



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE INFORMÁTICA (INF)
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CRISTIANE BASTOS ROCHA FERREIRA

Análise Multirresolução de Imagens Gigapixel para Detecção de Faces e Pedestres

Goiânia
2023



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES
ELETRÔNICAS DE TESES
E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFUG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFUG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Cristiane Bastos Rocha Ferreira

3. Título do trabalho

Análise Multirresolução de Imagens Gigapixel para Detecção de Faces e Pedestres

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Fabrizio Alphonso Alves De Melo Nunes Soares, Professor do Magistério Superior**, em 25/10/2023, às 13:17, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Bastos Rocha Ferreira, Discente**, em 25/10/2023, às 13:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4148441** e o código CRC **5D6E9B33**.

CRISTIANE BASTOS ROCHA FERREIRA

Análise Multirresolução de Imagens Gigapixel para Detecção de Faces e Pedestres

Tese apresentada ao Programa de Pós-Graduação em
Ciência da Computação do Instituto de Informática
(INF), da Universidade Federal de Goiás (UFG), como
requisito para obtenção do título de Doutora em Ciência
da Computação.

Área de concentração: Ciência da Computação.

Linha de Pesquisa: Sistemas Inteligentes e Aplicações.

Orientador: Professor Doutor Fabrizzio Alphonsus Alves
de Melo Nunes Soares

Coorientador: Professor Doutor Hélio Pedrini

Goiânia
2023

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Ferreira, Cristiane Bastos Rocha

Análise Multirresolução de Imagens Gigapixel para Detecção de Faces e Pedestres [manuscrito] / Cristiane Bastos Rocha Ferreira. - 2023.

CII, 102 f.: il.

Orientador: Prof. Fabrizzio Alphonsus Alves de Melo Nunes Soares; co-orientador Hélio Pedrini.

Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Informática (INF), Programa de Pós-Graduação em Ciência da Computação, Cidade de Goiás, 2023.

Bibliografia. Anexos.

Inclui gráfico, tabelas, lista de figuras, lista de tabelas.

1. Imagens Gigapixel. 2. Detecção de faces e pedestres. 3. Análise multirresolução. I. Soares, Fabrizzio Alphonsus Alves de Melo Nunes, orient. II. Título.

CDU 004



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA
ATA DE DEFESA DE TESE

Ata Nº **24/2023** da sessão de Defesa de Tese de **Cristiane Bastos Rocha Ferreira** que confere o título de Doutora em **Ciência da Computação**, na área de concentração em **Ciência da Computação**.

Aos vinte e sete dias do mês de setembro de dois mil e vinte e três, a partir das nove horas, na 150 do Instituto de Informática, realizou-se a sessão pública de Defesa de Tese intitulada **“Estudo de Adaptações de Algoritmos Convencionais no Processo de Varredura, Análise Multirresolução e Detecção de Objetos de Interesse em Imagens Gigapixel”**. Os trabalhos foram instalados pelo Orientador, Professor Doutor Fabrizzio Alphonsus Alves de Melo Nunes Soares (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professor Doutor Hélio Pedrini (UNICAMP), coorientador; Professor Doutor Edmilson Batista dos Santos (DCOMP/UFSJ), membro titular externo; Professor Doutor Dívio Leandro Borges (CIC/UnB), membro titular externo; Professora Doutora Deborah Silva Alves Fernandes (INF/UFG), membro titular interno; Professor Doutor Gustavo Teodoro Laureano (INF/UFG), suplente interno. Os membros externos, profs. Hélio Pedrini, Edmilson Batista dos Santos e Dívio Leandro Borges, participaram à distância através de videoconferência. Durante a arguição os **membros da banca fizeram sugestão de alteração do título do trabalho**, conforme especificado abaixo. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido a candidata **aprovada** pelos seus membros. Proclamados os resultados pelo Professor Doutor Fabrizzio Alphonsus Alves de Melo Nunes Soares, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos vinte e sete dias do mês de setembro de dois mil e vinte e três.

TÍTULO SUGERIDO PELA BANCA

Análise Multirresolução de Imagens Gigapixel para Detecção de Faces e Pedestres



Documento assinado eletronicamente por **Edmilson Batista dos Santos, Usuário Externo**, em 27/09/2023, às 15:55, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Helio Pedrini, Usuário Externo**, em 27/09/2023, às 16:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Bastos Rocha Ferreira, Discente**, em 27/09/2023, às 16:10, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fabrizzio Alphonsus Alves De Melo Nunes Soares, Professor do Magistério Superior**, em 27/09/2023, às 16:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gustavo Teodoro Laureano, Professor do Magistério Superior**, em 27/09/2023, às 16:17, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Deborah Silva Alves Fernandes, Professora do Magistério Superior**, em 27/09/2023, às 16:20, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Dívio Leandro Borges, Usuário Externo**, em 27/09/2023, às 16:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4078181** e o código CRC **060523C4**.

Referência: Processo nº 23070.046431/2023-99

SEI nº 4078181

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Cristiane Bastos Rocha Ferreira

Possui Graduação em Ciência da Computação pela Universidade Federal de Goiás (1999), Mestrado em Informática Aplicada pela Pontifícia Universidade Católica do Paraná (2002) e atualmente está cursando Doutorado em Ciência da Computação na Universidade Federal de Goiás. É Professora Adjunto IV do Instituto de Informática da Universidade Federal de Goiás. Atua na área de Ciência da Computação, com experiência em Processamento Digital de Imagens, Visão Computacional e Reconhecimento de Padrões.

Este trabalho é dedicado a Deus, à minha família e aos meus amigos.

Agradecimentos

A realização deste trabalho só foi possível, primeiramente, em virtude da permissão concedida pelo Grande Arquiteto do Universo, nosso Pai Maior, que nos capacita para superar todos os obstáculos da vida, e depois graças à colaboração de várias pessoas. Manifesto meus agradecimentos e minha eterna gratidão a todas elas e, de maneira particular:

- ao meu esposo William, pela ajuda, amor, companheirismo, paciência, empatia e compreensão dispensados em todos esses anos de Doutorado sendo, sempre, meu porto seguro. Agradeço por estar ao meu lado, me incentivando e ajudando em todos os sentidos;
- aos meus queridos e amados filhos, Beatriz e Gabriel, pelo amor, carinho, palavras e atitudes de incentivo. Agradeço por entenderem a minha ausência em vários momentos e, também, compreenderem que esse era um sonho a ser alcançado. Foi preciso muita resiliência, renúncia e grande esforço por parte de todos para atingirmos os objetivos traçados;
- aos meus pais, Julieta e Antônio, pelos inestimáveis valores e ensinamentos a mim transmitidos. São meus maiores exemplos de perseverança. Agradeço, também, ao meu irmão Zander por ser exemplo de virtudes como o compromisso com o trabalho, honestidade, estudo e persistência. Agradeço por me ajudar quando precisei, principalmente com relação à minha saúde. Agradecimentos também à minha cunhada Cristyane e sobrinhos por seu carinho e amor;
- ao amigo, irmão, Professor e Orientador Dr. Fabrízio Soares, por sua paciência infindável, amizade, empenho e compreensão. Agradeço por, em inúmeros momentos, ter me ensinado a não desistir e ser meu ombro amigo de maneira incondicional. Agradeço pelos conselhos, palavras e atitudes de motivação, momentos de descontração e conversas, por me mostrar os caminhos possíveis e trilhá-los junto comigo, respeitando minhas limitações e as inúmeras dificuldades ao longo dessa jornada. Agradeço por ter acreditado sempre que eu conseguiria concluir este trabalho;
- ao Professor Dr. Hélio Pedrini, que atuou como Coorientador deste trabalho, realizando contribuições relevantes e sendo sempre muito solícito em todas as in-

terações que foram estabelecidas. Obrigada por transmitir calma, tranquilidade e serenidade sempre;

- a todos colegas de pós-graduação, em especial ao integrantes do PixelLab, grupo coordenado pelo Professor Dr. Fabrizzio, que ajudou muito em vários momentos de estudos, discussões e confraternizações, nos quais havia o crescimento tanto pessoal quanto profissional de todos os envolvidos;
- aos Diretores, Vice-diretores e Coordenadores do Instituto de Informática (INF) da UFG que ocuparam esses cargos durante o período de desenvolvimento deste trabalho de doutorado, bem como aos colegas professores e técnicos administrativos do referido Instituto, por me apoiarem e me incentivarem em minhas atividades como docente do INF, buscando entender todas as dificuldades e limitações envolvidas nesta etapa;
- a todos os meus amigos e às pessoas que contribuíram direta ou indiretamente para o meu crescimento pessoal e profissional.

“O cientista não é o homem que fornece as verdadeiras respostas;
é quem faz as verdadeiras perguntas.”

(Claude Lévi-Strauss)

Resumo

FERREIRA, Cristiane Bastos Rocha. **Análise Multirresolução de Imagens Gigapixel para Detecção de Faces e Pedestres**. Goiânia, 2023. 102p. Tese de Doutorado. Instituto de Informática (INF), Universidade Federal de Goiás (UFG).

Imagens Gigapixel, também conhecidas como giga imagens, podem ser formadas a partir da junção de uma sequência de imagens individuais obtidas através de um processo de varredura de cena. Tais imagens podem ser entendidas como a construção de um mosaico baseado em um grande número de imagens digitais de alta resolução. Usando uma imagem gigapixel é possível observar detalhes mínimos que se encontram muito distantes do observador, proporcionando o desenvolvimento de pesquisas em diversas áreas como detecção de pedestres, vigilância, segurança, dentre outras. Como essa categoria de imagens possui um alto volume de dados capturados em sequência, sua geração está associada a vários problemas ocasionados no processo de produção e análise das mesmas, e assim, pode-se observar que utilizar algoritmos convencionais projetados para imagens não-gigapixel de uma maneira direta pode tornar-se inviável nesse contexto. Dessa forma, este trabalho propõe um método de varredura, manipulação e análise multirresolução de imagens Gigapixel para aplicações de identificação de pedestres e faces com o uso de algoritmos tradicionais. Tal abordagem é analisada utilizando-se tanto imagens Gigapixel com baixa quanto com alta densidade de pessoas e faces, apresentando resultados promissores.

Palavras-chave

Imagens Gigapixel. Detecção de faces e pedestres. Análise multirresolução.

Abstract

FERREIRA, Cristiane Bastos Rocha. **Multiresolution Analysis of Gigapixel Images for Face and Pedestrian Detection**. Goiânia, 2023. 102p. PhD. Thesis. Instituto de Informática (INF), Universidade Federal de Goiás (UFG).

Gigapixel images, also known as gigaimages, can be formed by merging a sequence of individual images obtained from a scene scanning process. Such images can be understood as a mosaic construction based on a large number of high resolution digital images. A gigapixel image provides a powerful way to observe minimal details that are very far from the observer, allowing the development of research in many areas such as pedestrian detection, surveillance, security, and so forth. As this image category has a high volume of data captured in a sequential way, its generation is associated with many problems caused by the process of generating and analyzing them, thus, applying conventional algorithms designed for non-gigapixel images in a direct way can become unfeasible in this context. Thus, this work proposes a method for scanning, manipulating and analyzing multiresolution Gigapixel images for pedestrian and face identification applications using traditional algorithms. This approach is analyzed using both Gigapixel images with low and high density of people and faces, presenting promising results.

Keywords

Gigapixel images. Face e pedestrian detection. Multiresolution analysis.

Conteúdo

Lista de Figuras	14
Lista de Tabelas	16
1 Introdução	17
1.1 Introdução	17
1.2 Descrição do Problema	19
1.3 Justificativa	21
1.4 Hipótese	22
1.5 Objetivos e Contribuições	23
1.6 Estrutura do Texto	23
2 Abordagens e Técnicas de Detecção de Objetos de Interesse em Imagens Convencionais e Gigapixel	25
2.1 Introdução	25
2.2 Imagens Gigapixel	25
2.3 Detecção de Pedestres e de Faces	28
2.4 Detecção de Pedestres e Faces em Imagens de Alta Resolução	31
2.5 Considerações Finais do Capítulo	33
3 Geração de Imagens Gigapixel por Processo de Varredura de Cena e Algumas Técnicas de Visualização de Imagens Panorâmicas	34
3.1 Introdução	34
3.2 Varredura de Cena para Captura de Imagens	38
3.3 Alinhamento e Combinação de Imagens	39
3.4 Visualização de Imagens Panorâmicas	42
3.5 Exemplo de Giga Imagem Gerada a Partir de um Processo de Varredura de Cena	43
3.6 Considerações Finais do Capítulo	48
4 Método de Análise de Imagens Gigapixel por Varredura	49
4.1 Introdução	49
4.2 Imagem Gigapixel Utilizada	50
4.3 Método Implementado	52
4.3.1 Etapa 1: Detecção de Pedestres	52
4.3.2 Etapa 2: Junção dos <i>Bounding Boxes</i>	55
4.3.3 Etapa 3: Detecção de Faces	55
4.3.4 Etapa 4: Análise Multirresolução	56
4.4 Resultados Alcançados	58

4.5	Considerações Finais do Capítulo	62
5	Abordagem de Varredura em Imagens Gigapixel com Agrupamento Denso de Pessoas	64
5.1	Introdução	64
5.2	Imagens Gigapixel de Multidões	65
5.3	Descrição da Base de Imagens	66
5.3.1	Base de Dados da Imagem Gigapixel	67
5.3.2	Base de Imagens Construída Manualmente	69
5.4	Método Proposto	70
5.4.1	Etapa 1: Detecção de Faces na Imagem Gigapixel	71
5.4.2	Etapa 2: Emparelhamento Entre a Base de Imagens Construída Manualmente e as Faces da Imagem Gigapixel	72
5.5	Resultados Alcançados	74
5.6	Considerações Finais do Capítulo	78
6	Conclusões, Trabalhos Futuros e Contribuições	79
6.1	Conclusões	79
6.2	Trabalhos Futuros	80
6.3	Contribuições, Publicações Derivadas desta Tese e Colaborações de Pesquisa	81
6.3.1	Trabalho Publicado em Periódico Internacional	81
6.3.2	Trabalhos Publicados em Conferências ⁴	81
6.3.3	Trabalhos Publicados como cooperação, mas não relacionados com esta tese	82
A	Termos de Direitos de Cópia de Publicações Utilizadas Nesta Tese	84
A.1	Artigo: Gigapixel Computational Imaging (Cossairt & Miau, 2011)	85
A.2	Artigo: Multiresolution Analysis on Searching for People in Gigapixel Images (Ferreira et al., 2018)	86
A.3	Artigo: A Study of Dimensionality Reduction Impact on an Approach to People Detection in Gigapixel Images (Ferreira et al., 2020a)	87
A.4	Artigo: Where’s Wally: A Gigapixel Image Study for Face Recognition in Crowds (Ferreira et al., 2020b)	88
	Bibliografia	95

Lista de Figuras

1.1	Mosaico de imagem adaptada de Schott (2009)	18
1.2	Imagem completa adaptada de Schott (2009)	18
1.3	Exemplo de imagem <i>Gigapixel</i> (da esquerda para a direita: Resistor, nota de dolar, código de barras 3D, impressão digital) (Cossairt & Miao, 2011). [©2011 IEEE]	20
1.4	Exemplo de equipamento Gigapan com câmera acoplada. Imagem extraída de Sargent et al. (2010)	22
3.1	Imagens do robô <i>Opportunity</i> , extraídas de (NASA et al., 2019).	37
3.2	Sistemas de captura de imagens GigaPan. Imagem extraída de Sargent et al. (2010)	38
3.3	Imagem GP1 como protótipo para as outras imagens Gigapixels (GP2 a GP8) no lapso de tempo. Figura adaptada do trabalho de Sargent et al. (2010)	40
3.4	Imagens GP1, GP3, GP5 e GP7 como quadro-chaves, selecionados para serem protótipos e com o objetivo de compensar os problemas de características fixas no lapso de tempo. Figura adaptada de Sargent et al. (2010)	41
3.5	Imagens do primeiro dia alinhadas em cadeia e usadas como protótipos para dias sucessivos. Figura adaptada de Sargent et al. (2010)	42
3.6	Imagem Gigapixel, com destaque para um jogador através de uma operação de <i>zoom in</i> , adaptada de Val (2013)	43
3.7	Exemplo de duplicidade de uma pessoa na mesma cena, extraído de Val (2013)	44
3.8	Exemplo de parte do corpo de uma pessoa aparecendo três vezes na mesma cena, extraído de Val (2013)	45
3.9	Exemplo de partes do corpo de uma pessoa aparecendo em localizações distintas em uma mesma cena, extraído de Val (2013)	46
3.10	Exemplo de desalinhamento no processo de combinação de imagens, com destaque para os corpos das pessoas, extraído de Val (2013)	46
3.11	Exemplo de desalinhamento no processo de combinação de imagens, com destaque para as linhas do campo de futebol, extraído de Val (2013)	46
3.12	Exemplo de perda de informação do rosto da pessoa na imagem, extraído de Val (2013)	47
3.13	Exemplo de perda de informação do tronco dos corpos das pessoas na imagem, extraído de Val (2013)	47

3.14	Exemplo de desalinhamento no processo de combinação e também perda da informação na imagem, extraído de Val (2013).	47
4.1	Imagem Gigapixel, com destaque para uma pessoa através de uma operação de <i>zoom in</i>	51
4.2	Caminhantes presentes na giga imagem avaliada.	53
4.3	Abordagem proposta para manipulação de imagem Gigapixel.	54
4.4	Exemplos de falsos positivos obtidos na análise da giga imagem avaliada.	57
4.5	Abordagem de casamento dos <i>bounding boxes</i> de face e corpo	58
4.6	Resultados da amostra depois do passo de análise multirresolução.	58
4.7	Falsos positivos <i>versus</i> taxa de perda para diferentes níveis de resolução com o ACF treinado com as bases do Inria e Caltech.	61
5.1	Onde está o Wally (<i>Where's Wally?</i>) ¹ - Conceito do estudo proposto (Adaptado de CNN).	66
5.2	Amostras de imagem Gigapixel para construir o conjunto de dados de faces.	68
5.3	Amostras da base de imagens Gigapixel: (a) de costas; (b) até (f) oclusão por acessório; (g) até (i) oclusão por outras faces e (j) distorção do processo de combinação de imagens.	69
5.4	Exemplos das faces do Obama na base de dados construída manualmente.	70
5.5	Passos da detecção de faces na imagem Gigapixel.	73
5.6	Emparelhamento entre a base de dados construída manualmente e as imagens de face da imagem Gigapixel.	75
5.7	<i>Ranking</i> de pessoas encontradas nos experimentos.	76
5.8	Histograma de <i>ranking</i> mínimo de poses de pessoas.	77
5.9	Amostras de faces em posições laterais e oclusas reconhecidas pela abordagem proposta.	77

Lista de Tabelas

3.1	Características e parâmetros de configuração para geração da giga imagem Jogo Final Brasil x Espanha no Maracanã em 2013.	44
4.1	Características e parâmetros de configuração para geração da giga imagem <i>Count the Cacti!!!</i>	51
4.2	Resultados alcançados com o uso do ACF detector treinado com as bases de dados do Inria e Caltech, considerando a imagem Gigapixel apresentada na Figura 4.1. Algoritmos de redução de resolução utilizados: Bicúbico (BC), Bilinear (BL), Vizinho Mais Próximo (VP) e Pirâmide Laplaciana (PL).	59
	(a) Número de <i>Bounding Boxes</i> Encontrados.	59
	(b) Tempo para encontrar os <i>bounding boxes</i>	59
	(c) Tempo para redução da resolução.	59
	(d) Tempo total de processamento.	59
4.3	Nível mais baixo de resolução para o casamento da face e da pessoa. .	62
5.1	Estatísticas para a base de imagens Gigapixel.	68
5.2	Resultado pelo <i>ranking</i> mínimo.	76

Introdução

Sumário

1.1	Introdução	17
1.2	Descrição do Problema	19
1.3	Justificativa	21
1.4	Hipótese	22
1.5	Objetivos e Contribuições	23
1.6	Estrutura do Texto	23

1.1 Introdução

Uma imagem digital pode ser entendida como uma representação de uma cena através de um conjunto de *pixels* que definem elementos discretos dispostos em uma forma bidimensional, o que permite seu armazenamento, transferência, impressão, reprodução e seu processamento através de técnicas e métodos de processamento de imagens. Em geral, imagens digitais podem ser encontradas em diversos tamanhos, tipicamente entre 0,3 e 15 Megapixels.

