra a sequência de imagens adquiridas e subamostradas para resolução de 576x432 pixels e degradadas por ruídos aleatórios no padrão RGB.

Figura 6.21: Imagens que compõem uma cena com ruídos coloridos.
Fonte: Adaptado de Almeida e Tommaselli (2003).

A Tabela 4 mostra que após a utilização do SIFT e SURF têm-se os parâmetros vinculados à matriz homográfica, que possibilitam a correspondência entre os pontos de controle de maior importância.

Tabela 4: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.20) em relação à imagem de referência Dcp01615.bmp.

		Escala em x	Escala em y	Rotação em x	Rotação em y	Translação em x	Translação em y	Pontos chave	Descritores
Dcp01612.bmp	SIFT	1.0331	1.0216	0.0155	0.0016	-32.8382	-5.6555	149	163
	SURF	1.0028	0.9875	0.0040	-0.0219	-25.9526	-0.0909	553	558
Dcp01613.bmp	SIFT	1.0209	1.0084	-0.0001	-0.0019	-27.0951	-3.1244	150	163
	SURF	1.0380	1.0642	0.0021	0.0242	-30.3592	-7.9727	554	558
Dcp01614.bmp	SIFT	0.9936	1.0055	-0.0042	0.0115	-22.2548	-0.2264	149	163
	SURF	1.0215	1.0201	0.0123	0.0032	-24.0856	-4.3712	548	558
Dcp01615.bmp	SIFT	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	163	163
	SURF	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	558	558
Dcp01616.bmp	SIFT	1.0277	1.0220	0.0096	0.0159	-8.9276	-4.7846	148	163
	SURF	1.0006	0.9651	0.0077	-0.0118	-4.3538	0.9323	569	558
Dcp01617.bmp	SIFT	1.0113	1.0027	0.0000	-0.0020	-18.5353	-2.6705	154	163
	SURF	1.0353	0.9867	0.0277	-0.0102	-20.2071	-3.2365	598	558
Dcp01618.bmp	SIFT	1.0124	0.9954	0.0117	0.0046	-25.3266	-1.1601	158	163
	SURF	0.9843	0.9887	-0.0060	0.0000	-21.6499	1.3784	550	558

Para a construção da imagem de super-resolução, foi utilizada a imagem de referência DCP01615.bmp subamostrada para que haja a possibilidade de validação do processo de construção. A imagem resultante é mostra na Figura 6.22, podendo-se observar que o ruído foi minimizado.

Figura 6.22: Imagem de super-resolução gerada, com uso do SIFT, a partir da sequência da Figura 6.21.

A Figura 6.23 mostra o resultado da subtração entre a imagem de super-resolução, com uso do SIFT, e a imagem original, destacando o valor da raiz do erro médio quadrático e o índice de similaridade, e ainda, a Figura 6.24, que representa o mesmo processo, mas com o uso do algoritmo SURF. Os resultados são muito próximos, pois a distorção entre as imagens são mínimas.

Em virtude do alto grau de ruídos são justificáveis os valores dessas métricas. Apesar da quantidade de ruídos adicionados, o valor do RMSE apresentou nível menor do que nos experimentos anteriores e o SSIM apresentou um índice de similaridade próximo de 70%, ou seja, menor do que dos experimentos anteriores. Esse valor é justificável em virtude do excesso de ruídos vinculados às imagens da sequência.

Figura 6.23: Diferença entre a imagem gerada, com uso do SIFT, e a imagem original com o erro médio quadrático e o índice de similaridade.

Figura 6.24: Diferença entre a imagem gerada, com uso do SURF, e a imagem original com o erro médio quadrático e o índice de similaridade.

6.3.4 Experimento 4 – Sequência com dados reais (Conjunto A1)

Neste experimento são utilizadas as imagens do Conjunto A1. A Figura 6.25 apresenta as imagens obtidas originalmente.

As imagens adquiridas (conforme Seção 5.2) foram subamostradas (378x252), ou seja, reduzidas à metade das dimensões da imagem original (756x504). Dessa forma a imagem gerada terá a mesma resolução geométrica da imagem original, possibilitando a comparação de ambas e determinação da raiz do erro médio quadrático e da relação de similaridade.

A imagem usada como referência para a construção da imagem de super-resolução é mostrada na Figura 6.25(b).

Figura 6.25: Sequência de imagens originais adquiridas com a câmera digital Kodak DC-40.
Fonte: Almeida e Tommaselli (2003).

A Tabela 5 mostra os parâmetros de translação, rotação e escala, além da quantidade de descritores que determinam o *matching* das imagens da Figura 6.25 em relação a imagem de referência (Figura 6.25(b) – Imagem02.bmp). Esses dados são encontrados na matriz homográfica de transformação.

Tabela 5: Matriz Homográfica com os devidos parâmetros estabelecidos pelo SIFT e SURF da sequência de imagens (Figura 6.25) em relação à imagem de referência Imagem02.bmp.

		Escala em x	Escala em y	Rotação em x	Rotação em y	Translação em x	Translação em y	Pontos chave	Descritores
Imagem01.bmp	SIFT	1.0024	1.0020	0.0013	0.0012	-0.1457	-0.2154	133	163
	SURF	0.9978	0.9976	-0.0004	-0.0006	0.2353	0.1450	512	499
Imagem02.bmp	SIFT	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	136	136
	SURF	1.000	1.0000	0.0000	0.0000	0.0000	0.0000	499	499
Imagem03.bmp	SIFT	1.0000	0.9946	-0.0230	0.0201	-3.2351	12.4777	136	136
	SURF	0.9965	0.9935	-0.0236	0.0192	-2.9234	12.4859	484	499
Imagem04.bmp	SIFT	0.9995	0.9941	-0.0223	0.0190	-5.0447	12.2251	123	136
	SURF	0.9985	0.9948	-0.0222	0.0188	-4.8807	12.1573	483	499

Na Figura 6.26 é mostrada a imagem de super-resolução gerada, com uso do SIFT, e na Figura 6.27 é ilustrada a diferença com a imagem original, juntamente com valores da

RMSE e do SSIM. Nota-se que o índice de similaridade para esse caso é superior ao do experimento da seção 6.3.4 e a raiz do erro médio quadrático é menor, evidenciando uma qualidade superior dessa imagem em relação à Figura 6.22, que possui a adição de ruídos.

Figura 6.26: Imagem resultante do registro da sequência evidenciada na Figura 6.25.

Figura 6.27: Diferença entre a imagem gerada, com uso do SIFT, e a imagem original com o erro médio quadrático e o índice de similaridade.

A Figura 6.28 ilustra a diferença entre a imagem original e a imagem gerada com o uso do SURF, e ainda, os valores estabelecidos pelas métricas de avaliação, que são muito próximos em relação à imagem gerada com o uso do SIFT, conforme Figura 6.26, e valores definidos na Figura 6.27.

Figura 6.28: Diferença entre a imagem gerada com uso do SURF, e a imagem original com o erro médio quadrático e o índice de similaridade.