É importante destacar que, um monitor de computador ou TV, com resolução *Full-HD* (1920×1080 pixels), proporciona em torno de 2,07 megapixels. Assim, imagens produzidas pelas câmeras digitais possuem resolução maior do que a resolução dos monitores/TVs comuns. Mesmo ainda monitores/TVs com resolução de *4K* proporcionam apenas $\approx 8,3$ megapixels, o que ainda é muito abaixo de muitas câmeras comuns. Entretanto, essas imagens podem ser facilmente exibidas na área da tela dos monitores convencionais, através do uso de algoritmos que promovam a redução de suas dimensões, usando um simples processo de descarte alternado de linhas e colunas. Embora isso possa caracterizar-se como um processo útil para a visualização humana, este pode não representar uma estratégia adequada para um

sistema de processamento, visto que dados relevantes podem ser perdidos, causando impacto nos resultados obtidos pelos algoritmos aplicados sobre a imagem.

Uma outra categoria de imagens que surgiu na última década é a denominada Imagens Gigapixel, também conhecidas como giga imagens, e que são imagens que possuem uma quantidade muito grande de dados, visto que possuem tamanhos que podem, em geral, ter sido produzidos em intervalos entre 0,3 a 300 gigapixels. Embora já existam sensores com resolução bastante alta, estes ainda se encontram em valores de 200 megapixels, e assim, para a produção de imagens de Gigapixel, estas são formadas por múltiplas capturas, ou seja, centenas ou até milhares de imagens de captura simples. A Figura 1.1 ilustra uma cena de múltiplas capturas, e que são organizadas e conectadas, formando uma enorme imagem única, resultando no que pode ser observado na Figura 1.2.

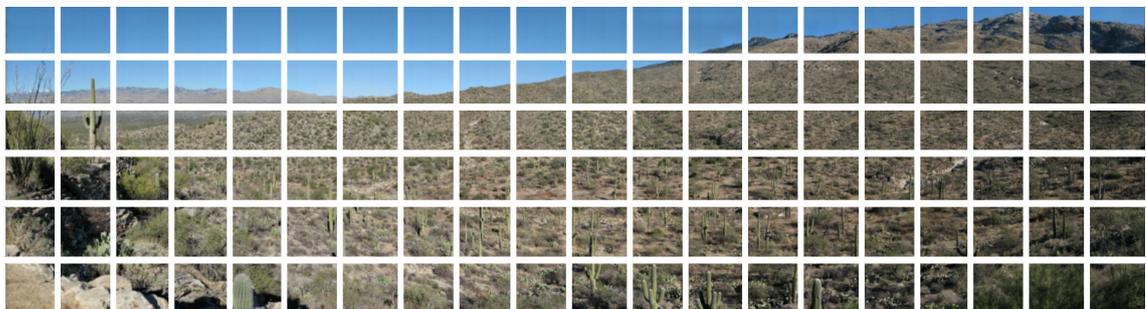


Figura 1.1: Mosaico de imagem adaptada de [Schott \(2009\)](#).

A Figura 1.2 mostra um exemplo do que foi evidenciado anteriormente. A imagem possui uma resolução de 4.067 megapixels (121.877×33.373 pixels), em que cada pixel, é composto por canais RGB de 8 bits (1 byte) cada, provendo uma imagem com aproximadamente 12,29 GB, sem compressão.



Figura 1.2: Imagem completa adaptada de [Schott \(2009\)](#).

A construção de uma imagem Gigapixel pode ser realizada usando uma câmera, que não necessariamente terá uma resolução tão alta, mas que poderá realizar a captura de várias imagens, uma ao lado da outra, em um plano ou eixo de varredura (tanto no eixo horizontal quanto vertical). Posteriormente, um mosaico é formado

através do uso de ferramentas (pacotes) computacionais de combinação de imagens. Buscando minimizar erros de perspectiva e combinação, as imagens geradas precisam possuir uma sobreposição de fronteira suficiente, além de serem capturadas com estabilidade, condições similares de iluminação, dentre outros. Assim, esse trabalho é geralmente feito por auxílio de equipamentos, uma vez que, a realização deste de forma manual pode se tornar desgastante e sujeito a problemas como, por exemplo, distorções e desalinhamentos.

1.2 Descrição do Problema

As câmeras de alta resolução podem capturar imagens com milhões de pixels, mas as imagens Gigapixel surgiram com a possibilidade de explorar detalhes mínimos de uma cena, podendo revelar elementos que são imperceptíveis ao próprio olho humano na geração das imagens.

O trabalho desenvolvido por [Cossairt & Miao \(2011\)](#) mostra uma arquitetura de geração de imagens Gigapixel que utiliza um dispositivo ótico simplificado, o qual compreende uma lente esférica compartilhada por diversos sensores planares e uma etapa de processamento de imagem após o estágio de captura da mesma. Algumas variantes da arquitetura proposta são evidenciadas ao capturar um campo de visão hemisférico contíguo, bem como outro esférico completo, utilizando duas câmeras comerciais para o estudo em questão, como o de formato de imagens grande ([Benzra, 2010](#)), arranjos de câmeras e óticas multiescalas ([Wilburn et al., 2005](#)) e óticas monocêntricas e sensores curvos ([Marks & Brady, 2010](#)).

A Figura 1.3 é mostrada no trabalho desenvolvido por [Cossairt & Miao \(2011\)](#) e apresenta um exemplo em que é possível observar quatro áreas que foram destacadas e ampliadas, evidenciando possíveis regiões de interesse para identificação e análise com relação ao contexto a ser trabalhado.

Nesse sentido e considerando o já discutido na Seção 1.1 deste capítulo, imagens Gigapixel podem ser de interesse em várias áreas do conhecimento, como, por exemplo: na pesquisa, consulta e monitoramento de desenvolvimento urbano, de hábitos de vidas de animais selvagens e de sítios arqueológicos, na astronomia com a exploração e dinâmica de corpos celestes, na área de segurança e vigilância, reconhecimento, entre outras. Os trabalhos relacionados às áreas anteriormente mencionadas serão detalhados no Capítulo 2.

É importante destacar que, algoritmos para processamento de imagens, que realizam tarefas como segmentação ou localização de objetos de interesse, necessitam muitas vezes de avaliar o conteúdo da imagem por completo, ou de

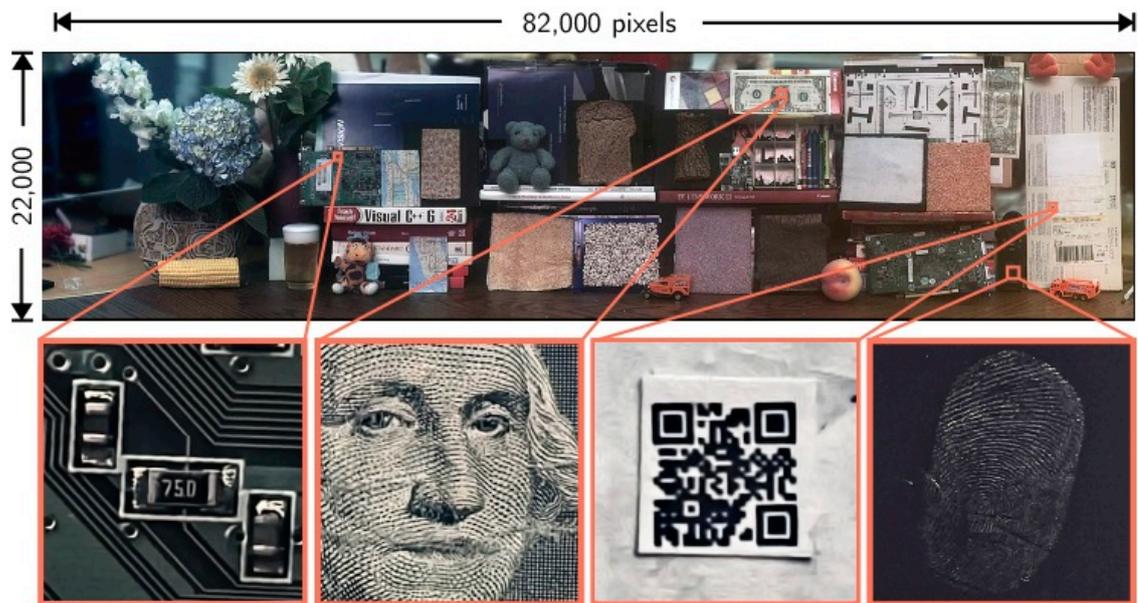


Figura 1.3: Exemplo de imagem *Gigapixel* (da esquerda para a direita: Resistor, nota de dolar, código de barras 3D, impressão digital) (Cossairt & Miao, 2011). [©2011 IEEE]

grandes agrupamentos de pixels, às vezes chamados de BLOBs¹. Contudo, aplicar algoritmos tradicionais nesse tipo de imagens para detecção de regiões de interesse, reconhecimento de pessoas, objetos, estruturas, dentre outros, é uma tarefa árdua, pois devido à sua alta resolução, essa estratégia requer um alto tempo de processamento, visto que tais algoritmos foram projetados pensando em imagens convencionais, as quais demandam um tempo mais razoável de processamento.

Além disso, o processo de combinação de imagens na geração da imagem Gigapixel pode ocasionar vários problemas que serão discutidos mais adiante no Capítulo 3, como, por exemplo, duplicidade de artefatos que se movem na cena, aparição de pedaços de um mesmo artefato em localizações distintas da giga imagem, desalinhamento nos eixos horizontal e/ou vertical de partes de um artefato no processo de combinação para gerar a imagem Gigapixel, geração de efeitos fantasma, perda de informação, entre outros.

Assim, considerando o alto custo computacional para análise de imagens, esta tese propõe um método de varredura, manipulação e análise multirresolução de imagens Gigapixel para aplicações de identificação de pedestres e faces com o uso de algoritmos tradicionais. Tal abordagem é analisada utilizando-se tanto imagens Gigapixel com baixa quanto com alta densidade de pessoas e faces.

¹BLOB – *Binary Large Objects*, do inglês, objetos grandes binários.

1.3 Justificativa

Inicialmente, nos sistemas de fotografias convencionais, a imagem era gerada em uma câmera analógica que possuía um filme fotográfico de haleto de prata. Para realizar a revelação fotográfica era feito um processo químico cujo objetivo principal era transformar a imagem registrada no filme em uma imagem visível, imprimindo-a, ao final, em um papel de haleto de prata, conforme evidenciado por [Parulski & Rabbani \(2000\)](#).

Com o passar do tempo, os sistemas de fotografias digitais foram surgindo e sendo aprimorados. As primeiras câmeras eletrônicas disponíveis comercialmente datam da década de 1980 e foram introduzidas pela Sony, conforme discutido por [Kihara et al. \(1982\)](#). Elas utilizavam um mecanismo de gravação analógica em um disquete (*floppy disk*). Entretanto, o sistema apresentou qualidade de imagem limitada e um custo elevado para os padrões da época, não se popularizando na ocasião, pois, também, se baseava na exibição de imagens estáticas no aparelho de televisão.

Já o primeiro sistema de fotografia digital voltado para os consumidores comuns foi desenvolvido pela Kodak na década de 1990 e também exibia as imagens na televisão, necessitando de um dispositivo de reprodução com alto custo para a época. Entretanto, com o armazenamento de imagens sendo realizado em discos ópticos (*Compact Disk - CD*) que podiam ser lidos por unidades de CD-ROM em computadores pessoais, esse sistema foi aceito em áreas como artes gráficas devido ao baixo custo para o armazenamento de imagens de alta qualidade.

Com o avanço da tecnologia, a indústria e os consumidores foram percebendo as vantagens dos sistemas de fotografia digital, devido a várias características, como, por exemplo, a capacidade de armazenamento, duplicação, edição e transmissão de imagens digitais. A evolução dos computadores pessoais e a diminuição do custo associado favoreceu o crescimento da fotografia digital. Além disso, o avanço da Internet ligado, também, ao aparecimento e desenvolvimento de dispositivos portáteis, tais como celulares e *tablets*, impulsionou ainda mais a geração e o compartilhamento de imagens digitais na rede para qualquer lugar do mundo de forma instantânea.

As imagens Gigapixel, por sua vez, chegaram ao mercado nos anos 2000, segundo [Nourbakhsh \(2010\)](#). O registro dessas imagens está ligado à geração de várias imagens individuais obtidas através de um processo de varredura de uma cena, sendo estas utilizadas, posteriormente, para a montagem de uma imagem Gigapixel final. Essas imagens podem ser entendidas como a montagem de um mosaico a partir de um grande número de imagens digitais de alta resolução, conforme descrito por [Powell et al. \(2010\)](#).

Através de uma imagem Gigapixel, é possível ao observador acessar detalhes que, a princípio, estão muito distantes de quem os está visualizando. Assim, com um equipamento gerador de imagens Gigapixel e, ainda, com um determinado nível de ampliação (*zoom*), ocorre a potencialização para permitir o acesso visual a detalhes mínimos presentes na imagem. As imagens Gigapixel podem ser produzidas através do uso de uma câmera acoplada a um equipamento denominado Gigapan, apresentado na Figura 1.4, o qual faz a varredura de uma cena tanto na horizontal quanto na vertical, produzindo várias imagens que serão, então, conectadas para formar uma única imagem Gigapixel.



Figura 1.4: Exemplo de equipamento Gigapan com câmera acoplada. Imagem extraída de [Sargent et al. \(2010\)](#).

O equipamento apresentado na Figura 1.4 permite gerar imagens com resoluções muito altas, 121.856×33.280 pixels, a título de exemplificação. Dessa forma, é importante salientar que, em geral, os algoritmos para visualização/deteção/reconhecimento de objetos e pessoas são desenvolvidos para imagens com baixas resoluções, como, por exemplo, 640×480 pixels. Assim, apenas uma janela pequena desta imagem pode conter o elemento de interesse e, dessa forma, os algoritmos clássicos revelam-se inviáveis de serem aplicados diretamente nestas imagens e, portanto, faz-se necessário um estudo e verificação de adaptação dos mesmos para uso em imagens Gigapixel.

1.4 Hipótese

Tendo em vista a natureza desafiadora de se trabalhar com imagens Gigapixel, devido ao alto volume de dados, uma distribuição espacial complexa e ao processo de formação da imagem total, a hipótese considerada neste trabalho é a de que usar fragmentos ou fatias de imagens Gigapixel e sua manipulação em versões em resoluções menores podem viabilizar o uso de técnicas e algoritmos tradicionais usados em imagens não-Gigapixel.

1.5 Objetivos e Contribuições

O objetivo geral deste trabalho compreende a proposta de um método de varredura, manipulação e análise multirresolução de imagens Gigapixel para aplicações de identificação de pedestres e faces com o uso de algoritmos tradicionais. Já como objetivos específicos, pretende-se:

- Propor um método de análise de imagens Gigapixel por varredura;
- Realizar análise em diferentes níveis de resolução;
- Experimentar algoritmos de detecção de pessoas e faces tanto em imagens Gigapixel com baixa presença delas, bem como em imagens que possuem agrupamento denso das mesmas.

É importante ressaltar que este trabalho leva em consideração os desafios em lidar com imagens Gigapixel, considerando as questões apresentadas nas seções anteriores, principalmente com relação ao seu alto volume de dados, bem como os problemas ocasionados na geração e análise de imagens dessa natureza ao se utilizar algoritmos projetados para imagens não-Gigapixel.

Dessa forma, o presente trabalho traz como contribuições o desenvolvimento de um método para lidar com as imagens Gigapixel no sentido de conseguir realizar um processo de varredura nas mesmas que possibilite encontrar pessoas e faces, considerando a possibilidade da existência ou não de um agrupamento denso das mesmas. Além disso, a análise em diferentes níveis de resolução constitui uma grande contribuição alcançada, visto que permite verificar se as características relevantes na detecção de pessoas e faces continuam presentes em resoluções diferentes da original, visto que outras resoluções podem conter menos dados e informações essenciais ao processo de detecção.

1.6 Estrutura do Texto

O restante deste documento encontra-se organizado da maneira a seguir. O Capítulo 2 versa sobre trabalhos que tratam de detecção de pedestres e faces, tanto em imagens convencionais quanto em imagens de alta resolução. O Capítulo 3, por sua vez, descreve a geração e visualização de imagens Gigapixel a partir de um processo de varredura, detalhando as etapas de captura, alinhamento e combinação das imagens, bem como a visualização das mesmas. O Capítulo 4 discute um método de varredura em imagens Gigapixel, abordando etapas de detecção de pedestres, combinação de retângulos envolventes (*bounding boxes*), detecção de faces e análise multirresolução. O Capítulo 5 analisa uma abordagem de varredura em imagens Gigapixel com

agrupamento denso de objetos de pessoas, passando por etapas de detecção de faces em imagens Gigapixel, bem como emparelhamento das faces para o processo de detecção. Dessa forma, os Capítulos 4 e 5 apresentam e discutem os resultados alcançados com as abordagens desenvolvidas. Por fim, o Capítulo 6 relaciona as conclusões, as propostas para trabalhos futuros e as principais contribuições deste trabalho.

Abordagens e Técnicas de Detecção de Objetos de Interesse em Imagens Convencionais e Gigapixel

Sumário

2.1	Introdução	25
2.2	Imagens Gigapixel	25
2.3	Detecção de Pedestres e de Faces	28
2.4	Detecção de Pedestres e Faces em Imagens de Alta Resolução	31
2.5	Considerações Finais do Capítulo	33

2.1 Introdução

As imagens Gigapixel possuem uma grande quantidade de informações e oportunizam o desenvolvimento de pesquisas em várias áreas, como, por exemplo, a detecção de pedestres, vigilância, segurança, entre outras. Este capítulo apresenta uma breve definição de Imagens Gigapixel e, em seguida, uma revisão de literatura, em que aborda trabalhos que tratam de imagens Gigapixel, além de detalhar o desenvolvimento de estratégias e modelos para detecção de pedestres e faces tanto em imagens convencionais quanto em imagens Gigapixel.

2.2 Imagens Gigapixel

As imagens Gigapixel são formadas a partir da captura de uma sequência de imagens que passam por um processo de junção para compor uma única imagem. O registro dessas imagens, bem como a organização das mesmas, é um processo

desafiador e que apresenta alguns problemas a serem discutidos no Capítulo 3. Vários trabalhos e pesquisas vêm sendo desenvolvidos utilizando imagens Gigapixel em diversas áreas, conforme mencionado no Capítulo 1 e alguns deles serão apresentados a seguir.

[Smith \(2010\)](#) desenvolveu um trabalho relacionado ao monitoramento do desenvolvimento do ambiente urbano, visto que os ambientes naturais existentes em alguns centros urbanos como florestas ou bosques estão sujeitos às pressões crescentes de urbanização, de destruição, invasões dos seres humanos, degradações, dentre outros. Neste trabalho, o autor discute que monitorar esses ambientes ao longo do tempo pode ser a chave para entender as mudanças ocorridas. Além disso, democratizar esse monitoramento, tornando-o público e disponível, pode levar a população humana a valorizar melhor esses espaços. Várias imagens Gigapixel são mostradas no trabalho mencionado e foram fotografadas num lapso de tempo de praticamente um ano (entre 2009 e 2010), buscando analisar os impactos causados pela ação humana nos espaços naturais, além de capturar a fenologia da flora e da fauna nativas da floresta estudada.

O trabalho desenvolvido por [Nichols et al. \(2009\)](#) mostra o monitoramento de pastagens objetivando integrar imagens panorâmicas de altíssima resolução com métodos convencionais de monitoramento, abordando três categorias de gerenciamento de recursos: áreas ribeirinhas, vida selvagem e espécies invasoras. O objetivo principal desse trabalho é apresentar as configurações dos equipamentos e parâmetros utilizados para geração das imagens Gigapixel, bem como mencionar sobre o software utilizado para a combinação das imagens geradas. Alguns problemas são discutidos como, por exemplo, a mobilidade de animais na coleta de imagens para o monitoramento da vida selvagem. Neste exemplo, como os animais se movem é necessário um cuidado maior com o intervalo de tempo entre a captura de uma imagem e outra. Alguns outros exemplos semelhantes a esses são discutidos ao longo desta tese. [Nichols et al. \(2009\)](#) mostram que o monitoramento pode ser a base para compreender e quantificar os impactos que fatores estressores como clima, gestão e desenvolvimento populacional, por exemplo, podem causar nas pastagens, além de fornecerem uma documentação para pesquisas em resoluções e nível de detalhe maiores do que as imagens tradicionais.

[Seidl & Breiteneder \(2011\)](#) apresentaram um trabalho de detecção de petróglifos em imagens Gigapixel. Petróglifo é o termo usado para caracterizar as gravuras rupestres, que são representações gravadas pelo homem em pedras ou rochas na pré-história. O trabalho mostra alguns resultados preliminares da segmentação de petróglifo através do uso de uma abordagem de classificação baseada em pixel. Primeiramente, para cada pixel é obtido um bloco da imagem de entrada

com o pixel em seu centro. Depois, são extraídas características de cor, borda e textura de cada um desses blocos e depois um SVM (*Support Vector Machine*) é treinado e, por último, a classificação de um novo dado é feito através do modelo treinado no passo anterior.

O trabalho proposto por [Gueguem et al. \(2011\)](#) apresenta uma ferramenta interativa de mineração de informações para manipulação de imagens Gigapixel com o objetivo de detectar estruturas de interesse como edifícios, estradas e tendas nas imagens, visto que isso pode ser interessante quando, por exemplo, ocorre a necessidade de avaliação de situações humanitárias como em desastres, conflitos, terremotos, dentre outros. A abordagem foi testada para a detecção de tendas de refugiados em uma imagem pancromática submétrica, que é produzida por satélites e caracterizada por usar uma única banda que combina as três bandas visíveis (vermelha, verde e azul), permitindo uma maior resolução espacial. Os autores propõem o uso de Decomposição Morfológica Diferencial para extração de características espectrais e geométricas e identificam as informações relevantes pelo princípio de *feedback* de relevância, fornecendo exemplos positivos e negativos para o sistema de aprendizagem. Por fim, é realizada uma classificação baseada em agrupamento e um processo de coleta de amostras de treinamento com o objetivo de atender a algumas restrições chaves (incremental, global e interativa) para uma ferramenta de mineração de dados com alto número de componentes extraídos, devido à natureza das imagens de sensoriamento remoto.

O processo de encontrar, rastrear e monitorar eventos e atividades de interesse de maneira contínua constitui os maiores requisitos nas tarefas de vigilância e reconhecimento inteligentes. Além disso, sistemas aéreos não tripulados constituem um dos meios primários mais responsivos para coleta de informações de inteligência e se tornam vitais em operações militares, por exemplo. [Leininger et al. \(2008\)](#) propõem em seu trabalho uma discussão sobre os elementos do programa ARGUS-IS (*Autonomous Real-time Ground Ubiquitous Surveillance – Imaging System*) que constitui um sistema capaz de gerar imagens de uma área maior do que 40 quilômetros quadrados com uma distância espacial do solo de 15 cm em taxas de vídeo maiores do que 12 Hz, sendo composto por três subsistemas, a saber: um sensor gigapixel, um sistema de processamento com sensores aéreos e um sensor terrestre de processamento.

O trabalho proposto por [Cao et al. \(2019\)](#) introduz um método para contagem de pessoas presentes em uma cena contendo uma multidão em uma imagem de alta resolução. A abordagem faz uso de uma estratégia de varredura com janela deslizante sobre a imagem. Uma rede neural CSRNet é treinada para gerar um mapa de densidade, utilizado, na sequência, para a contagem das pessoas presentes na cena.

[Fan et al. \(2022\)](#) propuseram uma abordagem de detecção de objetos em

imagens Gigapixel com arranjos de fragmentos da mesma. Estratégias de filtragem propostas por [Wang et al. \(2020b\)](#) apoiadas em Redes Neurais foram usadas para a detecção dos fragmentos, ocorrendo, em seguida, a segmentação das pessoas presentes nos mesmos.

O trabalho desenvolvido por [Lyu et al. \(2019b\)](#) propuseram um novo método para visualização de imagens Gigapixel panorâmicas. Este método foi apoiado no uso de GPU's, a fim de viabilizar operações, em tempo real, de deslocamento (*pan*), inclinação (*tilt*) e *zoom*. O método proposto apresentou bons resultados, mesmo com o uso de GPU's de baixo custo.

[Cruz & González-Villa \(2021\)](#) desenvolveram uma abordagem para estimar o tamanho de uma população presente em uma imagem Gigapixel. Para tal, os autores propuseram o uso de informações geométricas extraídas manualmente de 50 a 100 regiões aleatórias da imagem, obtendo baixas taxas de erros em torno de 5 a 10 por cento em seus resultados. O estimador proposto apresentou baixa influência dos erros de calibração e projeção.

Esta seção apresentou trabalhos relacionados a algumas das áreas de pesquisa que estão trabalhando com imagens Gigapixel. As seções seguintes mostram trabalhos que buscam detectar pedestres e faces e, na sequência, trabalhos com o mesmo objetivo em imagens de alta resolução.

2.3 Detecção de Pedestres e de Faces

As pesquisas relacionadas à detecção de pedestres vêm ganhando atenção especial nos últimos anos devido às necessidades trazidas, por exemplo, pelos carros autônomos, pela robótica, pela busca e identificação de pedestres, entre outras aplicações. O mesmo vem ocorrendo com as pesquisas de detecção de faces, as quais vêm se desenvolvendo com o objetivo de reduzir o esforço do observador para identificar um determinado rosto em um número cada vez maior de imagens digitais.

Nesse contexto, tem-se que a extração de características constitui uma etapa importante do processo de detecção de pedestres e de faces. Assim, observa-se a existência de dois agrupamentos de métodos de extração de características, a saber: métodos apoiados em extração manual de características (*Histogram of Oriented Gradients* - HOG, *Support Vector Machine* - SVM, etc.) e aqueles apoiados no uso de redes neurais convolucionais (*Convolutional Neural Networks* - CNNs).

Dessa maneira, [Viola et al. \(2005\)](#) apresentam um trabalho de detecção de pedestres que integra informações de intensidade da imagem com informações de movimento. Em seu sistema, eles utilizam um algoritmo de detecção de estilo que varre um detector em dois quadros consecutivos de uma sequência de vídeo.

Usando o algoritmo AdaBoost, o detector é treinado aproveitando as informações de movimento e aparência na detecção de uma pessoa andando, combinando ambas as fontes de informação em um único detector e identificando pedestres em escalas bem pequenas, como, por exemplo, 20×15 pixels. O detector dinâmico de pedestres construído é baseado nos filtros retangulares simples apresentados por Viola & Jones (2001) no problema de detecção de face estática, os quais combinam informações de movimento e aparência. Tais filtros foram estendidos para atuarem em pares de movimento e extraírem informações relacionadas à probabilidade de uma determinada região estar se movendo em uma direção, medindo a magnitude de movimento em uma das imagens em movimento, além do uso de filtros de aparência que são filtros retangulares que operam na primeira imagem de entrada.

Dalal & Triggs (2005) propõem o estudo de conjuntos de características para reconhecimento de objetos, adotando a detecção de pedestres baseada em características HOG e no classificador linear SVM. O estudo realizado utilizou o banco de dados de pedestres do MIT (*Massachusetts Institute of Technology*). A abordagem proposta considera uma janela deslizante através de blocos sobrepostos e são extraídos vetores de características HOG. Os vetores combinados alimentam um SVM para a classificação de objeto e não-objeto. A janela de detecção desliza através da imagem em todas as posições e escalas e a supressão não-máxima convencional é executada na pirâmide de saída para detectar instâncias de objetos.

Cao *et al.* (2008) propõem um sistema de detecção de pedestres que, primeiramente, obtém a aparência e a informação de movimento a partir de quadros (*frames*) sequenciais e, em seguida, procura realizar o reconhecimento de seres humanos a partir da combinação de aprendizagem estatística e SVM. Por fim, o método proposto estima a distância do pedestre ao veículo e identifica a sua direção através do uso do algoritmo proposto por Cao *et al.* (2005).

A detecção de faces em imagens digitais consiste em uma das tarefas mais fundamentais da visão computacional, tendo sido iniciada com o trabalho de Viola & Jones (2001) apoiados em características da imagem como aquelas extraídas a partir de *wavelets* de Haar, conforme descrito no trabalho de Papageorgiou *et al.* (1998), e *Local Binary Patterns* (LBP), como discutido por Zhang *et al.* (2007) e, também, *Speeded-Up Robust Features* (SURF), como detalhado por Li & Zhang (2013).

O método SURF caracteriza um algoritmo para extração de características locais da imagem que são robustas a eventuais transformações aplicadas na mesma. As principais aplicações dessa abordagem estão relacionadas com o processo de reconhecer e classificar objetos através da rápida computação de operadores usando filtros com janelas deslizantes na imagem. Já a abordagem LBP representa um tipo de descritor bastante utilizado para a caracterização de elementos de textura da imagem,

em que sua regularidade pode ser determinada através da forma de distribuição do histograma LBP.

Atualmente, as características da imagem para fins de detecção de faces estão sendo extraídas a partir do uso de aprendizado profundo (*deep learning*), mais especificamente, redes neurais convolucionais (*Convolutional Neural Networks* - CNNs), com bons resultados, conforme pode ser observado nos trabalhos de [Li et al. \(2015a\)](#) e [Zhang et al. \(2016\)](#) e, ainda, mais recentemente, no trabalho desenvolvido por [Luo et al. \(2019\)](#).

[Li et al. \(2015b\)](#) revisitam a abordagem popular de filtragem de características em múltiplos canais, conforme proposto por [Viola & Jones \(2004\)](#), aliado à utilização de CNNs a fim de propor um novo método denominado de *Convolutional Channel Features* (CCF), o qual procura enriquecer as características extraídas a partir de análise de múltiplos canais com características de baixo nível extraídas a partir de uma rede CNN e, dessa forma, obter melhorias na detecção de pedestres e faces, além de demonstrar potencial para ser utilizado em plataformas de *hardware* modesto como, por exemplo, *smartphones*.

Mais recentemente, um interessante *survey* foi produzido por [Feng et al. \(2022\)](#) sobre detecção de faces, o qual promove uma avaliação da acurácia e da eficiência dos principais detectores de faces baseados CNNs.

Por fim, [Mao et al. \(2017\)](#) propõem um método misto que procura agregar características intrínsecas da imagem à extração de características apoiada em CNN. Assim, são avaliados os canais semânticos aparentes, apoiados no uso de HOG, e os canais temporais (fluxo ótico e movimento). Além disso, são avaliados, ainda, os canais de profundidade de imagens estéreo. Os experimentos realizados mostram que os recursos do canal semântico podem ajudar os detectores de pedestres a discriminar amostras positivas e negativas em baixa resolução, enquanto que os recursos de canal aparente inibem falsos positivos de fundos e melhoram a precisão da localização em alta resolução.

O trabalho desenvolvido por [Hu & Ramanan \(2017\)](#) propõe o modelo *TinyFaces* para detecção de faces que considera em sua abordagem o papel da invariância de escala, a resolução da imagem e o contexto em que a imagem se encontra. O modelo proposto considera redes CNNs compartilhadas, que são treinadas separadamente para escalas diferentes e de uma maneira multi-tarefa em que características são extraídas de múltiplas camadas de uma hierarquia de características única, formando uma mapa de resposta. Por fim, é realizado um processo de interpolação na camada mais baixa da pirâmide de imagens com o objetivo de facilitar a detecção de objetos pequenos. O contexto espacial é levado em consideração e são definidos modelos que fazem uso massivo de campos receptivos

que estendem para além do objeto de interesse. Além disso, o papel da escala é explorado por redes neurais pré-treinadas que extrapolam escalas limitadas para faixas bem maiores.

2.4 Detecção de Pedestres e Faces em Imagens de Alta Resolução

A detecção de pedestres e faces em imagens Gigapixel apresenta grandes desafios em virtude da natureza da imagem, visto que a mesma é de alta resolução, com grande quantidade de dados e formada como um imenso mosaico de várias imagens, o que ocasiona vários problemas a serem discutidos no Capítulo 3. Entretanto, as imagens Gigapixel são de grande relevância em várias situações, como, a título de exemplo, para vigilância de áreas de grande aglomeração em que há potencial de ocorrência de crimes, entre outros.

Assim, de início, cumpre destacar o *survey* a respeito de técnicas de detecção e acompanhamento de pedestres escrito por [Brunetti et al. \(2018\)](#), o qual destina uma seção inteira para discutir o uso de redes neurais convolucionais como metodologia para detecção e acompanhamento de pedestres.

Os autores [Wang et al. \(2020a\)](#) apresentam um banco de dados (PANDA) de vídeo em nível Gigapixel para análise visual centrada no ser humano em larga escala e com vários objetos. Os vídeos em PANDA possuem ampla FoV e alta resolução espacial. O trabalho mostra o estado da arte de algoritmos para tarefas centradas em humanos, detecção de pedestres e rastreamento deles, considerando questões como variação significativa da pose dos pedestres, escala, oclusão, trajetória, dentre outros. O trabalho propõe um ambiente de global-local de *zoom-in* que combina trajetórias globais e interações locais, produzindo um desempenho promissor na detecção de grupos.

Já [Zhang et al. \(2021\)](#) propõem um algoritmo de aceleração de detecção de pedestres em imagens de alta resolução chamado de *Non-Pedestrian Area Estimation* (NPAE), o qual estima e remove áreas onde não existem pedestres e, então, promove a detecção de pedestres nas áreas da imagem onde os pedestres existem. O algoritmo RetinaNet, proposto por [Lin et al. \(2020\)](#), foi escolhido como detector de pedestres e dois conjuntos de dados com três imagens de resolução diferentes são usados para testar o algoritmo NPAE proposto nas plataformas GPU (*Graphics Processing Unit*) e CPU (*Central Processing Unit*). Os conjuntos de dados utilizados foram o *Joint Attention in Autonomous Driving* (JAAD), desenvolvido por [Rasouli et al. \(2017\)](#) e o Caltech, detalhada em [Dollar et al. \(2010b\)](#).

O trabalho apresentado por [Li et al. \(2022\)](#) traz, então, uma abordagem de detecção de pedestres em imagens de alta resolução em grande escala. É usada uma estratégia de janela deslizante para realizar um processo de pré-deteção e, em seguida, são feitos recortes que possuem o objeto de interesse no centro e em várias escalas. É utilizado um algoritmo de região de Supressão Não-Máxima (NMS), juntamente com uma estratégia de fusão, promovendo a eliminação de redundância de informações e possibilitando a detecção de pedestres. O método proposto tem sua eficácia testada com os modelos *Faster R-CNN*, *Cascade R-CNN*, *IterDet* e *Scale-aware Fast R-CNN* no conjunto de dados da base PANDA, proposta por [Wang et al. \(2020a\)](#). Especificamente, os autores desenvolveram seu modelo considerando um módulo de recorte para obtenção das imagens de entrada para a rede. Na sequência, foi utilizado um módulo de detecção no experimento e, por último, o algoritmo NMS é usado para a fusão dos resultados de detecção. O uso da base de dados PANDA é comparável à base Caltech, totalizando um número de 78 mil pedestres.

Os autores [Mo et al. \(2022\)](#) propõem o uso de uma rede de detecção de pedestres ADN (Rede profunda baseada em ATSS - Seleção de Amostras de Treinamento Adaptativa) em imagens de alta resolução, levando em consideração que imagens dessa natureza são mais difíceis de serem processadas diretamente, além de questões como oclusão de pedestres e variações de poses. É utilizada uma estratégia de divisão da imagem para que o algoritmo existente possa ser usado nas imagens. Em seguida, é usada uma rede ResNet50 como rede *backbone* para extrair características em diferentes escalas e evitar a degradação da rede. Na sequência é adicionada uma convolução deformável à rede *backbone*, que fortalece sua capacidade de modelagem de objetos deformáveis, melhorando a acurácia da detecção de pedestres em diferentes poses e em situações de oclusão.

O trabalho desenvolvido por [Mo et al. \(2023\)](#) propõe um detector multi-alvo de pedestres e veículos em imagens Gigapixel denominado PVDet. Os autores fazem uso de uma rede residual deformável (DPRNet) projetada como *backbone* para melhorar o campo perceptual efetivo e a representação de características da variação de poses e oclusões. Na sequência, a rede PAFPN (rede piramidal de características de agregação de caminhos) é usada para o processo de extração de características em várias escalas, buscando aumentar a capacidade de modelagem de alvos, bem como a localização de alvos pequenos. Além disso, módulos DyHead são utilizados para melhorar o processo de detecção e classificação de pedestres e veículos.

Os autores [Wang et al. \(2023\)](#) desenvolveram um método para detectar pedestres e veículos, utilizando uma estratégia da linha média para resolver o problema de um grande número de alvos em imagens recortadas através do uso de uma janela deslizante sobre a imagem Gigapixel. O trabalho também levou em consideração

que pedestres e veículos são separáveis em tamanho. A performance do detector com uma única classe de objetos foi melhorada para detectar pedestres e veículos, respectivamente. Foi utilizada a perda varifocal para resolver o desequilíbrio entre as amostras positivas e negativas. O método apresentou limitações no desempenho para se trabalhar com objetos grandes a serem recortados.

2.5 Considerações Finais do Capítulo

Este capítulo apresentou uma revisão de literatura com foco em detecção de pedestres em imagens de alta resolução, destacando o uso crescente de redes neurais convolucionais e de técnicas de seleção de áreas das imagens com maior probabilidade de existência de pedestres, reduzindo, dessa forma, o esforço computacional. O próximo capítulo discutirá sobre a geração e visualização de imagens de alta resolução a partir de um processo de varredura.