Na Tabela 6 é apresentado o resumo dos valores obtidos pelas métricas de avaliação utilizadas na comparação entre as imagens de super-resolução geradas, com uso dos algoritmos SIFT e SURF em conjunto com o RANSAC para eliminação dos *outliers*.

Tabela 6: Resumo dos resultados obtidos pelas métricas de avaliação.

Experimentos	Descritores	RMSE	SSIM	Similaridade (%)
Simulado (Seção 6.2)	SIFT	52.671546	0.700719	70
	SURF	85.650255	0.692591	69
Experimento 1 (Seção 6.3.1)	SIFT	51.334284	0.917799	92
	SURF	51.285093	0.913965	91
Experimento 2 (Seção 6.3.2)	SIFT	60.990451	0.830000	83
	SURF	61.010692	0.829456	83
Experimento 3 (Seção 6.3.3)	SIFT	58,711674	0.728875	73
	SURF	60.939315	0.697463	70
Experimento 4 (Seção 6.3.4)	SIFT	41.035120	0.905574	91
	SURF	40.684845	0.907512	91

6.3.5 Experimento 5 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de imagens aéreas urbanas (Conjunto C)

Neste experimento são utilizadas as imagens do Conjunto C. A Figura 6.29 apresenta as imagens aéreas obtidas originalmente. Esse conjunto de imagens (300x300 pixels) utilizado é referente a imagens urbanas com sobreposição, possibilitando assim a construção do mosaico. A partir do mosaico pode-se selecionar uma área do mesmo com o intuito de gerar imagens de melhor resolução a partir da região escolhida.

É realizada a extração de características de cada par de imagens utilizando o algoritmo SIFT com RANSAC. Por meio dessas características, é gerado o mosaico e é selecionada uma área, pelo operador, como ilustrado na Figura 6.30. A partir da área selecionada é construída a imagem de super-resolução mostrada na Figura 6.31.

Figura 6.29: Sequência de imagens aéreas para geração do mosaico.

A Figura 6.30, apresenta o mosaico gerado que possui uma estrutura de armazenamento para a identificação das imagens que fazem parte da fusão no mosaico em relação à área escolhida (em destaque na Figura 6.30) para a geração da imagem de super-resolução.

Figura 6.30: Imagem do mosaico gerado a partir do conjunto de imagens da Figura 6.29.

A região selecionada pelo operador, conforme Figura 6.30, é utilizada como imagem de referência no processo de determinação dos parâmetros de correlação com as imagens que pertencem à cena. E a partir disso, são construídas as listas cruzadas para vincular apenas os pontos das imagens que possuem sobreposição para posterior fusão dos mesmos e geração da imagem de super-resolução (Figura 6.31).

Figura 6.31: Imagem de super-resolução gerada a partir da seleção da região escolhida com uso do SIFT.

A Figura 6.32 mostra o mesmo processo, mas o algoritmo utilizado para extração dos descritores que determinam as correspondências de pontos entre as imagens foi o SURF.

E como pode ser verificado, as imagens apresentam mínima diferença visual. As métricas utilizadas nos experimentos anteriores não são utilizadas nos experimentos que envolvem a construção de mosaicos, pois não há uma imagem que representa o resultado ideal para comparação. E ainda, há a dificuldade de comparação das regiões extraídas, pois dificilmente apresentarão a mesma área no mosaico. Isso pode ser observado na Figura 6.31 e Figura 6.32 que possuem pequenas diferenças de tamanho entre ambas. E por isso, os valores gerados pelas métricas de avaliação não apresentarão precisão.

Figura 6.32: Imagem de super-resolução gerada a partir da seleção da região escolhida com uso do SURF.

Ampliando proporcionalmente regiões do mosaico (Figura 6.30) e da imagem de super-resolução gerada (Figura 6.31), pode-se verificar a minimização do efeito *aliasing* e a melhor definição da imagem gerada, conforme Figura 6.33.

Figura 6.33: Comparação da (a) região maximizada do mosaico (Figura 6.30) com (b) região maximizada da imagem de super-resolução obtida (Figura 6.31).

6.3.6 Experimento 6 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de imagens aéreas agrícolas (Conjunto C)

Neste experimento são utilizadas as imagens do Conjunto C. A Figura 6.34 mostra a sequência de imagens (530x629 pixels) aéreas obtidas originalmente. Esse conjunto de imagens é referente a áreas urbanas e rurais com sobreposição, possibilitando a geração do mosaico de imagens (Figura 6.35). A partir do mesmo pode-se selecionar uma área com o intuito de gerar uma imagem de super-resolução dessa região definida pelo operador.

Por meio do SIFT ou SURF são extraídas as características de cada imagem em relação à região selecionada pelo operador (Figura 6.36). Por meio das características extraídas, é gerada imagem de super-resolução conforme Figura 6.37 e Figura 6.38.

Figura 6.34: Sequência de imagens aéreas para geração do mosaico.

A Figura 6.35 mostra o mosaico a partir da Figura 6.34. E após selecionar uma porção da imagem (Figura 6.35) com o mouse obtém-se o recorte (Figura 6.36) que é utilizado como referência para a geração da imagem de super-resolução. As fusões dos pontos

correlacionáveis acontecem por meio dos parâmetros definidos pelos SIFT (ou SURF) e RANSAC, e o valor mediano entre os mesmos é o utilizado, acarretando uma leve diminuição das altas frequências conforme mostra a Figura 6.37. A Figura 6.38 evidencia o mesmo processo, mas com uso do SURF.

Figura 6.35: Mosaico obtido pelas imagens contidas na Figura 6.34.

Figura 6.36: Região selecionada na Figura 6.35 para geração da imagem de super-resolução.

Figura 6.37: Imagem de Super-resolução gerada a partir da seleção utilizando-se do SIFT.

Figura 6.38: Imagem de Super-resolução gerada a partir da seleção utilizando-se do SURF.

Ampliando proporcionalmente regiões do mosaico (Figura 6.35) e da imagem de super-resolução gerada (Figura 6.37), pode-se verificar a minimização do efeito *aliasing* e a melhor definição da imagem gerada, conforme Figura 6.39.

Figura 6.39: Comparação da (a) região maximizada do mosaico (Figura 6.35) com (b) região maximizada da imagem de super-resolução obtida (Figura 6.37).

6.3.7 Experimento 7 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de imagens aéreas em tons de cinza (Conjunto D)

Neste experimento são utilizadas as imagens do Conjunto D, que estão em níveis de cinza e estão no padrão BMP.

A Figura 6.40 mostra a sequência de imagens em tons de cinza (500x500 pixels) que é utilizada para gerar o mosaico de imagens, portanto é preciso selecionar as imagens de interesse da sequência para que haja a montagem do mosaico, conforme Figura 6.41.

Figura 6.40: Sequência de imagens aéreas em tons de cinza para geração do mosaico.

O mosaico obtido está com 1497x500 pixels e a fusão foi realizada adequadamente conforme Figura 6.41.