Geração de Imagens Gigapixel por Processo de Varredura de Cena e Algumas Técnicas de Visualização de Imagens Panorâmicas

Sumário

3.1	Introdução	34
3.2	Varredura de Cena para Captura de Imagens	38
3.3	Alinhamento e Combinação de Imagens	39
3.4	Visualização de Imagens Panorâmicas	42
3.5	Exemplo de Giga Imagem Gerada a Partir de um Processo de Varredura de Cena	43
3.6	Considerações Finais do Capítulo	48

3.1 Introdução

Antes de introduzir a discussão sobre a geração das imagens Gigapixel e as técnicas de visualização das mesmas neste capítulo, torna-se importante salientar alguns pontos relacionados com a geração de imagens panorâmicas, bem como as etapas e dificuldades encontradas neste processo de combinação das várias imagens geradas para a construção de uma única imagem final.

Nesse contexto, ao se pensar em gerar uma imagem de uma cena que é algo tridimensional, temos, de fato, um conjunto de imagens obtido de aproximadamente um mesmo ponto de vista, mas considerando a rotação da câmera utilizada para sua geração. Entretanto, à medida em que a câmera é rotacionada, existirão regiões das imagens individuais que serão capturadas com sobreposição e precisarão ser tratadas ao se gerar uma combinação das mesmas para uma única imagem.

Segundo evidenciado por [Lyu et al. \(2019a\)](#), existem duas categorias de técnicas de combinação de imagens que podem tratar o problema explicitado anteriormente, a saber: métodos *pixel-based*, também conhecidos como métodos diretos, e os métodos *feature-based*. A primeira categoria de métodos mencionada se utiliza de informações da imagem baseadas em *pixels*, tais como gradiente, cor, profundidade e geometria, estimando um modelo de transformação global para deformar e alinhar as imagens. Tais métodos aproveitam as informações da imagem, entretanto, exigem um pré-processamento e cálculos complexos, limitando-se a abordar as imagens com cenas simples contidas no mesmo plano.

Já a segunda categoria mencionada de métodos baseados em características usa descritores de características esparsas e sequencialmente implementa a correspondência das mesmas. Esses métodos são mais rápidos ao extrair e selecionar características esparsas ao invés de usar informações de todos os pixels das regiões sobrepostas, o que os faz mais robustos para a automatização das tarefas de cálculo de adjacências de imagens de entrada. Nesse sentido, para um desempenho mais rápido e eficiente, algumas técnicas são utilizadas para a etapa de combinação das imagens, a saber: *Scale-invariant Feature Transform* - SIFT ([Lowe, 2004](#)), SURF ([Li & Zhang, 2013](#)), Harris ([Harris & Stephens, 1988](#)), *Features from Accelerated Segment Test* - FAST ([Rosten & Drummond, 2006](#)), além de matrizes de transformações múltiplas, como homografia e afins, calculadas de maneira simultânea ao processo de combinação de imagens.

As imagens Gigapixel, por sua vez, constituem imagens panorâmicas e trazem uma quantidade de detalhes muito grande e relevantes para sua análise. Entretanto, o processo de geração das mesmas, caso seja feito, por exemplo, de forma manual, pode trazer inúmeros problemas, visto que é bastante sujeito a erros, como, por exemplo, alinhamento da câmera para produção das imagens individuais, preocupação com o espaçamento/deslocamento físico para geração de imagens próximas umas das outras, controle do tempo de geração das imagens separadas para minimizar a influência da iluminação na geração das imagens, ausência de equipamento de suporte para estabilizar o posicionamento da câmera, dentre outros, além de ser um processo tedioso e cansativo.

Com a expansão da robótica, a percepção das cenas se tornou diferente, permitindo novas pesquisas e desenvolvimento de aplicações. Com isso, a evolução dos equipamentos e a maneira como as imagens são geradas mudaram, e alguns equipamentos começaram a ser pensados e desenvolvidos para geração de várias imagens.

A título de exemplificação destacamos, o Projeto GigaPan¹, o qual se apresenta como sendo um experimento de uso da robótica para geração e visualização de imagens a partir de um processo de varredura, criando cenas em altíssima resolução, a saber: cenas panorâmicas Gigapixel, as quais constituem imagens digitais com bilhões de pixels.

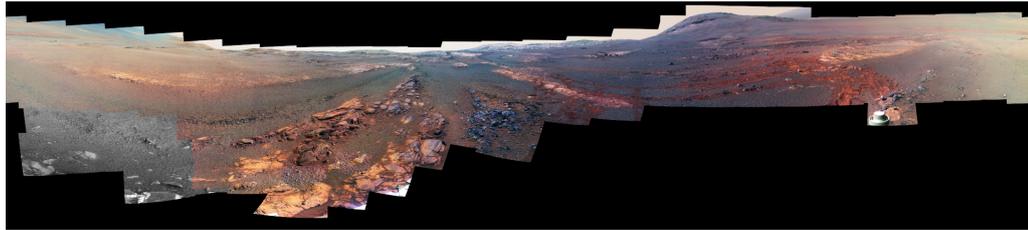
O Projeto GigaPan teve suas bases durante o desembarque dos Veículos de Exploração do planeta Marte *Spirit* e *Opportunity*, da Missão *Mars Exploration Rovers*², da *National Aeronautics and Space Administration* (NASA). Os veículos robóticos reuniam várias características bastante específicas e, dentre elas, a capacidade de produzir imagens panorâmicas de alta resolução, a partir de um processo de varredura de uma cena e correspondente montagem das múltiplas imagens derivadas do processo, uma ao lado da outra. Estas imagens panorâmicas de alta resolução permitiam aos cientistas selecionar uma determinada parte da mesma e, então, ampliar aquela parte para inspecioná-la, como se estivessem com uma lupa nas mãos. Um exemplo de imagem do robô *Opportunity* pode ser visto na Figura 3.1, a qual apresenta uma imagem panorâmica original de 360°, composta por 354 imagens.

Os resultados obtidos com as imagens panorâmicas produzidas pela Missão *Mars Exploration Rovers* inspiraram a aplicação da mesma tecnologia na Terra. Assim, conforme relatado por Nourbakhsh (2010), após quase dois anos de trabalho, o robô GigaPan foi desenvolvido, o qual, em vez de usar equipamento ótico caro, faz uso de uma câmera digital comercial de pequeno formato que realiza movimentos precisos a fim de realizar o processo de varredura de uma cena, permitindo que as imagens sejam obtidas com mínima sobreposição e com a calibração adequada.

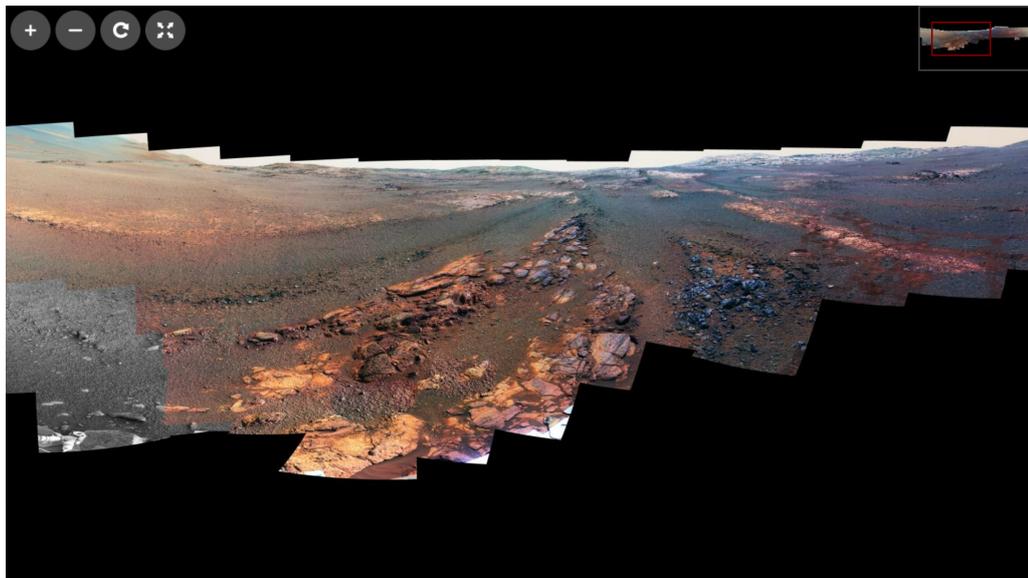
As imagens panorâmicas obtidas a partir de um robô GigaPan são conhecidas como imagens panorâmicas Gigapixel, ou simplesmente imagens Gigapixel, também referenciadas neste trabalho como giga imagens. O processo de obtenção destas imagens está ligado à geração de várias imagens individuais, que são usadas, posteriormente, para a montagem de uma imagem panorâmica Gigapixel final. Assim, estas imagens podem ser entendidas como a montagem de um mosaico a partir de um grande número de imagens digitais de alta resolução (Powell et al., 2010). Dessa forma, a montagem de uma imagem Gigapixel não se limita a enquadrar um cenário e registrá-lo em apenas um clique da câmera utilizada. Ao contrário disso, cada parte da imagem Gigapixel é obtida (fotografada) individualmente e, em seguida, é feito o processo de combinação de todas essas partes para formar a grande imagem Gigapixel. O resultado é a imagem Gigapixel que consegue trazer para perto dos

¹Projeto GigaPan: <http://gigapan.com/>

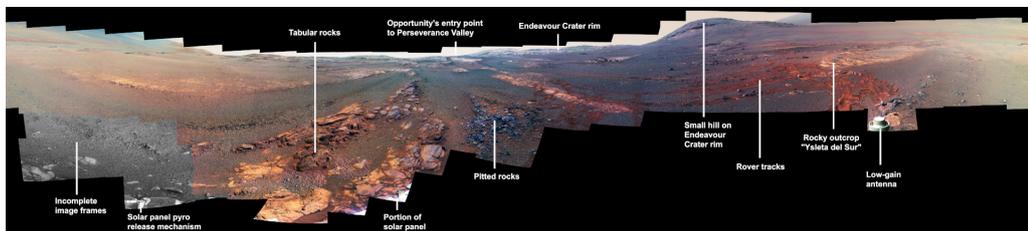
²Missão *Mars Exploration Rovers*: <https://mars.nasa.gov/mer/>



(a) Imagem original.



(b) Imagem com foco em uma região.



(c) Imagem com anotações.

Figura 3.1: Imagens do robô *Opportunity*, extraídas de (NASA et al., 2019).

olhos do observador detalhes mínimos, mesmo que estejam aparentemente longe e não possam ser observados com facilidade considerando uma imagem convencional.

O uso de imagens Gigapixel tem se mostrado bastante interessante em trabalhos de várias áreas como geologia, paleontologia, segurança, planejamento urbano, empreendimentos imobiliários, dentre outras em que seja necessária uma visualização mais detalhada de uma grande cena, visto que imagens Gigapixel trazem, em si, detalhes em alta resolução em um contexto amplo e, muitas vezes, não observados em outras situações (Nichols et al., 2009).

Entretanto, há passos anteriores à utilização dessas imagens, os quais compreendem a captura através do uso de um robô GigaPan, que serão explicados a

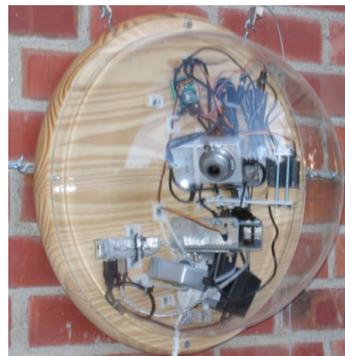
seguir e, posteriormente, a combinação e a visualização de imagens Gigapixel com dados espaço-temporais.

3.2 Varredura de Cena para Captura de Imagens

A primeira etapa para se obter uma imagem Gigapixel é a varredura de uma cena através do uso de um robô GigaPan com uma máquina fotográfica comercial acoplada a ele a fim de se obter um conjunto de imagens que irão compor a imagem panorâmica.

Durante o processo de captura das imagens, o robô GigaPan movimentava a câmera acoplada a ele no sentido horizontal e também muda sua inclinação na vertical para que uma sequência de fotos individuais seja obtida ao longo do tempo.

O protótipo original do robô GigaPan, bem como o *software* relacionado, foi desenvolvido por [Sargent et al. \(2010\)](#), os quais desenvolveram dois sistemas de captura de imagens Gigapixel com lapso temporal, mostrados na Figura 3.2.



(a) Protótipo inicial do GigaPan.



(b) GigaPan Epic Pro.

Figura 3.2: Sistemas de captura de imagens GigaPan. Imagem extraída de [Sargent et al. \(2010\)](#).

O protótipo apresentado na Figura 3.2a utilizava-se de um computador pessoal (PC) para sequenciar suas ações. O computador controlava uma câmera Canon S400, ligada através da entrada USB do computador. Esse protótipo conseguiu trabalhar na captura de imagens por um período de 6 meses. Entretanto, trouxe vários problemas associados como, por exemplo, o servo-motor falhando por diversas vezes, levando à necessidade de sua substituição. Além disso, a câmera travava em várias ocasiões, conseguindo capturar, em média, entre 2.000 e 5.000 imagens antes de falhar ([Sargent et al., 2010](#)).

Já o equipamento Gigapan Epic Pro, apresentado na Figura 3.2b, foi comercializado inicialmente no ano de 2010, incluindo um obturador controlado eletronicamente e uma bateria recarregável. As câmeras usadas foram a Canon G10 e G11 para operações de lapso de tempo, procurando a obtenção de imagens com a menor abertura antes de ocorrer embaçamento devido à difração. Essas câmeras possuem a capacidade de reverter o foco automático quando se encontram ociosas por um determinado tempo, visto que uma opção de *firmware* mantém o obturador ativo após a última foto de um panorama, evitando a perda da configuração de foco.

Quando o GigaPan Epic Pro é usado para captura de uma giga imagem é preciso selecionar a geometria de captura a ser utilizada, salvando-a a partir da opção do menu *Memória do Panorama*. Esse processo inclui definir um ponto de referência do ambiente, de modo que, caso a câmera seja substituída ou a montagem do dispositivo seja alterada, é possível que o usuário recalibre o sistema apontando a câmera novamente para o ponto de referência definido.

As imagens são armazenadas diretamente no cartão SDHC da câmera, com uma capacidade máxima de 32 GB, considerando a configuração de mais alta qualidade e sendo possível, então, armazenar 5.000 imagens para a G10 e 10.000 para a G11. Essa diferença de capacidade de número de imagens armazenadas se deve ao fato de níveis de compressão diferenciados, bem como contagens diversas de números de pixels.

É importante entender que armazenar e gerenciar centenas de milhares de imagens da câmera para uma sequência de lapso temporal é uma tarefa desafiadora. Uma aplicação precisa ser usada para importar automaticamente as imagens dos cartões, dividindo-as em panoramas separados e também permitindo que o usuário inspecione tais imagens antes de confirmar a operação adequada do sistema.

A seção seguinte discute o alinhamento e a combinação das imagens capturadas pelo processo de varredura realizado pelo robô GigaPan.

3.3 Alinhamento e Combinação de Imagens

Com o objetivo de tratar o lapso de tempo para combinação das imagens que formam uma imagem Gigapixel, é preciso tratar problemas como, por exemplo, o alinhamento de várias imagens obtidas em momentos diferentes e exposição variável.

As imagens Gigapixel de quadro único são montadas a partir de uma grade de imagens de entrada, sendo que cada imagem possui uma sobreposição com suas vizinhas. Assim, é preciso realizar a união dessas imagens, realizando uma tarefa de costura, que consiste em alinhar e combinar as imagens (Abadi *et al.*, 2021). Para tal, diversas técnicas são utilizadas, dentre elas a transformada de características

invariantes à escala (*Scale-invariant feature transform* - SIFT) de áreas sobrepostas descritas por [Lowe \(2004\)](#).

Já para a realização do processo de otimização das imagens, é utilizado o método proposto por [Levenberg \(1944\)](#), o qual consegue encontrar o conjunto ótimo de localizações e orientações de imagem e, ainda, parâmetros de lente da câmera como distância focal e distorção radial, com a finalidade de promover o melhor alinhamento possível das imagens que passarão pelo processo de combinação.

Na sequência, as imagens são projetadas em um sistema comum de coordenadas e combinadas através do uso de uma variação da abordagem de *spline* multiresolução proposta por [Burt & Adelson \(1983b\)](#). Além disso, para alinhar as imagens ao longo do tempo, é adicionada a capacidade de alinhar uma nova Gigapixel através de um ou mais protótipos Gigapixel já alinhados. As características SIFT entre as imagens da nova Gigapixel e os protótipos são encontradas e adicionadas às restrições para o processo de otimização de alinhamento. Após o alinhamento, os parâmetros de projeção são selecionados a partir do primeiro protótipo, procurando garantir que a imagem Gigapixel que passou pelo processo de combinação tenha as mesmas coordenadas do primeiro protótipo.

Uma maneira simples usada é a seleção de uma única imagem Gigapixel como protótipo de todas as outras imagens no lapso de tempo, conforme mostra a Figura 3.3, adaptada do trabalho de [Sargent et al. \(2010\)](#).

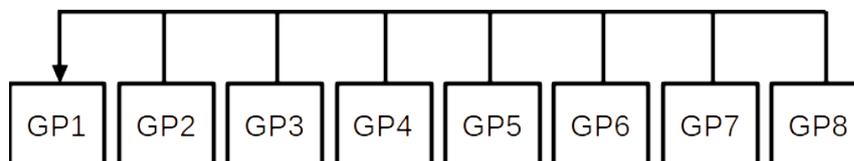


Figura 3.3: Imagem GP1 como protótipo para as outras imagens Gigapixels (GP2 a GP8) no lapso de tempo. Figura adaptada do trabalho de [Sargent et al. \(2010\)](#).

Entretanto, alguns problemas são ocasionados pela estratégia apresentada como, por exemplo, a existência de poucas partes nas imagens que permanecem constantes ao longo do tempo para servirem como marcas para o SIFT.

Outra dificuldade encontrada é que a incidência de iluminação muda muito ao longo do tempo como, por exemplo, em imagens diurnas externas. Assim, alinhar vários protótipos a partir de um único protótipo se mostra desafiador para essa abordagem.

Dessa forma, uma evolução do que foi proposto pode ser observada através da seleção de um novo protótipo para imagens sucessivas, conforme mostrado na Figura 3.4, adaptada de [Sargent et al. \(2010\)](#) e baseada na seleção de quadros-chave para compressão de vídeo digital desenvolvida por [Gall \(1991\)](#).

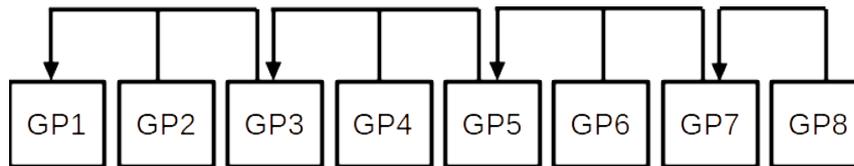


Figura 3.4: Imagens GP1, GP3, GP5 e GP7 como quadro-chaves, selecionados para serem protótipos e com o objetivo de compensar os problemas de características fixas no lapso de tempo. Figura adaptada de [Sargent et al. \(2010\)](#).

Quando se tem um lapso de tempo de uma cena externa que inclui imagens tanto diurnas quanto noturnas, ocorrem diferenças importantes nas características SIFT. Assim, pode se tornar uma tarefa árdua a tentativa de alinhamento de uma imagem Gigapixel obtida ao meio-dia com outra obtida à meia-noite. Assim sendo, uma árvore de protótipos de alinhamento pode ser usada, selecionando-se um protótipo mestre para o alinhamento em cadeia de imagens Gigapixel do mesmo dia. Com isso, a alteração de iluminação entre elas é diminuída e a quantidade de características SIFT pode se tornar suficiente. A cadeia de imagens Gigapixel se torna, então, protótipos para imagens dos dias seguintes, tendo cada panorama sucessivo alinhado para o protótipo com a iluminação mais semelhante, conforme pode ser observado na Figura 3.5, adaptada do trabalho de [Sargent et al. \(2010\)](#).

Em geral, para uma única imagem Gigapixel, o fotógrafo escolhe uma configuração de exposição única para todas as imagens da câmera que estão sendo unidas em um único mosaico. Com isso, as transições entre as imagens vizinhas são minimizadas.

Entretanto, em muitos lapsos de tempo com a iluminação se alterando, é comum não se ter o fotógrafo presente para verificar uma boa exposição para cada captura de imagem Gigapixel individual. A solução está em permitir que a própria câmera selecione a exposição em um modo automático. Em geral, as câmeras gravam a exposição selecionada para cada imagem em metadados EXIF que estão no próprio arquivo da imagem.

Na iniciativa de procurar compensar as diferenças de exposição, o processo de combinação das imagens procura considerar as informações de velocidade do obturador, da abertura e das configurações ISO de cada imagem de entrada, para, então, calcular a exposição média da cena e definir a correção de exposição de cada imagem baseada na média mencionada.

A seção seguinte realiza o detalhamento do processo de visualização das imagens Gigapixel.

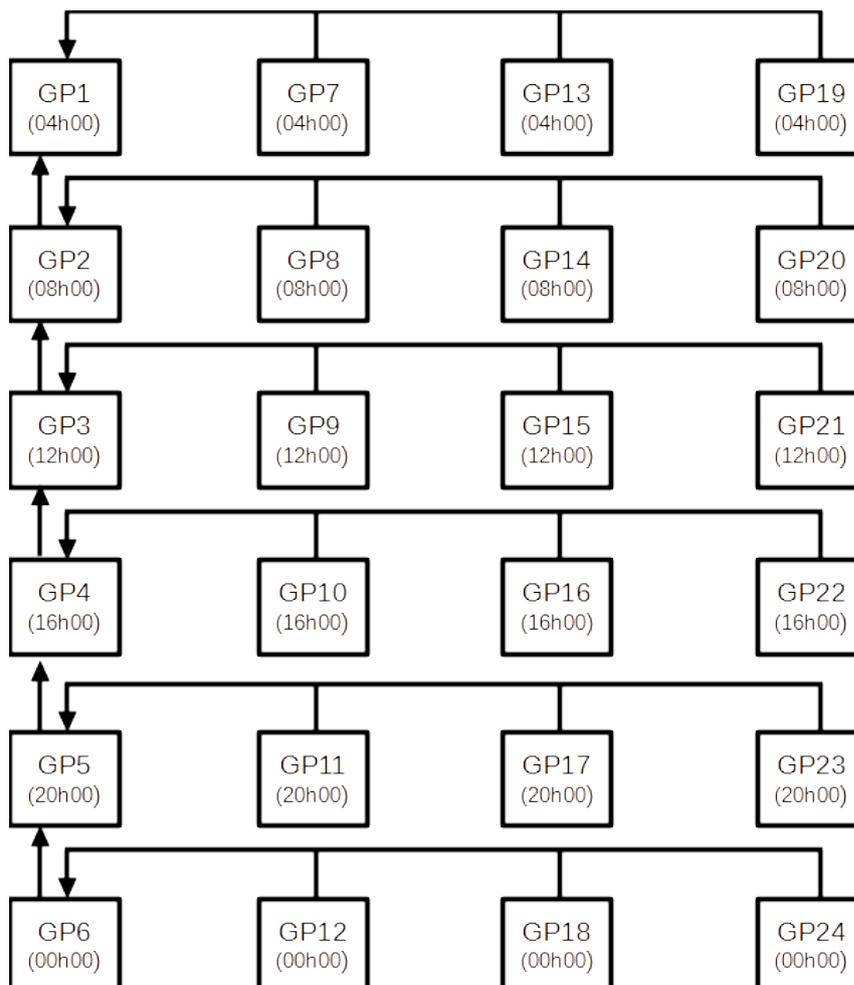


Figura 3.5: Imagens do primeiro dia alinhadas em cadeia e usadas como protótipos para dias sucessivos. Figura adaptada de Sargent et al. (2010).

3.4 Visualização de Imagens Panorâmicas

Normalmente, imagens Gigapixel são muito grandes para serem compartilhadas, podendo ocasionar um tempo grande ou a inviabilidade para serem, por exemplo, transferidas completamente através de uma conexão à Internet. Isso pode se tornar um problema maior quando consideramos um lapso de tempo em escala gigapixel, visto que a quantidade de dados pode, facilmente, alcançar os Terabytes.

Nesse sentido, o Gigapan busca transmitir as imagens individuais, em tempo real, para o visualizador, baseando-se na área atual de foco dentro da imagem. Para realizar esse processo de *streaming* em tempo real, a imagem é dividida e renderizada em pequenos blocos de várias resoluções. Dessa forma, de maneira análoga ao realizado pelo Google Maps³, por exemplo, o visualizador se utiliza apenas dos blocos necessários para uma visualização específica.

³Google Maps: <https://maps.google.com/>

Com o objetivo de mostrar imagens Gigapixel ao longo do tempo, o visualizador poderia alternar entre conjuntos de quadros individuais, mas isso leva a questões como a necessidade de grande largura de banda e de poder computacional da Unidade Central de Processamento (CPU) da máquina. Dessa maneira, uma solução foi dividir cada Gigapixel individual em blocos multi-resolução com as mesmas dimensões de pixels.

Em seguida, os grupos de blocos são classificados por tempo e compactados em uma pilha de blocos em um arquivo de vídeo, através de um *codec* H.264 (Richardson, 2011). O visualizador permite explorar com fluidez a Gigapixel, tanto no espaço, quanto no tempo. À medida em que ocorre a ampliação de uma imagem, o visualizador solicita novos blocos de vídeo do servidor, descartando os antigos que não são mais visíveis. Dessa forma, cada bloco de vídeo é exibido como um vídeo individual.

3.5 Exemplo de Giga Imagem Gerada a Partir de um Processo de Varredura de Cena

Vários exemplos de imagens geradas a partir de um processo de varredura se encontram disponíveis em Val (2013). Um deles é a giga imagem “Jogo Final Brasil x Espanha - 2013”, que tem uma versão miniaturizada mostrada na Figura 3.6.



Figura 3.6: Imagem Gigapixel, com destaque para um jogador através de uma operação de *zoom in*, adaptada de Val (2013).

Essa imagem ilustra o jogo da final da Copa das Confederações no ano de 2013, realizado no estádio do Maracanã, no Rio de Janeiro, Brasil, no qual as seleções do Brasil e da Espanha se enfrentaram. A giga imagem original se encontra disponível em Val (2013), está sob a licença Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License⁴ e foi gerada por Mauricio Val, FV Imagem⁵,

⁴Creative Commons BY-NC-SA 3.0 US: <http://creativecommons.org/licenses/by-nc-sa/3.0/us>

⁵Gigapan profile page: <http://www.gigapan.com/profiles/fvimagem>

Parâmetros	Valores
Modelo da câmera	Nikon D7000
Tamanho da imagem	4.928 x 3.264 (16,1 megapixels)
Abertura da lente	f/4,8 - f/9
Tempo de exposição	0,004 segundos
Tempo de captura	19:58:28 a 20:21:21 de 30/06/2013
Sensibilidade	ISO 2.000
Comprimento focal	300,0 mm
Zoom digital	×0 - <i>off</i>
Balanço branco	Automático
Modo de exposição	Automático
Sobreposição horizontal	24,0 a 29,7 por cento
Sobreposição vertical	17,4 a 35,2 por cento
Estatísticas do computador	8 GB de RAM, 2 CPUs
Tempo total	26:44 (7,4 segundos por imagem)
Alinhamento	2:40
Projeção	2:23
<i>Blending</i>	21:31
Versão do <i>software</i> de combinação de imagens	Versão Efx 2.1.0160 (Macintosh)
Tamanho do panorama	2121 megapixels (100.224 x 21.172 pixels)
Imagens de entrada	216 (27 colunas × 8 linhas)
Campo de visão	137,4 graus de largura × 29,0 graus de altura

Tabela 3.1: Características e parâmetros de configuração para geração da giga imagem Jogo Final Brasil x Espanha no Maracanã em 2013.

em 30 de junho de 2013. Algumas características dessa giga imagem, bem como parâmetros de configurações de equipamentos, são mostrados na Tabela 3.1.

Vários problemas podem ser encontrados ao se desenvolver este trabalho com as imagens Gigapixel, como, por exemplo:



Figura 3.7: Exemplo de duplicidade de uma pessoa na mesma cena, extraído de Val (2013).

- a) duplicidade de artefatos que se movem na cena como, por exemplo, pessoas caminhando ou correndo enquanto o dispositivo GigaPan está se movendo para gerar as imagens. Dessa forma, como a dimensão de um campo de futebol não



Figura 3.8: Exemplo de parte do corpo de uma pessoa aparecendo três vezes na mesma cena, extraído de Val (2013).

é consideravelmente grande e também as pessoas se deslocam rapidamente na cena, visto que essa é a dinâmica imposta por esse tipo de esporte, são muito comuns situações onde as pessoas aparecem em mais de um ponto, tanto no sentido horizontal, quanto no sentido vertical da giga imagem. Esse problema é evidenciado na Figura 3.7, na qual o jogador que veste a camisa de número 2 da seleção brasileira aparece destacado no canto esquerdo inferior da imagem e, também, no canto superior direito da mesma. Nessa última posição, é possível perceber outro problema: seu tronco aparece desconectado de suas pernas. Já no caso da Figura 3.8, pode-se perceber a parte inferior das pernas de uma mesma pessoa em três localizações distintas próximas umas das outras;

- b) pedaços de um mesmo artefato separados em partes de localização distintas, como o tronco do corpo de uma pessoa em uma localização e suas pernas em outras, conforme pode ser percebido na Figura 3.9;
- c) desalinhamento no eixo vertical e horizontal de partes do corpo de uma ou mais pessoas ao passar pelo processo de combinação de imagens, conforme pode ser visto na Figura 3.10. Outro exemplo de situação semelhante pode ser visto através do desalinhamento das linhas no chão do campo de futebol apresentado na Figura 3.11.
- d) geração de efeito *fantasma*, sombra ou perda de informação, conforme mostrado nas Figuras 3.12 e 3.13;
- e) a Figura 3.14 apresenta tanto o problema do desalinhamento das imagens no processo de combinação das mesmas, quanto o de perda de informação evidenciado no item anterior.

Dessa forma, é possível perceber que grande parte dos problemas evidenciados são ocasionados devido à divergência de velocidades entre o processo de captura e digitalização das imagens através do dispositivo GigaPan e o deslocamento das pessoas na cena. Estes problemas voltarão a ser discutidos mais adiante nesse texto.



Figura 3.9: Exemplo de partes do corpo de uma pessoa aparecendo em localizações distintas em uma mesma cena, extraído de Val (2013).



Figura 3.10: Exemplo de desalinhamento no processo de combinação de imagens, com destaque para os corpos das pessoas, extraído de Val (2013)..



Figura 3.11: Exemplo de desalinhamento no processo de combinação de imagens, com destaque para as linhas do campo de futebol, extraído de Val (2013).



Figura 3.12: Exemplo de perda de informação do rosto da pessoa na imagem, extraído de Val (2013).



Figura 3.13: Exemplo de perda de informação do tronco dos corpos das pessoas na imagem, extraído de Val (2013).



Figura 3.14: Exemplo de desalinhamento no processo de combinação e também perda da informação na imagem, extraído de Val (2013).

3.6 Considerações Finais do Capítulo

Neste capítulo, foi apresentada uma contextualização sobre um exemplo de geração e visualização de imagens a partir de um processo de varredura, as giga imagens. Foi evidenciado como é o processo de captura, alinhamento e combinação das imagens que as formam, além do processo de visualização das mesmas. Também foi apresentado um exemplo de giga imagem, com suas características de geração, bem como alguns problemas encontrados, que serão revisitados posteriormente.

O capítulo seguinte discutirá um exemplo de método de extração e localização de características de interesse em imagens geradas a partir do processo de varredura de imagens.

Método de Análise de Imagens Gigapixel por Varredura

Sumário

4.1	Introdução	49
4.2	Imagem Gigapixel Utilizada	50
4.3	Método Implementado	52
4.3.1	Etapa 1: Detecção de Pedestres	52
4.3.2	Etapa 2: Junção dos <i>Bounding Boxes</i>	55
4.3.3	Etapa 3: Detecção de Faces	55
4.3.4	Etapa 4: Análise Multirresolução	56
4.4	Resultados Alcançados	58
4.5	Considerações Finais do Capítulo	62

4.1 Introdução

Este capítulo descreve um método para detecção de pedestres em imagens Gigapixel, sem conhecimento prévio sobre suas características, tais como quantas pessoas estão presentes na imagem, suas dimensões em termos de pixels, sua distância do observador, se elas estão de frente para a câmera ou não, dentre outras.

Devido à alta resolução de imagens Gigapixel, algoritmos tradicionais comumente usados para encontrar pessoas (Hosang *et al.*, 2015; Dalal & Triggs, 2005), faces (Viola & Jones, 2004; Lienhart *et al.*, 2003), dentre outros, requerem um alto tempo de processamento, haja vista que eles foram normalmente idealizados para trabalhar com um conjunto pequeno de imagens que demandam um tempo de processamento razoável.

Dessa forma, uma alternativa é adaptar os processos de aplicação dos algoritmos tradicionais às imagens de altíssima resolução ou reduzir o tamanho

das imagens Gigapixel, o que pode causar perda de informação. Assim, diversos estudos experimentais para a verificação dos impactos de mudança de resolução foram conduzidos nesta tese.

Ressalta-se que, na fase inicial desta pesquisa, foram realizados estudos experimentais de multirresolução com diversos algoritmos, como a pirâmide gaussiana, por exemplo, com comparações de algoritmos com abordagem sequencial em CPU e abordagem em paralelo com GPU (CUDA) (Ferreira *et al.*, 2016), para validar o custo computacional e tempo para realização das operações. Todavia, a abordagem e resultados descritos neste capítulo estendem de forma significativa os estudos iniciais, e estão detalhados também nos trabalhos (Ferreira *et al.*, 2018) e (Ferreira *et al.*, 2020a), produzidos durante esta pesquisa.

Assim, este capítulo avalia também as consequências da redução do tamanho da imagem no processo de detecção de pessoas e o esforço computacional associado. Além disso, uma análise multirresolução é desenvolvida em imagens Gigapixel para avaliar o tempo de processamento gasto e seus impactos no algoritmo de detecção de pessoas.

4.2 Imagem Gigapixel Utilizada

Para a condução deste estudo, foi utilizada a imagem Gigapixel denominada “*Count the Cacti!!!*”¹, a qual tem uma versão miniaturizada mostrada na Figura 4.1. Essa imagem ilustra uma enorme paisagem de um campo de cactos com alguns caminhantes dispersos com a vegetação da paisagem. Para exemplificar a dimensão da imagem e grau de dificuldade de observação a olho nu, foram inseridas na imagem uma ampliação de um caminhante detectado e sua localização aproximada. A giga imagem original se encontra disponível em Schott (2009) e foi gerada por Ron Schott, em dezembro de 2008, no deserto do Arizona, nos Estados Unidos da América. A câmera utilizada foi uma Canon PowerShot S5 IS, acoplada a um GigaPan Epic para varredura de uma cena, contemplando a geração de suas imagens componentes, bem como o processo de combinação das mesmas para produzir a imagem Gigapixel usando o software específico para essa finalidade, o GigaPan Stitcher versão 0.4.3510 para Windows.

Algumas características dessa giga imagem, bem como parâmetros de configurações dos equipamentos, são mostrados na Tabela 4.1. É importante ressaltar que os parâmetros apresentados na referida tabela estão relacionados ao processo de

¹Giga imagem “*Count the Cacti!!!*” profile page: <http://www.gigapan.com/profiles/rschott>



Figura 4.1: Imagem Gigapixel, com destaque para uma pessoa através de uma operação de *zoom in*.

captura da imagem, antes da etapa de junção das mesmas para posterior utilização neste trabalho.

Tabela 4.1: Características e parâmetros de configuração para geração da giga imagem *Count the Cacti!!!*

Parâmetros	Valores
Modelo da câmera	Canon PowerShot S5 IS
Tamanho das imagens individuais	3.264x2.448 (8,0 megapixels)
Abertura da lente	f/4
Tempo de exposição	0,00125 segundos
Sensibilidade	ISO 80
Comprimento focal	434,0 mm
Zoom digital	×0 (<i>Off</i>)
Balanco branco	Fixo
Modo de exposição	Manual
Sobreposição horizontal	23,6 a 52,8 %
Sobreposição vertical	21,4 a 51,7 %
Estatísticas do computador	3 GB de RAM, 2 CPUs
Tempo total	18:18:17 (0:53 por imagem)
Alinhamento	9:24:35
Projeção	58:38
<i>Blending</i>	7:55:03
Versão do <i>software</i> de combinação de imagens	0.4.3510 (Windows)
Tamanho do panorama	4.067 megapixels (121.877x33.373 pixels)
Imagens de entrada	1.239 (59 colunas × 21 linhas)
Campo de visão	176,1 graus de largura × 48,2 graus de altura

A imagem da Figura 4.1 possui alguns caminhantes que, em média, possuem a altura média de aproximadamente 522 ± 130 pixels, visto que a imagem foi tirada em perspectiva e a altura em pixels não reflete a real altura dos caminhantes. Além disso, a imagem conta com a presença de um caminhante sentado, o que implica em uma altura menor (em termos de número de pixel) do que a esperada para uma pessoa real.

Para realizar o download dessa imagem, foi usado um script em Python e se obteve um total de 477×131 imagens, ou seja, 62.487 imagens, com 256×256 pixels RGB cada bloco, resultando em uma imagem total de 122.112×33.536 pixels, o que leva a um total de aproximadamente 12,29 GB quando descompactado.

A Figura 4.2 apresenta excertos de todos os caminhantes disponíveis na giga imagem mostrada na Figura 4.1 e todos eles foram marcados pela comunidade em geral. Todos os excertos de (a) até (i) mostram o comprimento total dos caminhantes e os mesmos estão com seus rostos borrados para fins de preservação de sua privacidade. Devido a um dos problemas já mencionados nessa tese, algumas pessoas aparecem várias vezes na imagem completa e também nos excertos. Tal fato, se deve ao tempo de atraso de digitalização do dispositivo GigaPan enquanto as pessoas caminham pela cena.

4.3 Método Implementado

Para manipular e estudar uma imagem Gigapixel, foi construído um modelo para deslocar uma janela deslizante, com 4 linhas \times 5 colunas de blocos com 256×256 pixels cada, resultando em uma janela deslizante de 1024×1280 pixels. A detecção de faces e pedestres foi realizada através de análise multirresolução com varredura da imagem Gigapixel com a janela deslizante. Também foi feito um processo de junção de *bounding boxes* (caixas delimitadoras) para reduzir a redundância e casar a face com os pedestres correspondentes.