Figura 6.41: Mosaico obtido pelas imagens contidas na Figura 6.40.

A partir da seleção de uma porção do mosaico (Figura 6.41) gera-se a imagem de super-resolução realizando a fusão pela Mediana dos pontos descritos pelo SIFT (Figura 6.42) ou pelo SURF (Figura 6.43), obtendo-se assim uma imagem de 1082x442 pixels de tamanho e preservando dados reais das imagens de menor resolução.

Figura 6.42: Imagem de Super-resolução gerada a partir da seleção na Figura 6.41 utilizando-se do SIFT.

Como evidenciado nos experimentos realizados, a diferença entre as imagens representadas pela Figura 6.42 e Figura 6.43 é praticamente nula, pois neste caso as imagens estão em um plano que minimizam os efeitos de distorções.

A Figura 6.43 é gerada a partir de uma região similar a região verificada na Figura 6.42.

Figura 6.43: Imagem de Super-resolução gerada a partir da seleção na Figura 6.41 utilizando-se do SURF.

Ampliando proporcionalmente regiões do mosaico (Figura 6.41) e da imagem de super-resolução gerada (Figura 6.42), pode-se verificar a minimização do efeito *aliasing* e a melhor definição da imagem gerada, conforme Figura 6.44.

Figura 6.44: Comparação da (a) região maximizada do mosaico (Figura 6.41) com (b) região maximizada da imagem de super-resolução obtida (Figura 6.42).

6.3.8 Experimento 8 – Construção do mosaico e geração da imagem de super-resolução por meio de uma sequência de recortes de imagens com elementos textuais (Conjunto B)

Neste experimento são utilizadas as imagens do Conjunto B. A Figura 6.45 mostra a sequência de recortes de imagens coloridas de 3 megapixels, adquiridas pela câmera Nikon

P500, que é utilizada para gerar o mosaico das mesmas para posterior seleção de elementos textuais para análise do quanto é possível recuperar informações por meio dessa metodologia. Esse tipo de experimento possibilita também, a aplicação das imagens geradas em algoritmos de realizam o reconhecimento de caracteres.

Figura 6.45: Sequência de imagens com elementos textuais.

A seleção e fusão das imagens de 3 megapixels evidenciadas na Figura 6.45 utilizando-se do método de interpolação Bicúbica e da Mediana dos pontos que apresentam correlação, obtém uma nova imagem de 12 megapixels, conforme evidenciada na Figura 6.46.

Figura 6.46: Resultado da seleção e fusão da área com elementos textuais.

A partir disso, é possível comparar a imagem resultante (Figura 6.46) com uma imagem da mesma cena adquirida com 12 megapixels, conforme Figura 6.47.

Figura 6.47: Imagem adquirida com 12 megapixels com elementos textuais.

A Figura 6.46 e Figura 6.47 possibilitam a comparação entre imagens obtidas de diferentes maneiras. A Figura 6.46, foi gerada a partir de uma sequência de imagens de menor resolução e a Figura 6.47 foi adquirida com a resolução equivalente à imagem gerada (Figura 6.46). E o resultado da recuperação de informações pode ser observado confrontando a Figura 6.48 e a Figura 6.49, onde apresentam resultados muito significativos, pois elimina ruídos e destaca informações relevantes que podem ser aplicados no reconhecimento de caracteres, por exemplo.

Figura 6.48: Parte da imagem de Super-resolução gerada a partir da seleção na Figura 6.46 utilizando-se do SIFT.

Figura 6.49: Parte da imagem de 12 megapixels adquirida, conforme Figura 6.47.

Foi elaborado um experimento com uso de um algoritmo de reconhecimento de caracteres (SILVA et al., 2011) aplicado nas imagens completas representadas pelas porções mostradas na Figura 6.47, Figura 6.48 e Figura 6.49 como uma forma de análise da qualidade. O experimento foi dividido em duas partes, onde cada parte contém a região da imagem que contém o título de um dos dois livros fotografados nas imagens. Na primeira parte do experimento, foram utilizadas imagens contendo o título de um dos livros ("Inteligência Artificial"). O algoritmo de reconhecimento de caracteres poderia identificar corretamente todos os caracteres contidos na imagem de alta resolução. As regiões que contêm as características avaliadas pelo algoritmo são mostradas na Figura 6.50.

Figura 6.50: Regiões de imagens que contêm os caracteres avaliados na primeira parte do experimento: a) de alta resolução (HR), b) uma das 15 imagens de baixa resolução, usadas para obter uma imagem de super-resolução (SR) obtido com o método proposto.

Nas 15 imagens de baixa resolução, o algoritmo de reconhecimento de caracteres apresentou uma taxa média de acerto de 77,27%, excluindo em média 5 de 22 caracteres, enquanto que para as imagens de super-resolução a taxa de acerto foi de 95,45%, não reconhecendo somente um caractere ('R').

A Tabela 7 mostra a quantidade de erros obtidos em cada uma das imagens 15 de baixa resolução e também a resolução super (SR) e de alta resolução (HR) imagens.

Tabela 7: Número de caracteres incorretamente identificadas nas 15 imagens de baixa resolução, super-resolução (SR) e imagens de alta resolução (HR).

Imagem N°	N° de erros	Imagem N°	N° de erros	Imagem N°	N° de erros
1	4	7	7	13	5
2	6	8	4	14	6
3	3	9	5	15	9
4	4	10	5	SR	1
5	5	11	5	HR	0
6	4	12	3		

Na segunda parte do experimento, as imagens foram usadas contendo o título do segundo livro ("Processamento de imagens, análise e visão de máquina"). A Figura 6.51 mostra em (a) uma imagem contendo um pedaço do título do livro, obtidos em alta resolução, e (b) uma imagem obtida em baixa resolução. Em (c) tem a imagem de super-resolução (SR) obtida com o método proposto.

Figura 6.51: Regiões de imagens que contêm os caracteres avaliados na segunda parte do experimento: a) de alta resolução (HR), b) uma das 15 imagens de baixa resolução, usadas para obter uma imagem de super-resolução, c) imagem de super-resolução (SR) obtido com o método proposto.

Neste experimento foi utilizado o algoritmo de reconhecimento de caracteres de Silva et al. (2011) aplicado nas imagens como uma forma de análise de qualidade, o algoritmo teve uma taxa de sucesso de 87,18%, deixando apenas 5 dos 39 caracteres para bater. Em 15 imagens de baixa resolução, o algoritmo apresentou uma taxa média de acerto de 62,89%, não conseguindo acertar uma média de 14,47 dos 39 caracteres. Na imagem de super-resolução, o algoritmo apresentou uma taxa média de acerto de 82,05%, não conseguindo acertar 7 de 39 caracteres, uma taxa muito maior do que a obtida com as imagens de baixa resolução e muito próximo ao obtido à imagem de alta resolução (HR). A Tabela 8 mostra os erros obtidos em cada uma das 15 imagens de baixa resolução.

Tabela 8: Número de caracteres incorretamente identificados nas 15 imagens de baixa resolução, super-resolução (SR) e imagens de alta resolução (HR).