A Figura 4.3 mostra uma visão geral do método implementado e, ainda, ilustra um exemplo da janela deslizante que se encontra desenhada como um retângulo vermelho na Figura 4.3 (a).

4.3.1 Etapa 1: Detecção de Pedestres

Na etapa de detecção de pedestre apresentada na Figura 4.3 (a), foi utilizada a *toolbox* disponibilizada por Dollar (2018) e descrita em Dollar *et al.* (2009) e Dollar *et al.* (2010a). Os autores desenvolveram um método multi-escala para detecção de pedestre com base em tipos de características múltiplas, incluindo *Aggregate Channel Features* (ACF), detalhado em Dollar *et al.* (2014), no qual foi desenvolvida uma abordagem híbrida que possui um *framework* usado em características de canais baseadas no descritor HOG (Dalal & Triggs, 2005) e que emprega uma pirâmide de imagens com amostragem esparsa para aproximar características em escalas intermediárias. Os resultados apresentados alcançaram o estado da arte na detecção

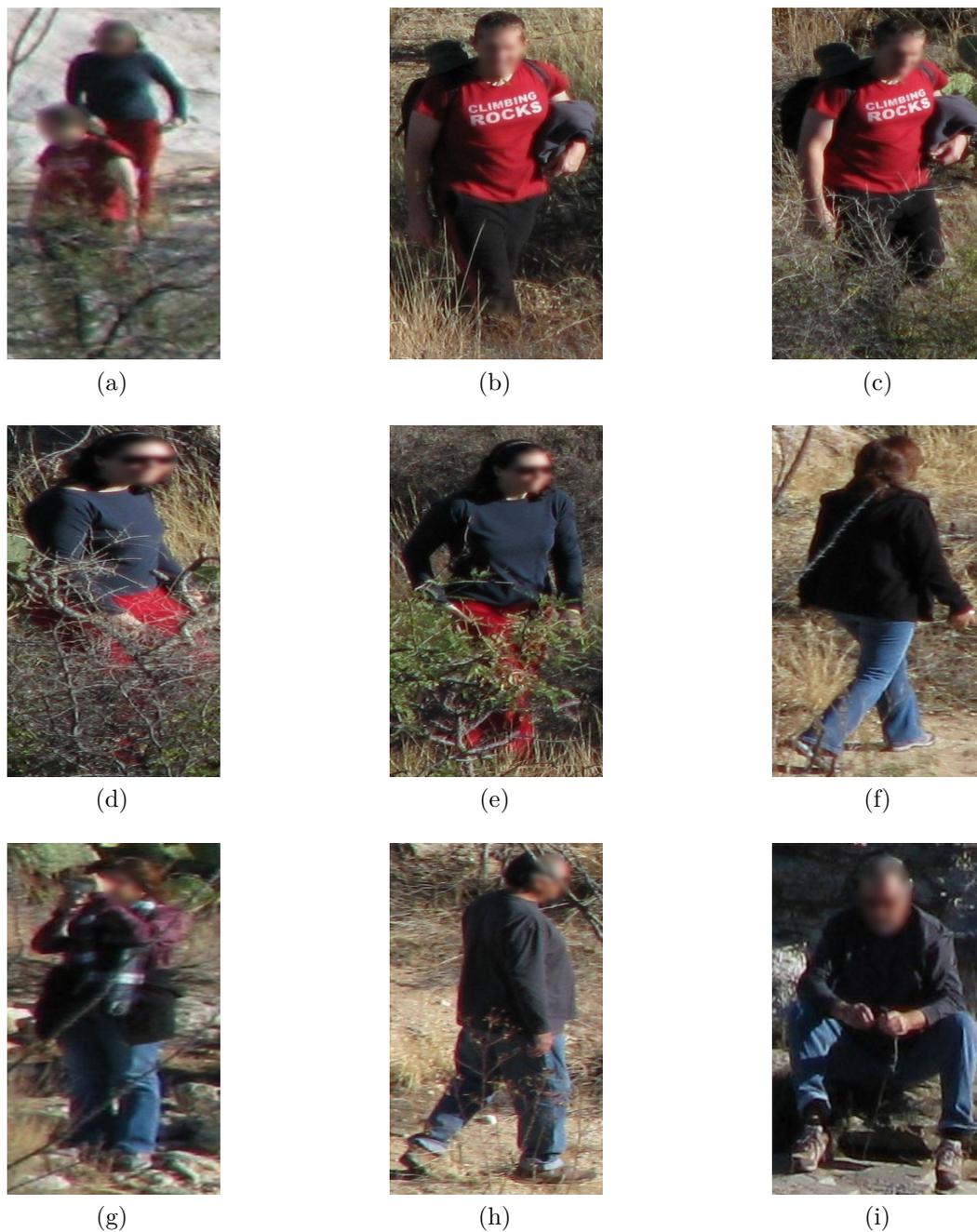


Figura 4.2: Caminhantes presentes na giga imagem avaliada.

de pedestres em termos de velocidade pela ordem de magnitude com pouca perda em acurácia.

O modelo ACF é treinado com a base de dados de pessoas do Inria, apresentada por Dalal & Triggs (2005), e, ainda, com a base de dados de pedestres da Caltech (Dollar et al., 2012). A base do Inria é composta de dois subconjuntos, com o primeiro contendo 1805 imagens de 128×64 pixels, onde os pedestres possuem aproximadamente 96 pixels de altura, e o segundo contém janelas cortadas em

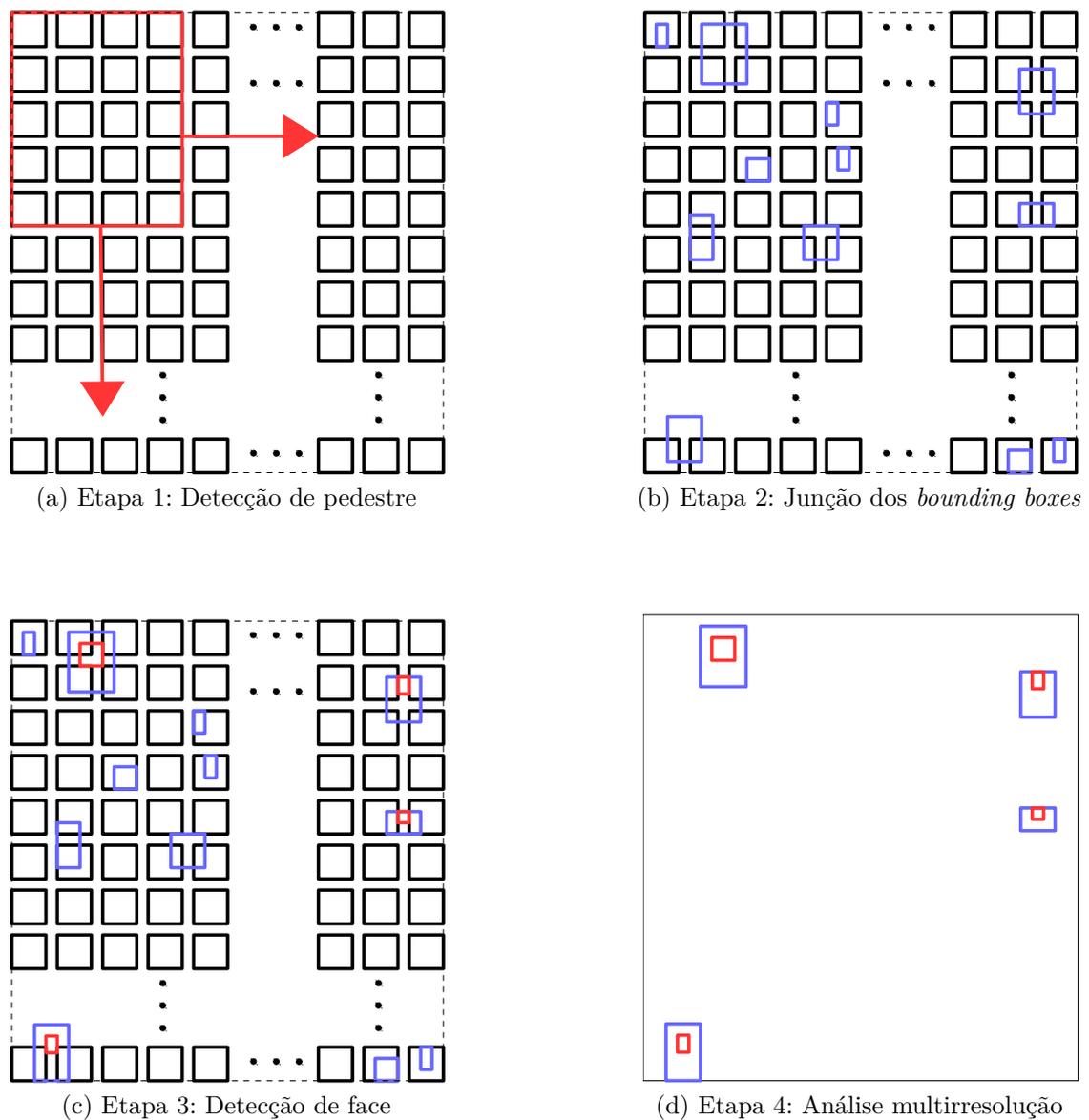


Figura 4.3: Abordagem proposta para manipulação de imagem Gigapixel.

posições randômicas de 1218 imagens no conjunto de treinamento negativo (com ausência de pedestres) do Inria. Já a base de dados da Caltech contém 2.300 pedestres únicos, divididos em três subconjuntos: próximos, a uma distância média e distantes, motivados por uma distribuição de tamanhos na base de dados. Pedestres que se encontram mais próximos possuem altura de 80 pixels ou mais, pedestres com uma distância média, com altura entre 30 e 80 pixels, e pedestres mais distantes com 30 pixels ou menos.

O resultado da aplicação do método desenvolvido nessa etapa é um conjunto de *bounding boxes* com as coordenadas da localização de áreas com possíveis pedestres, como mostrado na Figura 4.3 (b).

4.3.2 Etapa 2: Junção dos *Bounding Boxes*

A abordagem da janela deslizante é de um ladrilho (ou bloco) e, com isso, o processo de escaneamento pode encontrar mais *bounding boxes* do que o necessário, visto que muitas delas estão sobrepostas. Embora seja possível adotar a estratégia de passos maiores do que um bloco, isso poderia causar perda de dados, visto que não é usada nenhuma informação prévia sobre o tamanho e a posição das pessoas.

Dessa forma, com o objetivo de reduzir os *bounding boxes* desnecessários, é realizado um processo de junção de *bounding boxes* através do cálculo da interseção sobre a união (IoU) baseada no critério de supressão não máximo (Dollar et al., 2012; Hosang et al., 2015). Esse processo consiste em calcular a área sobreposta entre pares de *bounding boxes* e suas respectivas razões de proporção. Se a razão IoU é maior do que um valor definido, acontece a junção dos *bounding boxes*.

As áreas são calculadas de acordo com as Equações 4-2 e 4-3, enquanto as proporções com a Equação 4-1, expressas como:

$$P_n = \frac{O}{A_n}, \quad (4-1)$$

$$O = (\max(X) - \min(X+W))(\max(Y) - \min(Y+H)), \quad (4-2)$$

$$A_n = W_n \cdot H_n, \quad (4-3)$$

onde Y , X , W e H são pares do canto superior esquerdo do *bounding box* e seus valores de largura e altura, respectivamente, O e A_n correspondem à área sobreposta e área da n ésima (n) caixa delimitadora, respectivamente, enquanto P_n é a proporção da área entre os *bounding boxes*. Para o estudo em questão foi realizada a junção de dois *bounding boxes* quando $\max(P_n) \geq 0,7$. Este parâmetro foi definido empiricamente. Valores mais baixos do que esse causaram um crescimento artificial do tamanho dos *boundings boxes* e tiveram impacto negativo na abordagem de casamento de pedestre com a face.

4.3.3 Etapa 3: Detecção de Faces

A etapa de detecção de face é mostrada na Figura 4.3(c) e considerou o uso do método de análise de árvore de regressão e classificação (CART) proposto por Lienhart et al. (2003), o qual se apresenta como sendo um ambiente de detecção de face capaz de processar imagens de maneira extremamente rápida, ao mesmo tempo que alcança altas taxas de detecção.

O CART considera características faciais codificadas usadas com um conjunto de classificadores fracos, com base na classificação e análise de árvore de regressão. Neste trabalho, os coeficientes de Haar foram considerados como características

adotadas para a entrada dos classificadores. O modelo CART foi treinado com o conjunto de imagens de testes de faces frontais da *Carnegie Mellon University* (CMU) com imagens de tamanho 20×20 (Lienhart *et al.*, 2003).

4.3.4 Etapa 4: Análise Multirresolução

Nesta etapa, foi avaliada a capacidade de detecção de pedestres e faces em diferentes níveis de resolução dentro dos *bounding boxes* encontrados com o detector ACF (Dollar *et al.*, 2009, 2010a) e os algoritmos CART (Viola & Jones, 2004) para análise multirresolução.

A análise multirresolução foi desenvolvida neste trabalho com as técnicas de Pirâmide Laplaciana de Burt (1981) e de Burt & Adelson (1983a), Vizinho mais Próximo (*Nearest Neighbor*), Interpolação Bilinear e Interpolação Bicúbica (Gonzalez & Woods, 2008).

Na abordagem da Pirâmide Laplaciana, a imagem original é usada para gerar os níveis da pirâmide Gaussiana através da repetição da média local. Os níveis da Pirâmide Laplaciana, então, são computados como as diferenças entre os níveis Gaussianos adjacentes.

Na Interpolação do Vizinho Mais Próximo, o pixel de saída é assinalado como o valor do pixel onde o ponto se encontra. Nenhum outro pixel é considerado. Já na Interpolação Bicúbica, o valor do pixel de saída é uma média ponderada dos pixels em uma vizinhança 4×4 . Na Interpolação Bilinear, por sua vez, o valor do pixel de saída é uma média ponderada dos pixels dentro de uma vizinhança 2×2 .

Durante a análise da imagem, as coordenadas dos *bounding boxes* foram mapeadas entre o mosaico e a imagem Gigapixel inteira, visto que durante o processo de escaneamento elas são mapeadas para a janela deslizante. Isso pode ser visto com a Equação 4-5 e revertido com a Equação 4-4,

$$CR_i = \lfloor BB_i / 256 \rfloor, \quad (4-4)$$

$$BB_i = CR_i \cdot 256 + bb_i, \quad (4-5)$$

$$bb_i = BB_i \% 256, \quad (4-6)$$

onde BB and bb são pares de coordenadas x e y do *bounding box* i na imagem Gigapixel e no mosaico, respectivamente, enquanto CR corresponde ao par de coluna e linha do *bounding box* i no mosaico.

Estudos de medidas antropométricas físicas contribuem para a análise da proporção entre o corpo e a cabeça de uma pessoa em que ambos são centrados no eixo de simetria humano (Preedy, 2012).

Foi desenvolvido um processo de casamento de *bounding boxes* de pedestre/face para determinar se os *bounding boxes* de pedestres e faces estavam desalinhados no eixo-x em uma razão maior do que um fator de tolerância. Alguns exemplos de situações em que o modelo encontra um *bounding box* de pessoa e de face, mas não representa realmente uma pessoa são mostrados na Figura 4.4.

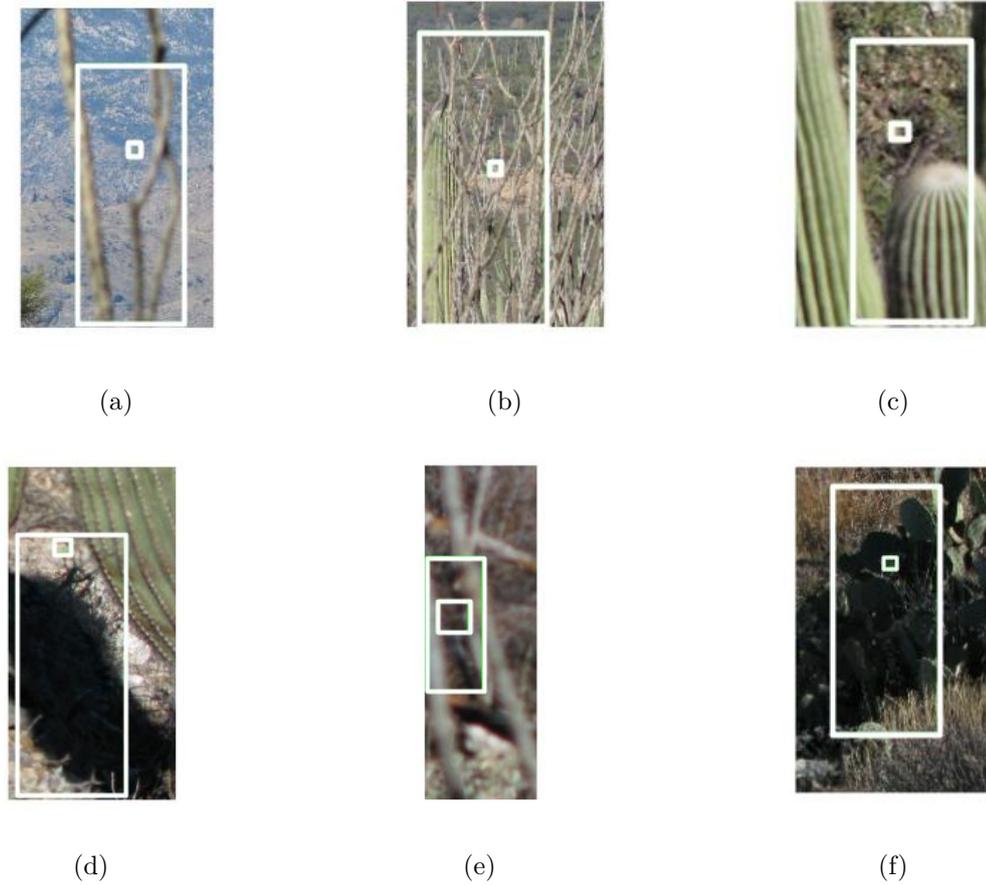


Figura 4.4: Exemplos de falsos positivos obtidos na análise da giga imagem avaliada.

Dessa forma, para eliminar os falsos positivos foi assumido que esses achados não contêm uma pessoa e, dessa maneira, ambos foram descartados com base nas considerações do desalinhamento e no fator de tolerância mencionados. Esse modelo é mostrado na Figura 4.5. Para fazer o casamento do par pedestre-face, a Equação 4-7 foi utilizada:

$$tf = \frac{\max(d_L, d_R)}{\min(d_L, d_R)}, \quad (4-7)$$

tal que

$$\begin{aligned} d_L &= bb_x(face) - bb_x(corpo), \\ d_R &= bb_x(corpo) + bb_{largura}(corpo) - bb_x(face) + bb_{largura}(face), \end{aligned}$$

onde tf é o fator de tolerância para a proporção entre as distâncias entre os cantos esquerdo e direito (d_l, d_r) dos *bounding boxes* da face e do pedestre. Foi definido, empiricamente, que os *bounding boxes* com $tf > 2$ seriam descartados. A Figura 4.6 ilustra os resultados das amostras de todos os passos desenvolvidos (a, b, c, e d) do método implementado.