Imagem Nº	Nº de erros		Imagem Nº	Nº de erros		Imagem Nº	Nº de erros
1	11		7	11		13	21
2	18		8	12		14	11
3	15		9	10		15	12
4	17		10	12		SR	7
5	16		11	19		HR	5
6	14		12	18			

6.4 Discussões

Uma característica importante da metodologia proposta neste trabalho, é que não há a necessidade de se ter um número grande de imagens que compõe a cena. E não há necessidade de se ter um controle específico de como as imagens foram adquiridas, pois algoritmos de detecção e extração de descritores como o SIFT e SURF, determinam pontos-chaves em todas as regiões das imagens, garantindo assim, a simetria ideal para a determinação do registro das imagens, e assim, minimizando os problemas de distorção que podem ocorrer durante esse processo, que é considerado essencial para a metodologia desenvolvida.

Com relação aos experimentos 1, 2, 3 e 4, que são os experimentos simulados, observa-se pelos valores da RMSE e do SSIM que mesmo com a adição de ruídos nas imagens originais, conseguiu-se obter um índice de similaridade acima de 70% em relação à imagem original. E pode ser confirmado pela avaliação visual das imagens de super-resolução resultantes.

Com relação aos experimentos 5, 6 e 7, realizados com os mosaicos, a avaliação visual mostra um resultado satisfatório por não haver degradação das áreas ampliadas.

Com relação ao experimento 8, realizado com imagens de elementos textuais, a avaliação visual mostra que reduz o efeito *aliasing* e a recuperação da informação é realizada de forma satisfatória, podendo-se concluir que a técnica pode ser utilizada no reconhecimento de caracteres, conforme evidenciam a Tabela 7 e a Tabela 8.

Os resultados obtidos utilizando-se SIFT ou SURF para a extração dos descritores são muito similares no que tange a tempo de processamento e visualização dos resultados, principalmente em imagens planares, ou seja, sem grandes distorções.

Capítulo 7

Conclusões e Sugestões

Este trabalho apresentou a proposta de uma metodologia para construir automaticamente imagens de super-resolução a partir de sequências de imagens com sobreposição. A partir dessa metodologia foi implementada uma ferramenta que possibilita a construção de mosaicos com esse tipo de imagem e permite ao operador a seleção de uma área de interesse nos mesmos, a partir da qual, é gerada a imagem de super-resolução.

Para a validação da ferramenta, foram utilizados quatro conjuntos de imagens (A, B, C e D), sendo que o conjunto de imagens A foi usado sem o intuito da geração do mosaico, mas com o objetivo de testar os algoritmos para a obtenção de uma imagem de super-resolução.

O conjunto A possui dois subgrupos A1 e A2, cujas imagens foram obtidas com equipamentos diferentes. Para o experimento envolvendo o conjunto A1 e A2, foram utilizadas câmeras de baixa resolução e para os experimentos 1, 2, 3 e 4 houve o redimensionamento das imagens da sequência para a comparação da imagem de super-resolução com a imagem original. Para os experimentos 1, 3 e 4 também foram adicionados ruídos às imagens com o intuito de gerar uma sequência de imagens degradadas.

Para o conjunto B foi utilizado uma câmera de alta resolução acoplada a um suporte externo de um avião e contém áreas de pastagens, matas e área urbana. Para o conjunto C foi utilizada uma câmera de alta resolução para capturar cenas com elementos textuais. E para o conjunto D, foram utilizadas imagens aéreas em níveis de cinza disponibilizadas pelo departamento de Engenharia Cartográfica da FCT/UNESP – Campus de Presidente Prudente.

Para a geração dos mosaicos, a metodologia proposta baseou-se em um algoritmo de extração de descritores de pontos chave (no trabalho foi utilizado o algoritmo SIFT ou algoritmo SURF), e em seguida foi utilizada a BBF para acelerar a busca por correspondência dos pontos chave, e também, o algoritmo RANSAC para eliminar falsas correspondências (*outliers*) e estimar a matriz homográfica. Os parâmetros fornecidos pela matriz homográfica possibilitam o registro das imagens que pertencem à mesma cena para posterior fusão dos pontos de correlação utilizando o valor mediano entre esses pontos, minimizando assim, o

efeito *aliasing* contidos nas imagens. A escolha desses algoritmos foi baseada nos vários trabalhos científicos (LOWE, 2004) (MIKOLAJCZYK; SCHMID, 2005) (BAY et al., 2008) (MACHADO et al., 2008) (IHARA et al., 2009) (SILVA, 2012), que constata um desempenho superior em qualidade da descrição dos pontos chave em relação a outros algoritmos para a mesma finalidade.

Durante a construção do mosaico são utilizadas estruturas de dados, baseadas em listas dinâmicas cruzadas, que organizam os dados de sobreposição para posterior fusão de áreas selecionadas pelo operador. Estas estruturas de dados agilizam o acesso aos pontos que realmente devem ser utilizados, melhorando o desempenho durante a geração da imagem de super-resolução. Esses dados também são armazenados em arquivos organizados com o intuito de reutilização do mosaico, ou seja, não é necessária a construção do mosaico novamente a fim de gerar novas imagens de super-resolução a partir do mesmo, sendo necessário apenas, associar o mosaico aos dados que vinculam as imagens que fazem parte do mesmo, para que novas áreas sejam selecionadas para obtenção de imagens de super-resolução. Estas estruturas de dados constituem uma das principais contribuições deste trabalho.

A ferramenta desenvolvida neste trabalho foi idealizada para construção de imagens de super-resolução a partir da seleção de uma determinada área de mosaicos de forma automática, onde o operador define as imagens, os algoritmos de detecção, de interpolação e de fusão que deverão ser utilizados durante o processamento. A partir daí, não é necessária qualquer intervenção do operador. E ainda, a ferramenta também não faz uso de, somente métodos de interpolação e filtros para gerar uma imagem de maior resolução, mas analisa mais de uma imagem de um mesmo local, fazendo com que um determinado pixel que não se apresenta com qualidade em uma imagem, seja compensado por outro que se localiza em outra imagem da mesma cena.

Os experimentos mostraram que na metodologia proposta são recuperadas as informações quando da geração da imagem de super-resolução que é analisada por meio das métricas RMSE e SSIM, aplicadas às imagens originais e às de super-resolução obtidas pela ferramenta. A RMSE possibilita uma avaliação quantitativa e a SSIM uma avaliação qualitativa. Além das métricas utilizadas, a avaliação visual também possibilita a verificação dos detalhes fornecidos pelas imagens de super-resolução, comprovando a eficiência da metodologia aplicada. O limitador em relação à quantidade de vezes que se deseja aumentar a resolução geométrica de uma imagem, obtendo resultados aceitáveis radiometricamente, está

vinculado à quantidade de memória disponível e a resolução original das imagens que compõem a sequência.