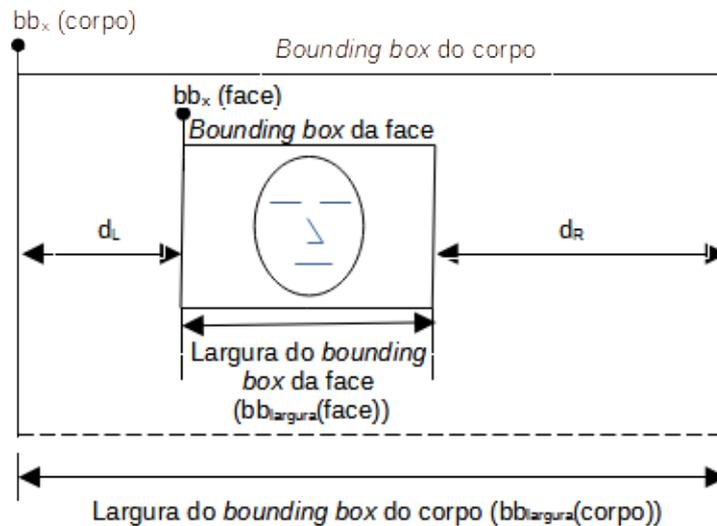


Figura 4.5: Abordagem de casamento dos *bounding boxes* de face e corpo

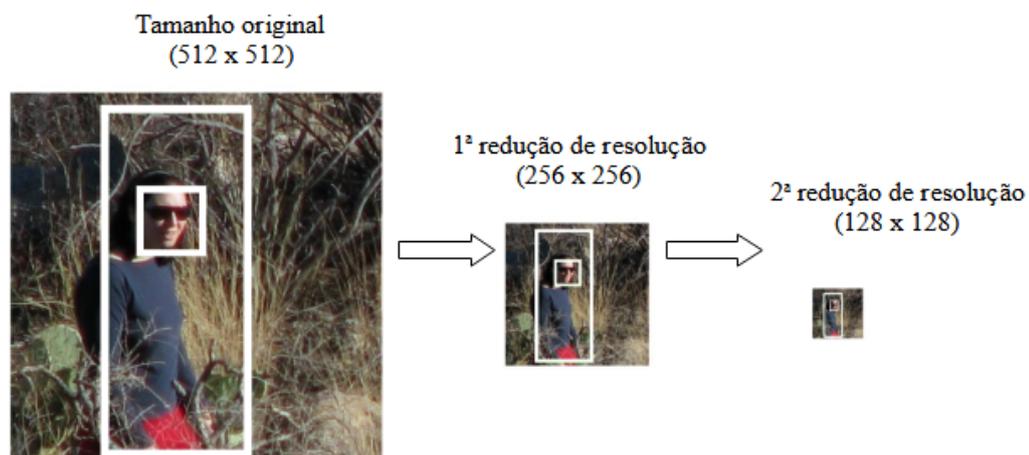


Figura 4.6: Resultados da amostra depois do passo de análise multirresolução.

4.4 Resultados Alcançados

O método apresentado neste capítulo foi aplicado à imagem Gigapixel apresentada na Figura 4.1. A Tabela 4.2 apresenta o número de *bounding boxes*

encontrados, o tempo computacional para encontrá-los, bem como o de redução de resolução e de processamento total para os experimentos com o detector ACF treinado com as bases do Inria e da Caltech, considerando quatro níveis de resolução.

Tabela 4.2: Resultados alcançados com o uso do ACF detector treinado com as bases de dados do Inria e Caltech, considerando a imagem Gigapixel apresentada na Figura 4.1. Algoritmos de redução de resolução utilizados: Bicúbico (BC), Bilinear (BL), Vizinho Mais Próximo (VP) e Pirâmide Laplaciana (PL).

(a) Número de *Bounding Boxes* Encontrados.

Base	Nível	Número de Blocos	Número de <i>Bounding Boxes</i> Encontrados			
			BC	BL	VP	PL
Inria	1/2	15.774	1.754	1.682	1.785	1.621
	1/4	3.960	363	332	343	335
	1/8	1.020	64	60	66	61
	1/16	270	10	6	7	13
Caltech	1/2	15.774	7.134	5.675	7.452	5.321
	1/4	3.960	1.996	1.481	1.812	1.341
	1/8	1.020	496	337	516	315
	1/16	270	92	73	92	66

(b) Tempo para encontrar os *bounding boxes*.

Base	Nível	Número de Blocos	Tempo para Encontrar o <i>bounding box</i> (min'seg ^s)			
			BC	BL	VP	PL
Inria	1/2	15.774	43'57"	42'38"	42'16"	42'21"
	1/4	3.960	8'26"	8'50"	8'36"	8'52"
	1/8	1.020	1'53"	1'50"	1'56"	1'50"
	1/16	270	0'21"	0'21"	0'21"	0'21"
Caltech	1/2	15.774	218'02"	173'34"	224'58"	167'37"
	1/4	3.960	31'48"	27'43"	29'38"	26'35"
	1/8	1.020	5'35"	5'18"	5'38"	5'10"
	1/16	270	0'59"	0'59"	1'02"	0'56"

(c) Tempo para redução da resolução.

Base	Nível	Número de Blocos	Tempo para Redução da Resolução (min'seg ^s)			
			BC	BL	VP	PL
Inria	1/2	15.774	21'57"	20'40"	16'09"	20'17"
	1/4	3.960	4'57"	5'12"	1'40"	1'51"
	1/8	1.020	0'31"	0'20"	0'16"	0'29"
	1/16	270	0'10"	0'05"	0'04"	0'07"
Caltech	1/2	15.774	21'57"	20'40"	16'09"	20'17"
	1/4	3.960	4'57"	5'12"	1'40"	1'51"
	1/8	1.020	0'31"	0'20"	0'16"	0'29"
	1/16	270	0'10"	0'05"	0'04"	0'07"

(d) Tempo total de processamento.

Base	Nível	Número de Blocos	Tempo Total de Processamento (min'seg ^s)			
			BC	BL	VP	PL
Inria	1/2	15.774	65'53"	63'17"	58'25"	62'37"
	1/4	3.960	35'19"	34'41"	26'25"	30'59"
	1/8	1.020	29'16"	28'01"	20'00"	24'25"
	1/16	270	27'54"	26'36"	18'29"	23'02"
Caltech	1/2	15.774	239'59"	194'14"	241'07"	187'54"
	1/4	3.960	58'42"	53'45"	47'27"	48'53"
	1/8	1.020	32'10"	31'40"	23'43"	27'57"
	1/16	270	27'44"	27'26"	19'11"	23'50"

A imagem usada nos experimentos tem 62.487 imagens blocos em seu

tamanho original. Nos experimentos, foram obtidos 7.382 *bounding boxes* depois da aplicação do processo de detecção de pedestres com a base de dados do Inria e 17.274 com a base da Caltech, sendo o tempo total de processamento de 288'23" e 1.124'15", respectivamente, uma vez que não há redução de resolução neste nível. Foram experimentados quatro algoritmos de redução em quatro níveis de resolução, onde o número de *bounding boxes* reduziu aproximadamente a um quinto de um nível para o outro.

Embora o tempo de redução de resolução seja o mesmo com os experimentos das bases de dados do Inria e da Caltech, o algoritmo detector de pedestre se comporta de maneira diferente em cada experimento, mudando tanto o custo computacional do tempo de encontrar os *bounding boxes* bem como o tempo total de processamento. A quantidade de *bounding boxes* apresentados na Tabela 4.2 é o valor obtido depois do processo de junção dos *bounding boxes*, e eles são metade, ou até mesmo um sexto, do total anterior do mencionado processo.

A Figura 4.7 apresenta uma taxa de falha média logarítmica para resumir o desempenho do detector calculado pela média das pessoas perdidas e pelas taxas de falsos positivos espaçadas uniformemente no espaço logarítmico no intervalo de 0 a 10^5 para experimentos com as bases de dados do Inria e Caltech na mudança de resolução (alta taxa de falso positivo significa maior nível de resolução e vice-versa).

É importante ressaltar que a taxa de falsos positivos é diretamente dependente da mudança do nível de resolução, enquanto que a taxa de perda de pessoas é inversamente dependente. Dessa forma, é possível observar que quando a resolução é menor, a taxa de falso positivo também é menor, enquanto a taxa de perda é maior. Entretanto, se a resolução é maior, a taxa de falso positivo também é maior, enquanto a taxa de perda é menor. É possível observar (Figura 4.7 (b)) que os experimentos com a base de dados da Caltech fornecem 10 vezes mais falsos positivos do que os experimentos com a base do Inria (Figura 4.7 (a)) para o mesmo nível de resolução.

Como o número de caminhantes na imagem Gigapixel é de apenas 10, para reduzir a quantidade total de *bounding boxes* encontrados, foi realizada a detecção de face dentro de cada um deles, verificando se havia um casamento entre as distâncias calculadas com a Equação 4-7. No primeiro nível de redução, o número de *bounding boxes* é reduzido para 17 e 89, para as bases do Inria e Caltech, respectivamente, incluindo os 7 caminhantes com faces detectáveis. As outras 3 imagens de caminhantes não são detectáveis na abordagem apresentada, pois estão de costas ou com faces oclusas.

Embora a base de dados da Caltech forneça imagens de pedestres menores do que as do Inria, o que supostamente poderia fornecer melhores resultados de detecção de pedestres, tal fato não apresenta diferença para o algoritmo de detecção, visto

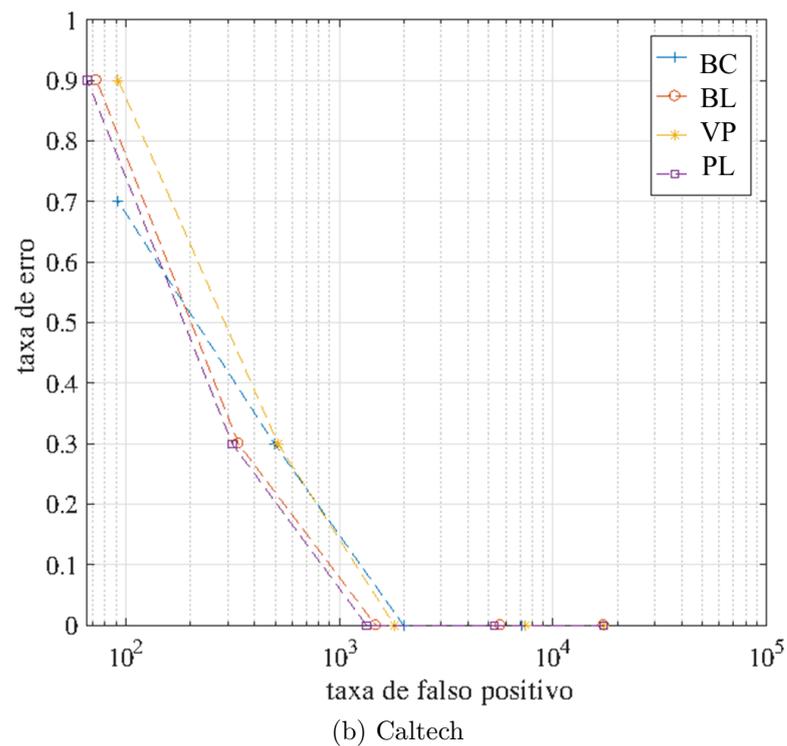
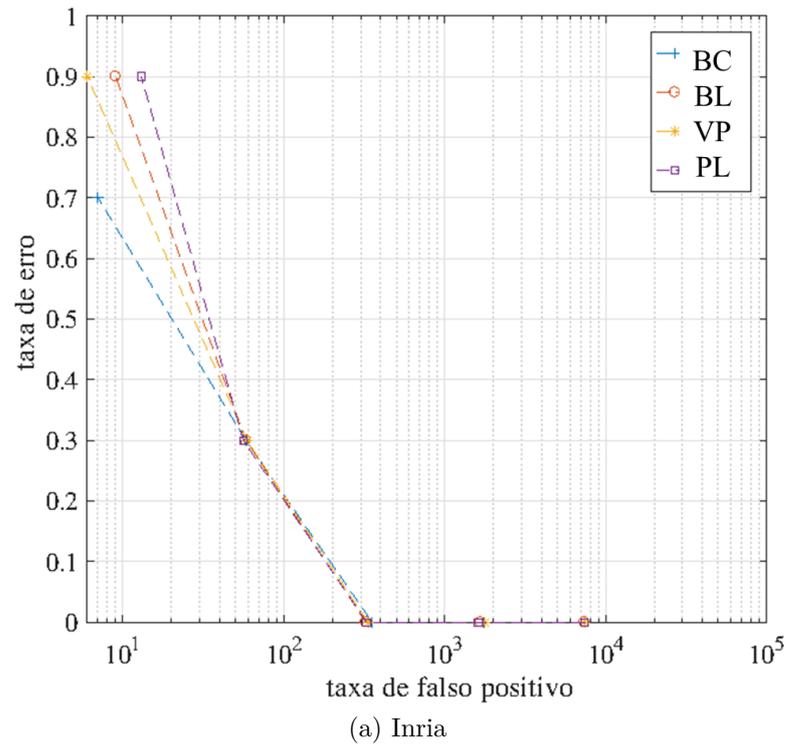


Figura 4.7: Falsos positivos *versus* taxa de perda para diferentes níveis de resolução com o ACF treinado com as bases do Inria e Caltech.

que todos os pedestres foram encontrados em ambos os experimentos. Entretanto, os experimentos com a Caltech levaram muito mais tempo e encontraram muito mais *bounding boxes* sem utilidade.

A Tabela 4.3 apresenta o nível de resolução mais baixo que a abordagem desenvolvida foi capaz de fazer o casamento de pessoas e faces. Os excertos listados nessa tabela são os mesmos da Figura 4.2. Na fase de detecção de pessoas, o algoritmo BC detectou mais pessoas no nível $1/16$ do que os outros algoritmos. Entretanto, todos os algoritmos detectaram pessoas entre os níveis $1/4$ e $1/16$, a não ser o excerto (h) para PL, que foi detectado apenas no primeiro nível de redução. Na etapa de detecção de face, os excertos (d) e (e) alcançaram o nível $1/2$ em todos os algoritmos, exceto BC para (d), que atingiu o nível $1/4$. Os excertos (a), (b), (c) e (i) tiveram as faces detectadas somente no nível de resolução original.

Como é realizada uma correspondência entre face e corpo com o objetivo de filtrar falsos positivos, o nível mais baixo a ser considerado é o nível de resolução da face, pois, quando uma face não é encontrada, não há correspondência e é assumido que não existe pessoa dentro do *bounding box*. Entretanto, isso implica em um crescimento da taxa de falsos negativos. Além disso, como esperado, independentemente do método de redução, a abordagem proposta não detectou faces de caminhantes de costas ou com faces oclusas. Esse fato é enfatizado como zeros na Tabela 4.3.

Tabela 4.3: Nível mais baixo de resolução para o casamento da face e da pessoa.

Excerto	Detecção de Pessoa				Detecção de Face			
	BC	BL	VP	PL	BC	BL	VP	PL
(a)	$1/4$	$1/4$	$1/4$	$1/4$	$1/1$	$1/1$	$1/1$	$1/1$
(b)	$1/8$	$1/8$	$1/8$	$1/4$	$1/1$	$1/1$	$1/1$	$1/1$
(c)	$1/16$	$1/16$	$1/8$	$1/8$	$1/1$	$1/1$	$1/1$	$1/1$
(d)	$1/8$	$1/8$	$1/8$	$1/8$	$1/4$	$1/2$	$1/2$	$1/2$
(e)	$1/4$	$1/4$	$1/8$	$1/4$	$1/2$	$1/2$	$1/2$	$1/2$
(f)	$1/16$	$1/8$	$1/8$	$1/8$	0	0	0	0
(g)	$1/8$	$1/8$	$1/8$	$1/8$	0	0	0	0
(h)	$1/16$	$1/8$	$1/16$	$1/2$	0	0	0	0
(i)	$1/4$	$1/4$	$1/4$	$1/8$	$1/1$	$1/1$	$1/1$	$1/1$

Além dos testes com janelas deslizantes 4×5 , foram realizados outros experimentos com diferentes tamanhos de janelas, os quais demonstraram que os mesmos podem influenciar nos resultados. Em alguns casos, os *bounding boxes* se encaixaram muito bem nas pessoas, enquanto em outros testes eles se encaixaram apenas em partes do corpo. Além disso, parece que o algoritmo do detector ACF é sensível ao conteúdo do cenário.

4.5 Considerações Finais do Capítulo

Este capítulo apresentou um estudo e um método para lidar com a detecção de pedestres em imagens Gigapixel com base na redução de suas dimensões. O

número de *bounding boxes* e o custo computacional associados foram apresentados e analisados.

Foi possível observar que o tempo para redução de resolução do método do Vizinho mais Próximo é menor quando comparado aos outros métodos, sendo possível verificar que o mesmo padrão ocorre no tempo de processamento total dos experimentos, exceto no nível $1/2$ para experimentos com a base de dados da Caltech.

Também ficou claro que algoritmos de redução de resolução mostraram um comportamento similar no estágio de detecção de face. Entretanto, os resultados variaram no estágio anterior de detecção de pessoas. Por outro lado, a maioria dos caminhantes foi detectada no nível de redução $1/4$ ou superior. O caminhante (h), mostrado na Tabela 4.3 foi detectado apenas até o nível de redução $1/2$ com o algoritmo da Pirâmide Laplaciana.

Neste capítulo também ficou evidente que a busca por pessoas, sem conhecimento prévio da imagem Gigapixel, é uma tarefa desafiadora, pois não se sabe qual nível de resolução pode ser suficiente para os algoritmos. Além disso, não há garantia alguma de que exista uma pessoa em um determinado local na imagem, por exemplo.

O próximo capítulo discutirá um exemplo proposto para auxiliar o trabalho de um observador humano, que, ao olhar imagens maiores com multidões, busca por uma pessoa específica. A ideia desenvolvida é reduzir o espaço de busca de imagens para um *ranking* de 10 imagens possíveis relacionadas à pessoa sendo buscada.

Abordagem de Varredura em Imagens Gigapixel com Agrupamento Denso de Pessoas

Sumário

5.1	Introdução	64
5.2	Imagens Gigapixel de Multidões	65
5.3	Descrição da Base de Imagens	66
5.3.1	Base de Dados da Imagem Gigapixel	67
5.3.2	Base de Imagens Construída Manualmente	69
5.4	Método Proposto	70
5.4.1	Etapa 1: Detecção de Faces na Imagem Gigapixel	71
5.4.2	Etapa 2: Emparelhamento Entre a Base de Imagens Construída Manualmente e as Faces da Imagem Gigapixel	72
5.5	Resultados Alcançados	74
5.6	Considerações Finais do Capítulo	78

5.1 Introdução

Este capítulo descreve uma abordagem para auxiliar um observador humano a encontrar pessoas em imagens maiores que contenham multidões, através da redução do espaço de busca de várias imagens relacionadas a uma pessoa específica para um *ranking* de 10 imagens da busca da pessoa mencionada. O desenvolvimento desse tipo de abordagem é desafiador, visto que é necessário realizar um processo de varredura na imagem Gigapixel em busca de faces em multidões.

No modelo desenvolvido neste capítulo, várias faces existentes na imagem Gigapixel são coletadas e, em seguida, uma pessoa específica é procurada utilizando-se imagens da mesma em três poses distintas: de frente, com a face virada para a esquerda e com a face virada para a direita, que, aqui, para efeitos de simplificação, são denominadas de faces de frente, à esquerda e à direita. Destaca-se que o método e resultados apresentados neste capítulo estão também detalhados no trabalho ([Ferreira et al., 2020b](#)) produzido durante esta pesquisa.