Os resultados apresentados foram significativos, mostrando a relevância do trabalho na geração de imagens de super-resolução seja utilizando imagens de referências ou selecionando uma área no mosaico gerado, ou ainda, em imagens panorâmicas. Mesmo porque as imagens de super-resolução geradas apresentam melhorias no que diz respeito à restauração das mesmas para futura análise de alvos de interesse, sem ter o retrabalho de adquirir novas imagens da cena, pois dependendo da cena analisada não será possível nova aquisição de imagens.

Como ilustrado pelos experimentos, o presente trabalho pode ser utilizado para geração de imagens de maior resolução a partir de uma sequência de imagens com sobreposição, nas seguintes aplicações:

- utilizando uma imagem como referência;
- selecionando a área de interesse no mosaico;
- utilizando a técnica para reconhecimento de texto.

Embora existam vários trabalhos publicados nessa área que propõem metodologias para a geração de imagens de super-resolução a partir de sequências de imagens (TANG et al., 2008) (NEMRA; AOUF, 2009) (NASIR et al., 2011) (CAMPONEZ et al., 2012) (PURKAIT; CHANDA, 2012) (SALARI; BAO, 2012), as metodologias não se aplicam para uma região específica do mosaico, como no presente trabalho. Outros trabalhos relacionados à geração de imagem de super-resolução utilizam uma imagem de referência, a partir da qual é gerada a imagem de super-resolução (PELEG; IRANI, 1991) (BALLARD; BROWN, 1992) (JOYEUX et al., 2001) (ALMEIDA; TOMMASELLI, 2003) (BAUMANN; SEALES, 2009).

Dessa forma, pode-se destacar que as principais contribuições deste trabalho são as seguintes:

- geração de imagem de super-resolução a partir da sequência que compõe um mosaico, pois é necessário o conhecimento de quais imagens possuem dados que estão vinculados com a área do mosaico que está sendo analisada, para que haja descritores ideais para que o *matching* tenha resultado satisfatório;

- desenvolvimento de estruturas de dados e algoritmos que possibilitam a análise de regiões específicas do mosaico, sem que o mesmo tenha que ser processado integralmente.

7.1 Sugestões de trabalhos futuros

Sugere-se como trabalhos futuros para melhoria da geração e validação das imagens de super-resolução:

- otimização de processamento e memória para diminuir o tempo de geração das imagens de super-resolução;
- uso de técnicas estatísticas para classificação da qualidade das imagens geradas;
- possibilitar ao operador, a seleção de uma região em qualquer formato, ou seja, selecionar uma região na forma de qualquer polígono;
- utilizar as imagens de super-resolução geradas para detecção e reconhecimento de plantação específica, por exemplo, plantação de lavoura proibida, áreas de desmatamento, ou ainda, plantação em locais proibidos;
- minimizar o efeito *ghosting* que ocorre durante a junção das imagens que formam o mosaico;
- reconhecimento de caracteres vinculados a uma cena;
- utilização de processamento paralelo.

Sobre o conjunto de imagens, sugere-se a aquisição de imagens utilizando aeronaves não tripuladas em altitudes que possibilite obter maior qualidade visual e nitidez de detalhes de lavouras, para melhor avaliação da existência de pragas ou doenças.

7.2 Artigos Publicados e Submetidos

7.2.1 Artigos Publicados

Almeida, L. L.; Paiva, M. S. V.; Silva, F. A.; Jorge, L. A. C. Super Resolution Images from Mosaics. In: IX WORKSHOP DE VISÃO COMPUTACIONAL (WVC 2013). **Anais...** Rio de Janeiro: FGV-UFF, 2013.

Almeida, L. L.; Paiva, M. S. V.; Silva, F. A.; Artero, A. O. Super-Resolution Images Enhanced for Applications to Character Recognition. **The The Standard International Journals (SIJ) – Transactions on Computer Science Engineering & its Applications (CSEA)**. Manuscript ID: CSEA-0103520101 – Print ISSN: 2321–2373 | Online ISSN: 2321–2381. 2013.

Almeida, L. L.; Silva, F. A.; Artero, A. O.; Paiva, M. S. V.. Super-resolution image created from a sequence of images with application of character recognition. **International Journal of Intelligent Systems and Applications (IJISA 2013)** – Modern Education and Computer Science Publisher (MECS) – ISSN: 2074-904X (Print), ISSN: 2074-9058 (Online) – DOI: 10.5815/ijisa.

7.2.2 Artigos Aceitos para Publicação

Almeida, L. L.; Paiva, M. S. V.; Silva, F. A.; Jorge, L. A. C. Automatic Image Mosaicing and Super Resolution with SIFT and RANSAC. In: 11th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION AND IMAGE ANALYSIS (PRIA-11-2013). Samara – Russia.

Almeida, L. L.; Paiva, M. S. V.; Silva, F. A.; Jorge, L. A. C. Super resolution image created automatically from mosaics. In: INTERNATIONAL ACADEMIC CONFERENCE ON COMPUTER SCIENCES AND APPLICATIONS (ICCSA 2013). Wuhan – China.

7.2.3 Participação como co-autor em Artigos Publicados

MASUTANI, Victor Hugo; ARTERO, Almir Olivetti; ALMEIDA, L. L.; SILVA, Francisco Assis da. Reconhecimento de Objetos Usando Coloridos e Mãos usando Cores e Formas. **Revista Colloquium Exactarum**, v.5, n.1, Jan-Jun. 2013, pp.01–11. DOI: 10.5747/ce.2013.v05.n1.e050.

MASUTANI, Victor Hugo; ARTERO, Almir Olivetti; ALMEIDA, L. L. ; SILVA, Francisco Assis da; PAIVA, M. S. V.. Reconhecimento de Mãos usando Cores e Formas Aplicado no Projeto de Interface Gestual. Anais do VIII WORKSHOP DE VISÃO COMPUTACIONAL (WVC 2012), v. 1, pp. 1-1, 2012.

SANTOS, Fagner Fernandes; ALMEIDA, L. L.. JAVA ADVANCED IMAGING, uma aplicação em processamento digital de imagens. **Revista Colloquium Exactarum**, v. 2, n.2, Jul-Dez. 2010, pp. 32–41. DOI: 10.5747/ce.2010. v02. n2. e015.

Referências Bibliográficas

Ali, S.; Hussain, M. Panoramic Image Construction using Feature based Registration Methods. In: IEEE Multitopic Conference (INMIC), 2012.

Almeida, L. L. **Geração de Imagens de Melhor Resolução a Partir de Sequências de Imagens**. 2001. 132 f. Dissertação (Mestrado em Ciências Cartográficas) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente, 2001.

Almeida, L. L.; Tommaselli, A. M. G., Melhoria da Resolução a partir de Sequências de Imagens. Boletim de Ciências Geodésicas, 9(2):163–178, Julho-Dezembro 2003.

Andrews, H. C.; Hunt, B. R. Digital Image Restoration. Prentice-Hall, Inc., 1977.

Anscombe, F. J. The Transformation of Poisson, Binomial and Negative-Binomial Data. Biometrika, vol. 15, pp. 246-254, 1948.

Atkinson, K.B. Developments in Close Range Photogrammetry. Essex, England: Applied Science Publishers LTD. 1996.

Arya, S. Nearest neighbor searching and applications. Technical Report CAR-TR-777, Center for Automation Research, 1995.

Baker S.; Kanade T. Super-Resolution Optical Flow. [s.l.]: CMU-RI-TR-99-36, 1999.

Ballard, D.H.; Brown, C.M. **Computer Vision**. New Jersey, USA: Prentice-Hall, 1992.

Bauer, J.; Sünderhauf, N.; Protzel, P. Comparing Several Implementations of Two Recently Published Feature Detectors. In: PROC. OF THE INTERNATIONAL CONFERENCE ON INTELLIGENT AND AUTONOMOUS SYSTEMS. IAV, Toulouse, France, 2007.

Baumann, R.; Seales, W. B. JCDL'09. In: ACM, Austin, Texas, USA, 2009, pp.263-266.

Bay, H.; Tuytelaars, T.; and Gool, L. V. Surf: Speeded up robust features. In: 9th European CONFERENCE ON COMPUTER VISION, 2006, pp. 404-417.

Bay, H.; Ess, A.; Tuytelaars, T.; Gool, L. V. SURF: Speeded Up Robust Features. **Computer Vision and Image Understanding (CVIU)**, v. 110, n. 3, p. 346-359, 2008.

Bei, L.; Haizhen, Z. **An algorithm of fabric image mosaic based on SIFT feature matching**. IEEE International Conference on Artificial Intelligence and Computational Intelligence. p. 435 – 438, 2009.

Beis, J.; Lowe, D. G. Shape indexing using approximate nearest-neighbour search in highdimensional spaces. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. Puerto Rico, p. 1000-1006, 1997.

Belo, F. A. W. **Desenvolvimento de Algoritmos de Exploração e Mapeamento Visual para Robôs Móveis de Baixo Custo**. Rio de Janeiro, 276 p. Dissertação (Mestrado). Pontifícia Universidade Católica do Rio de Janeiro, 2006.

Bentley, J. L. Multidimensional binary search trees used for associative searching. **Communications of the ACM**, 18(9): p. 509-517, 1975.

Berg, M.; Cheong, O.; Kreveld, M.; Overmars, M. Computational Geometry: Algorithms and Applications. **Springer-Verlag**. 3rd. ed. 2008. 386p.

Brown, L. G. A Survey of Image Registration Techniques. *Computing Surveys*; 24(4); 1992.

Brown, M.; Lowe, D .G. Invariant features from interest point groups. In: BRITISH MACHINE VISION CONFERENCE. Cardiff, Wales, 2002, pp. 656-665.

Brown, M.; Lowe, D. G. Automatic Panoramic Image Stitching using Invariant Features. **International Journal of Computer Vision**, v. 74, Issue 1, pp. 59-73, 2007.

Camponez, M. O.; Salles, O. T.; Sarcinelli-Filho, M. Super-Resolution Image Reconstruction Using Nonparametric Bayesian INLA Approximation. **IEEE TRANSACTIONS ON IMAGE PROCESSING**. v. 21, n. 8. 2012. pp. 3491-3501.

Capel, D. Image Mosaicing and Super-Resolution. (CPHC/BCS Distinguished Dissertations). SpringerVerlag. London. 2004.

Chen, J.; Tian, J. Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor. **Progress in Natural Science**. v. 19, pp. 643-651, 2009.

Cho, P.; Bae, S.; Durand, F. **Image-based querying of urban knowledge databases**. XVIII Signal Processing, Sensor Fusion, and Target Recognition (*The International Society for Optical Engineering – SPIE*), 2009.

Farag, A. A.; Abdel-Hakim, A. E. Detection, Categorization and Recognition of Road Signs for Autonomous Navigation. In: ADVANCED CONCEPTS FOR INTELLIGENT VISION SYSTEMS (ACIVS 2004). **Proceedings...** Brussels, Belgium, Aug. 31-Sept. 3, 2004. p. 125-130.

Fischler, M. A.; Bolles, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. **Communications of the ACM**, v. 24, p. 381-395, 1981.

Felzenszwalb, P. F.; Huttenlocher, D. P. Efficient belief propagation for early vision. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR 2004). **Proceedings...** vol. 1, 2004, pp. 261-268.

Freeman, W. T.; Jones, T. R.; Paztor, E. C. Example-based super-resolution. Mitsubishi Electric Research Laboratories – MERL., TR-2001-30.

Friedman, J. H.; Bentley, J.L.; Finkel, R.A. An algorithm for finding best matches in logarithmic expected time. **ACM Transactions on Mathematical Software**, v. 3, Issue 3, p. 209-226, 1977.

Goltz, G. A. M.; Shiguemori, E. H. **Aplicação do Algoritmo SIFT em Imagens de Navegação Autônoma**. Workshop Anual de Pesquisa e Desenvolvimento do IEAv, 2008.

Gross, D.; Ur, A. Improved Resolution from Subpixel Shifted Pictures. CVGIP - Graphical Models and Image Processing; vol. 54, n. 2, Março, 1992, pp. 181-186.

Hardie, R.; Barnard, K.; Armstrong, E. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. **IEEE Transactions on Image Processing**, 6(12):1621-1633, Dec. 1997.

Hartley, R.; Zisserman, A. **Multiple View Geometry in Computer Vision**. University Press, Cambridge, 2 ed., 2003.

Hathaway, D.; Meyer P. NASA: Video Enhancer For Law Enforcement & More. Advanced Imaging: Solutions For The Electronic Imaging Professional; Maio, 1999.

Heckel, D. Y. **Método Híbrido de correspondência para pares estereoscópicos de imagens aéreas e de satélite de alta definição.** 2008. 115 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Hiraga, A. K.; Silva, F. A. **Algoritmo Para a Construção de Panoramas Contemplando a Minimização das Distorções Causadas nas Junções das Imagens.** 20º Simpósio Internacional de Iniciação Científica da USP, 2011.

Hossain, T.; Teng, S. W.; Lu, G.; Lackmann, M. An Enhancement to SIFT-Based Techniques for Image Registration. In: INTERNATIONAL CONFERENCE ON DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS (DICTA), Sydney, NSW, 2010. pp. 166-171.

Hsu, Chih-Chung; Lin, Chia-Wen; Image Super-Resolution via Feature-Based Affine Transform. In IEEE 13th INTERNATIONAL WORKSHOP ON MULTIMEDIA SIGNAL PROCESSING (MMSP 2011), Hangzhou, China, October 17-19, 2011.

Ihara, A.; Fujiyoshi, H.; Takagi, M.; Kumon, H. Tamatsu, Y. Improved Matching Accuracy in Traffic Sign Recognition by Using Different Feature Subspaces. **Machine Vision Applications (MVA2009)**, 3-26, p. 130-133, 2009.

Irani, M., Anandan, P., Hsu, S. **Mosaic Based Representations of Video Sequences and Their Applications.** *Proceeding of Fifth International Conference on Computer Vision (ICCV'95)*, 1995.

Joyeux, L., et al. Reconstruction of degraded image sequences. Application to film restoration. *Image and Vision Computing*. n. 19, France, 2001, pp. 503-516.

Kang, M.J, et al. The Digital Image Acquisition of High-Resolution by Enhancement the Multiple Images. *ASPRS-Proceedings - Launching the Geospatial Information Age*, Washington, 2000, pdf226.pdf.

Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In IEEE CONF. ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), v. 2, 2004. p. 506-513.

Kennedy, J. A.; Israel, O.; Frenkel, A.; Bar-Shalom, R.; Azhari, H. Super-resolution in PET imaging. **IEEE Transactions on Medical Imaging**. 25(2): pp. 137-147, 2006.

Kim, S.P.; Bose, N.K.; Valenzuela, H.M. Recursive reconstruction of high resolution image from noisy undersampled multiframes. **IEEETrans. on Accoustics, Speech and Signal Processing**, v. 18, n. 6, pp. 1013-1027, June 1990.

Kim, S.P.; Su, W.Y. recursive high-resolution reconstruction of blurred multiframe images. **IEEE Trans. On Image Processing**, v. 2, pp. 534-539, Oct. 1993.

Koenderink, J. J. The structure of images. **Biological Cybernetics**, Springer-Verlag, v. 50, 1984, p. 363-396.

Li, K.; Zhou, S. A Fast SIFT Feature Matching Algorithm for Image Registration. In: **INTERNATIONAL CONFERENCE ON MULTIMEDIA AND SIGNAL PROCESSING (CMSP)**, Guilin, Guangxi, 2011, v. 1, pp. 89-93.

Li, L.; Geng; N. **Algorithm for Sequence Image Automatic Mosaic based on SIFT Feature**. **IEEE – International Conference on Information Engineering**. pp. 203 -206. 2011.

Li, K.; Zou, S.. **A Fast SIFT Feature Matching Algorithm for Image Registration**. **IEEE – Computer Science – International Conference on Multimedia and Signal Processing (CMSP 2011)**. pp. 89-93. 2011.

Lindeberg, T. Scale-space theory: A basic tool for analyzing structures at different scales. **Journal of Applied Statistics**, 21(2): 1994, pp. 224-270.

Lillesand, T. M.; Kiefer, R. W. *Remote Sensing and Image Interpretation*. 4th ed. Ed John Wiley and Sons, 724. 724p. ISBN 0-471-25515-7. 2000.

Liu, J.; Chen, Z.; Guo, R. **A Mosaic Method for Aerial Image Sequence by R/C Model**. **International Conference on Computer Science and Software Engineering – IEEE**, 2008.

Lowe, D. G. Object recognition from local scale-invariant features. In: **INTERNATIONAL CONFERENCE ON COMPUTER VISION**, Corfu, Greece, 1999. p. 1150-1157.

Lowe, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n.2, p. 91-110, 2004.

Machado, W. R. S.; Louro, A. H. F.; Gonzaga, A.; Boaventura, M. Estudo e proposta de adaptação do algoritmo SIFT em relação ao problema de iluminação em imagens. In: **IV**

WORKSHOP DE VISÃO COMPUTACIONAL (WVC 2008). **Anais...** Bauru: Canal 6 Projetos Editoriais, v. 1, 2008.

Meitav, N.; Ribak, E. N. Improving retinal image resolution with iterative weighted shift-and-add. Vol. 28, No. 7, July 2011. IEEE, Optical Society of America. p.1395-1402. 2011

Molina, R.; Mateos, J.; Katsaggelos, A. K.; Zurita-Milla, R. A new super resolution Bayesian method for pansharpening Landsat ETM+ imagery. In 9th International SYMPOSIUM ON PHYSICAL MEASUREMENTS AND SIGNATURES IN REMOTE SENSING (ISPMSRS), edited by S. Liang, J. Liu, X. Li, R. Liu and M. Schaepman, 280-283, Beijing (China). 2005.

Mikolajczyk, K.; Schmid, C. A Performance Evaluation of Local Descriptors. **IEEE Trans. on Pattern Analysis and Machine Intelligence**. v. 27, Issue 10, p. 1615-1630, 2005.

Mikolajczyk, K.; Leibe, B.; Schiele, B. Local Features for Object Class Recognition. In: 10th IEEE INT. CONF. COMPUT. VISION (ICCV'05). **Proceedings...** 17-21 October, 2005. p. 1792-1799.

Muad, A.M.; Foody, G. M.; Super-Resolution Analysis For Accurate Mapping Of Land Cover And Land Cover Pattern. IEEE International Conference, IGARSS 2010. pp. 502-505. 2010.

Nasir, H.; Stankovic, V.; Marshall, S.; Image Registration For Super Resolution Using Scale Invariant Feature Transform, Belief Propagation And Random Sampling Consensus. 18th European Signal Processing Conference (EUSIPCO 2010). 2010. p. 299-303.

Nasir, H.; Stankovic, V.; Marshall, S.; Singular value decomposition based fusion for super-resolution image reconstruction. IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2011, pp. 393-398.

Nemra, A.; Aouf, N.; Robust Invariant Automatic Image Mosaicing and Super Resolution for UAV Mapping. IEEE, 2009.

Nguyen, N., Milanfar, P. An efficient wavelet-based algorithm for image superresolution. In Proc. Int. Conf. Image Processing, 2: 351-354. 2000.

Papa, J. P., Mascarenhas, N. D. A., Fonseca, L. M. G., Bensebaa, K. Convex restriction sets for CBERS-2 satellite image restoration. International Journal of Remote Sensing, 29(2): 443-458. 2008.

Peleg, S.; Keren, D.; Schweitzer, L. Improving Image Resolution Using Subpixel Offsets (Motion). **Pattern Recognition Letters**, 1987, pp. 223-226.

Peleg, S.; Irani, M. Improving Resolution by Image Registration. **Graphical Models and Image Processing**. v. 53, n. 3, 1991, p. 231-239.

Purkait, P.; Bhabatosh, C.. Super Resolution Image Reconstruction Through Bregman Iteration Using Morphologic Regularization. **IEEE TRANSACTIONS ON IMAGE PROCESSING**. v. 21, n. 9, 2012, pp. 4029-4039.

Richards, J. A. Remote Sensing Digital Image Analysis - In Introduction. Berlin-Heidelberg. Springer-Verlag, 5^a edition, 2013. 494p.

Rubert, C. **Super-resolução aplicada a imagens de sensores remotos**. Exame de qualificação em computação aplicada. Instituto Nacional de Pesquisas Espaciais – INPE. São José dos Campos - SP, 2004.

Rubert, C.; Fonseca, L.; Velho, L., **Super-Resolução de Imagens de Sensores Remotos Usando a Transformada de Cor YUV para Estimação das Altas Frequências**. In IV Workshop dos Cursos de Computação Aplicada, Instituto Nacional de Pesquisas Espaciais. São José dos Campos - SP, 2004.

Salari, E.; Bao, G., **Super-resolution using an enhanced Papoulis–Gerchberg algorithm**. The Institution of Engineering and Technology – IET Image Process., vol. 6, Iss. 7, pp. 959–965, 2012.

Sezer, O. G. **Super-resolution Techniques for Face Recognition from Video**. Master of Science Thesis, Sabanci University, Istanbul, Turkiye, 62p. 2003.

Sezer, O. G.; Altunbasak, Y. Super-resolution Reconstruction of Multichannel Images. **SUPER-RESOLUTION IMAGING**. Livro Editado por Peymar Milafar. pp. 355-381. 2011.

Shrestha, M. S.; Arai, I. Signal Processing of Ground Penetrating Radas Using Spectral Estimation Techniques to Estimate the Position of Buried Target. **EURASIP Journal on Applied Signal Processing**. 12: 1198-1209. Tokyo. 2003.

Silpa-Anan, C.; Hartley, R. Optimised kd-trees for fast image descriptor matching. In: **CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR 2008)**, p. 1-8. 2008.

Silva, F. A.; Artero, A. O.; Paiva, M. S. V.; Barbosa, R. L. “Reconhecimento de Caracteres Baseado em Regras de Transições entre Pixels Vizinhos”, *Avanços em Visão Computacional*. Omnipax Editora Ltda, Curitiba - PR, 2011.

Silva, F. A. **Georreferenciamento automático de placas de sinalização com imagens obtidas com um sistema móvel de mapeamento**. 2012. 211 f. Tese (Doutorado - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Processamento de Sinais e Instrumentação), Escola de Engenharia de São Carlos da Universidade de São Paulo, 2012.

Silva, F. A.; Hiraga, A.K.; Artero, A.O.; Paiva, M.S.V.; Barbosa, R.L. **Um Novo Algoritmo Para A Construção de Imagens Panorâmicas Usando os Algoritmos Sift e Ransac**. Anais do III Simpósio Brasileiro de Geomática, Presidente Prudente - SP, 25-27 de julho, v.1, 2012.

Stelzer, M. 2005. **Restauração de Imagens de Microscopia Confocal Usando POCS**. Dissertação de Mestrado em Computação. Universidade Federal de São Carlos. 72p. 2005.

Stark, H. Theory of convex projections and its application to image restoration. **IEEE International Symposium on Circuits and Systems**, pp. 963-964, 1988.

Stark, H.; Oksui, P. High-resolution image recovery from image-plane arrays using convex projections. *J. Optical Society of America*, v.6, n. 11, pp 1715-1726, Nov. 1989.

Suzuki, T.; Amano, Y.; Hashizume, T. Vision Based Localization of a Small UAV for Generating a Large Mosaic Image. In: *SICE Annual Conference*, p 2960-2964, 2010.

Tang, C.; Dong, Y.; Su, X. Automatic Registration based on Improved SIFT for Medical Microscopic Sequence Images. In: *SECOND INTERNATIONAL SYMPOSIUM ON INTELLIGENT INFORMATION TECHNOLOGY APPLICATION*. IITA '08. vol. 1. 2008, pp. 580-583.

Tarallo, A. S., **Mosaico Automático de Imagens Digitais Aéreas Agrícolas Através da Transformada SIFT e Aplicação de Processamento Paralelo**. 2012. 210 f. Qualificação (Doutorado - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Processamento de Sinais e Instrumentação), Escola de Engenharia de São Carlos da Universidade de São Paulo, 2012.

Telles Jr, M. A. B. G.; Rosa, A. N. C. S. Super-resolução de imagens CBERS 2. In: XIII SIMPÓSIO DE SENSORIAMENTO REMOTO. **Anais...** Florianópolis, Brasil, INPE, pp 1197-1204, Abr. 2007.

Telles Jr, M. A. B. G.; **Super-resolução de imagens de sensoriamento remoto**. 2008. 123f. Tese (Doutorado em Geologia) – Brasília, Brasil, Universidade de Brasília (UnB). 2008.

Thévenaz, P.; Blu, T.; Unser, M., Interpolation revisited. *IEEE Transactions on Medical Imaging*, v. 19, n. 7, p. 739–758, jul. 2000.

Tsai, R. Y.; Huang, T. S. Multiframe image restoration and registration. **Advances in Computer Vision and Image Processing**. pp. 317-339, JAI Press Inc., 1984.

Viola, P.; Jones, M.; Robust Real-time Object Detection. In: SECOND INTERNATIONAL WORKSHOP STATISTICAL AND COMPUTATIONAL THEORIES OF VISION, Vancouver, Canada, 2001.

Wang, H. S. Z.; Bovik A., Simoncelli, E. Image Quality Assessment: From Error Visibility to Structural Similarity. **IEEE Transactions on Image Processing**, 13(4):600–612, Abr. 2004.

Wang, Y.; Shen, J.; Liao, W; Zhou, L. **Automatic Fundus Images Mosaic Based on SIFT Feature**. International Congress on Image and Signal Processing, 2010.

Wang Z.; Bovik, A. C. A Universal Image Quality Index. *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, 2002.

Willett, R.; Jermyn, I.; Nowak, R.; Zerubia, J. Wavelet-Based Superresolution in Astronomy. In: XIII ASTRONOMICAL DATA ANALYSIS SOFTWARE AND SYSTEMS. **Proceedings...** v.314, 2004.

Wiman, H. Improvement of Digital Resolution by Oversampling. ISPRS Commision II, Washington, 1992.

Witkin, A. P. Scale-space filtering. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, KARLSRUHE, GERMANY, 1983. p. 1019-1022.

Wittman, T. Mathematical Techniques for Image Interpolation. Department of Mathematics, Universidade do Misshott. Department of Mathematics, University of Minnesota. 33p. 2005.

Xing, C.; Huang, J. **An improved mosaic method based on SIFT algorithm for UAV sequence images**. IEEE International Conference on Computer Design and Applications (ICCD), p. 414-417, 2010.

Yang, F; Lu, Y. M.; Sbaiz, L.; Vetterli, M.; Bits From Photons: Oversampled Image Acquisition Using Binary Poisson Statistics. IEEE Transactions on Image Processing. Vol. 21, n.4, April 2012. 1421-1436.

Yaroslavsky, L.; Fast Signal Sinc-Interpolation and its Applications in Signal and Image Processing. Image Processing: Algorithms and Systems, **Proceedings of SPIE**, v. 4667, 2002.

Zhan-Long, Y.; Bao-Long, G. **Image Mosaic Based On Sift**. International Conference on Intelligent Information Hiding and Multimedia Signal Processing - IEEE, 2008.

Zhang, Q; Zhang, Z.; Zeng, D. A fast Automatic and Robust Image Registration Algorithm. In: INTERNATIONAL CONFERENCE ON VIRTUAL REALITY AND VISUALIZATION (ICVRV), Beijing, pp. 307-310, 2011.