Também é mostrado como a base de dados foi construída, o método proposto envolvendo a detecção de faces na imagem Gigapixel e o casamento entre elas e as faces da base de dados contendo faces de frente, à esquerda e à direita, além da apresentação dos resultados alcançados. Assim sendo, a seção seguinte descreve imagens Gigapixel de multidões e suas respectivas características.

5.2 Imagens Gigapixel de Multidões

Uma multidão é uma situação comum em que pessoas se aglomeram em torno de um evento por algum motivo. Por exemplo, as pessoas geralmente se reúnem em lugares específicos para eventos comuns, como shows, teatros, estádios, *shoppings*, aeroportos, dentre outros. A título de exemplificação, a Autoridade Portuária de Nova York e Nova Jersey ([Port Authority of New York and New Jersey, 2018](#)) relata que aeroportos como o Aeroporto Internacional Hartsfield-Jackson Atlanta teve cerca de 103.902.992 passageiros apenas no ano de 2017, o que representa aproximadamente 284.665 pessoas por dia ou 11.861 por hora.

Quando há centenas ou até mesmo milhares de pessoas juntas, podem haver situações em que seja importante e necessário um monitoramento de segurança, bem como detectar e identificar pessoas na cena, como uma criança desaparecida ou um suspeito de crime. Desta forma, os equipamentos podem capturar imagens ou vídeos de cenas tipicamente povoadas por um grande número de pessoas e, em seguida, deve ser realizado um processo de identificação das pessoas no local a fim de se viabilizar, por exemplo, a localização de um indivíduo suspeito na cena. Identificar várias faces em uma multidão não é uma tarefa simples, pois requer separação prévia de rostos e posterior identificação e classificação de um banco de dados facial já treinado.

Um exemplo de imagem considerando multidões é apresentado na Figura 5.1, que representa uma imagem Gigapixel do discurso de posse do Presidente Donald Trump como Presidente dos Estados Unidos da América, no dia 20 de janeiro do ano de 2017. Na imagem mencionada é possível perceber a presença de várias autoridades, integrantes da família, líderes religiosos, além de inúmeros cidadãos que estavam presentes para assistir à cerimônia.

Em se tratando de imagens dessa natureza, é importante salientar que alguns estudos mostram que o uso de cenas visuais que aproximam o detalhe e a complexidade das cenas naturais ajuda a compreender as propriedades de uma cena visual complexa e influencia sua compreensão completa, como mostrado por [Clarke et al. \(2013\)](#). Assim, os seres humanos têm grandes dificuldades em encontrar padrões em imagens sobrecarregadas de informações, densas, como imagens de multidões. Considerando isso, no estudo apresentado neste capítulo foi desenvolvido um modelo de auxílio visual para um observador humano trabalhar com imagens Gigapixel de multidões, através da redução do espaço de busca de várias imagens para um *ranking* de dez imagens relacionadas a uma pessoa específica. A abordagem proposta é capaz de separar faces nessas imagens, bem como a busca e reconhecimento de pessoas por meio de um conjunto de dados de teste, considerando um conjunto de três poses diferentes para cada pessoa a ser encontrada e identificada na imagem Gigapixel. Dessa forma, o modelo contribui para uma melhor visualização e redução do espaço de busca para a análise de um observador.

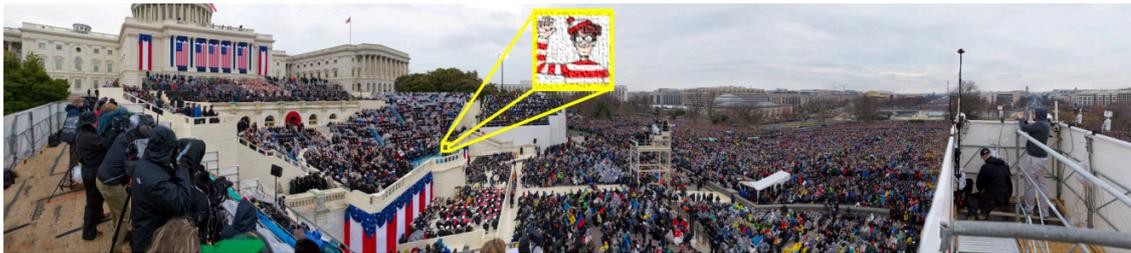


Figura 5.1: Onde está o Wally (*Where's Wally?*)¹- Conceito do estudo proposto (Adaptado de CNN).

A seção seguinte traz, então, a descrição das bases de imagens utilizadas neste estudo.

5.3 Descrição da Base de Imagens

No desenvolvimento deste estudo de caso, foram usados dois conjuntos de dados para experimentos combinados. O primeiro deles compreende uma imagem Gigapixel da posse de Trump, mencionada na Seção 5.2. Já o segundo, é um conjunto de dados artesanal de diferentes poses dos convidados especiais de posse de Trump, em eventos diversos, sendo os mesmos apresentados na Subseção 5.3.2.

¹Os livros da série “Where’s Wally” são conhecidos como “Where’s Waldo” na América do Norte.

5.3.1 Base de Dados da Imagem Gigapixel

A imagem Gigapixel avaliada nesta abordagem chama-se *The Inauguration of Donald Trump*, produzida pela CNN².

Essa imagem foi gerada em Washington DC em 20 de janeiro de 2017 às 15h GMT-2, com cerca de 300.000 a 600.000 pessoas participando da cerimônia pública de posse. Embora não haja informações disponíveis sobre o equipamento usado para capturar a cena, em um fórum de fotografia do StackExchange ([StackExchange, 2017 \(acessado em 15 de Março de 2022\)](#)), há uma discussão especulando que ela foi gerada com um Gigapan Epic Pro. Entretanto, mais detalhes sobre como a CNN capturou a imagem da posse de Trump podem ser encontrados em [Arthur \(2017 \(acessado em 3 de Agosto de 2020\)\)](#). A imagem é armazenada em uma representação de face de cubo, onde cada face de cubo significa uma direção de mira da câmera. A imagem foi extraída da URL www.fanpic.co com um *script* Python, via URL apresentado da seguinte forma:

```
http://europe.tiles.fanpic.co/749-2017-cnn/  
mres_{s}/l{l}/v/l{l}_{s}_{v}_{h}.jpg
```

onde $\{s\}$ é a face do cubo composta por f , b , u , d , l e r , que são frente, trás, cima, baixo, esquerda e direita, respectivamente, l é o nível de resolução, onde 1 é a menor e 7 é a maior resolução, e v e h são as posições vertical e horizontal da câmera, respectivamente, variando de 1 a 125 na resolução mais alta.

Para realizar esse estudo de caso, foram baixadas as imagens de todas as faces do cubo no nível 7, que é a resolução mais alta, fornecendo blocos RGB com 512×512 pixels e um total de 93.750 blocos. No entanto, foi utilizada uma pequena porção da imagem Gigapixel que corresponde ao local ao redor do palco do discurso do Presidente, onde se sentam autoridades, como juízes da Suprema Corte, Senadores e Deputados, ex-Presidentes e Primeiras Damas, líderes religiosos, familiares e amigos mais próximos, dentre outros. Foram consideradas as colunas de 92 a 124 na face esquerda do cubo, colunas de 00 a 30 na face frontal do cubo e linhas de 73 a 90 em ambas as faces do cubo, resultando em um total de 1.152 blocos.

Os blocos são projetados para evitar o efeito esférico do processo de aquisição da imagem e já conectadas. No entanto, não há informações sobre o *software* e algoritmos usados para isso. Além disso, é possível observar que, em vários casos, ocorre o achatamento da imagem, o que pode trazer alguns efeitos de distorção radial do rosto das pessoas e dificultar o processo de reconhecimento facial.

²Disponível em <https://edition.cnn.com/interactive/2017/01/politics/trump-inauguration-gigapixel>. (Acesso feito em 01/09/2022).

Também é importante mencionar que, na construção dessa imagem em Gigapixel, diversos problemas dificultam o processo de identificação e reconhecimento das pessoas, devido a fatores como, por exemplo, sua posição na cena. Várias pessoas aparecem por trás de outras, várias de lado e diversas com o rosto escondido. Assim, surgem vários problemas como a variância da postura das pessoas e a oclusão, que são grandes desafios a serem resolvidos.

A estratégia utilizada para percorrer a imagem Gigapixel é detalhada na Seção 5.4, com a obtenção de 432 imagens de faces com a abordagem desenvolvida após o descarte de algumas imagens detectadas sem faces. A Figura 5.2 mostra essas faces e a Tabela 5.1 traz algumas estatísticas para essa base de dados de faces.

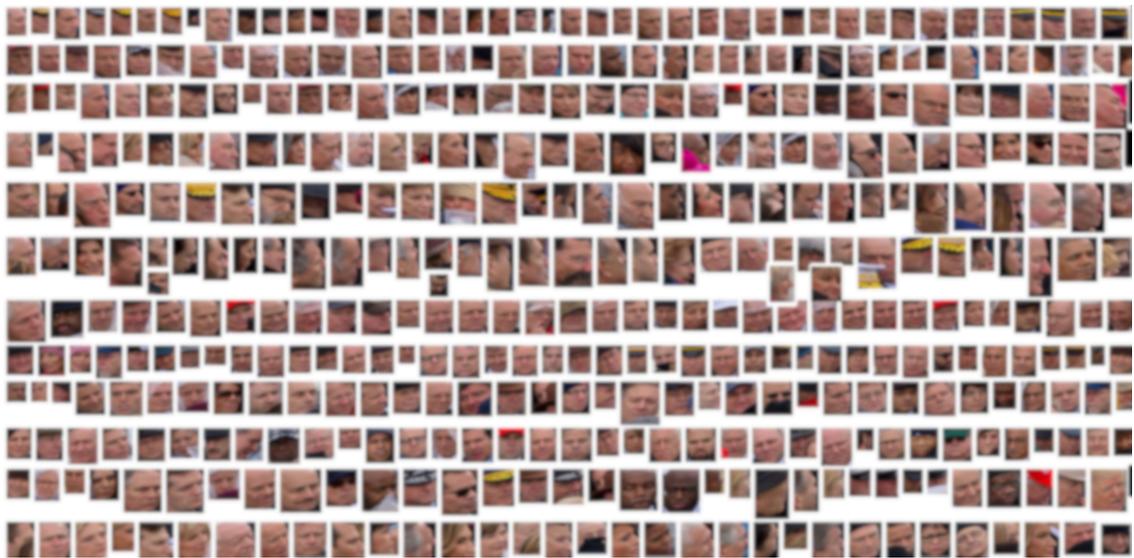


Figura 5.2: Amostras de imagem Gigapixel para construir o conjunto de dados de faces.

Tabela 5.1: Estatísticas para a base de imagens Gigapixel.

Imagens	Total	Posição da Face				Oclusão por		Nenhuma Oclusão
		Frontal	Direita	Esquerda	De costas	Acessórios	Outras Faces	
Todas	432	111	284	11	26	123	41	268
Selecionadas	42	12	25	1	4	7	9	26

A Tabela 5.1 agrupa faces com base em dois critérios principais: a posição da face e a visibilidade da face. Para o primeiro caso, observamos quatro posições frequentes nestas imagens: frente, direita, esquerda e de costas. No segundo caso, observamos imagens com oclusão de acessórios, imagens com oclusão de face e sem oclusão. Usando parte da imagem Gigapixel, o modelo encontrou 432 faces. Desse total, e considerando o primeiro critério, foram encontradas 111 faces na posição frontal, 284 faces voltadas para a direita, 11 faces para a esquerda e 26 de costas. Há diversas pessoas presentes na imagem que não são autoridades ou pessoas públicas

conhecidas. Assim, foram identificadas 42 pessoas (autoridades, líderes religiosos, familiares, entre outros). Dessa forma, considerando esse total de pessoas conhecidas, verificou-se que 12 pessoas aparecem com o rosto na posição frontal, 25 com o rosto virado para a direita, 1 com o rosto virado para a esquerda e 4 de costas. Ao considerar o segundo critério, a preocupação se concentrou em verificar se as faces estão ou não oclusas. A maioria das faces não estava oclusa (26) e 16 faces estavam oclusas por acessórios (7) ou por outra face (9).

A Figura 5.3 mostra alguns dos problemas encontrados e que dificultam o reconhecimento das faces extraídas da imagem Gigapixel. Já na Figura 5.3(a), podemos ver uma face de costas. Nas Figuras de 5.3(b) até 5.3(f) podemos observar que as faces estão oclusas por diversos acessórios, como chapéu, óculos, capa de chuva e mão na face. Nas Figuras 5.3(g), (h) e (i), temos oclusão por outras faces. Por fim, na Figura 5.3(j), temos um problema complexo causado pela distorção do processo de combinação dos blocos de imagens que formam a imagem Gigapixel.

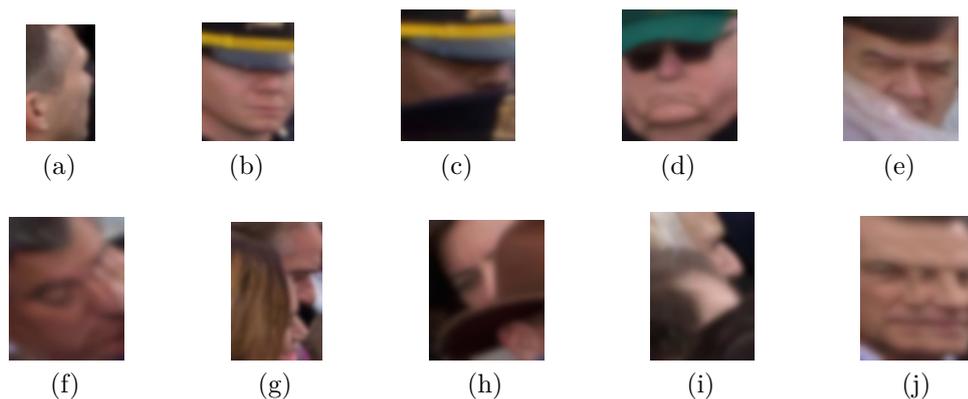


Figura 5.3: Amostras da base de imagens Gigapixel: (a) de costas; (b) até (f) oclusão por acessório; (g) até (i) oclusão por outras faces e (j) distorção do processo de combinação de imagens.

5.3.2 Base de Imagens Construída Manualmente

Essa base de dados foi criada a partir de imagens baixadas de pessoas que estiveram na posse do Presidente Donald Trump, como, por exemplo, membros da Suprema Corte dos EUA, ex-presidentes e primeiras-damas, o próprio Presidente Donald Trump e seus familiares, algumas celebridades religiosas, deputados e senadores.

Foram utilizadas imagens das poses frontal, esquerda e direita da face de cada indivíduo, em diferentes situações e datas, conforme mostrado na Figura 5.4, sendo (a) a imagem com a face para a esquerda, (b) a com a face frontal e (c) com

a face para a direita. Na base de dados construída manualmente, há 42 indivíduos diferentes, com 3 imagens de cada, resultando em um total de 126 imagens.

Foi realizada uma pesquisa em vários sites de uma lista completa de todas as autoridades, familiares, amigos, líderes religiosos, juízes da Suprema Corte, representantes do presidente da Câmara, entre outros, que estiveram presentes na posse do Presidente Donald Trump. Como não foi encontrado este documento completo, foram realizadas buscas em diversos sites, como BBC, Wikipedia e Suprema Corte dos Estados Unidos, para encontrar alguns nomes de pessoas convidadas e que compareceram ao evento.



Figura 5.4: Exemplos das faces do Obama na base de dados construída manualmente.

Além disso, como a posse do Presidente Trump não contou apenas com a presença de autoridades oficiais, foi difícil encontrar uma base pública de imagens para coletar imagens de todos os presentes na cerimônia. Conseguiu-se encontrar um conjunto de imagens de membros do Congresso Americano³, que possui apenas em imagens na pose frontal de cada membro. No entanto, na imagem Gigapixel, várias pessoas aparecem com a face escondida ou em pose lateral, devido ao posicionamento do equipamento que captou as imagens. Assim, a opção foi construir uma base de dados própria, considerando os critérios das três posições: frente, direita e esquerda, bem como a questão dos direitos autorais das imagens. A construção do conjunto de dados foi uma tarefa trabalhosa, pois nem todas as imagens disponíveis na Internet estão disponíveis para uso público.

5.4 Método Proposto

O modelo proposto foi desenvolvido em duas etapas, ambas usando a biblioteca DLIB⁴ para detecção e correspondência de faces. A primeira está relacionada à detecção de todas as faces existentes em uma determinada região

³Repositório GitHub – Contacting Congress: <https://github.com/unitedstates/contact-congress/>

⁴DLIB é uma biblioteca para visão computacional e aprendizagem de máquina.

de interesse na imagem da posse do Presidente Trump. Já a segunda, está relacionada ao uso de 126 imagens da base de dados construída com imagens de pessoas conhecidas em três poses diferentes e a tentativa de localizar essas pessoas nas faces detectadas na imagem Gigapixel. Nas Subseções 5.4.1 e 5.4.2 serão descritos detalhes da abordagem.

5.4.1 Etapa 1: Detecção de Faces na Imagem Gigapixel

Nesta etapa, ocorreu, primeiramente, a seleção de uma pequena parte da imagem Gigapixel que corresponde à localização ao redor do palco onde o presidente discursa. A Figura 5.5(a) mostra uma imagem Gigapixel inteira e a Figura 5.5(b) mostra a parte da imagem Gigapixel que foi considerada neste trabalho. Assim, um modelo foi construído para rolar uma janela deslizante através da imagem, com 2 colunas por 2 linhas de imagens de blocos com 512×512 pixels RGB cada, resultando em uma janela deslizante de 1024×1024 pixels, como visto na Figura 5.5(c). Essa medida foi usada considerando o tamanho médio do rosto na imagem Gigapixel.

Em seguida, houve a detecção de face escaneando a parte da imagem Gigapixel com a janela deslizante. Como o passo adotado para a abordagem da janela deslizante é de 1 bloco, o processo de digitalização pode encontrar mais *bounding boxes* do que o necessário, pois muitas delas estão sobrepostas. Embora fosse possível adotar passos maiores que 1 bloco, isso poderia causar perda de dados, pois não foi utilizada nenhuma informação prévia sobre as faces nas imagens. Assim, para reduzir *bounding boxes* desnecessários e a redundância de faces, foi realizado um processo de junção através do cálculo da interseção sobre união (IoU) com base em um critério de supressão não-máximo e foi adotada uma abordagem semelhante à desenvolvida por [Ferreira et al. \(2018\)](#) e [Ferreira et al. \(2020a\)](#) e explicada no Capítulo 4. Se a razão IoU for maior que um valor definido, ocorre a junção dos *bounding boxes*. Neste estudo foi escolhido um valor limite de 0,9 para a fusão.

Para encontrar uma face em cada *bounding box* foram experimentadas duas abordagens disponíveis na biblioteca DLIB. A primeira envolvendo o uso do Histograma de Gradientes Orientados (Histogram of Oriented Gradients – HOG). As características extraídas são as distribuições (histogramas) das direções dos gradientes da imagem. Gradientes, em geral, são grandes em torno das bordas e cantos, permitindo, assim, a detecção dessas regiões. Em seguida, os histogramas são colocados em um vetor de características que é submetido a um algoritmo de classificação, uma *Support Vector Machine* (SVM) neste caso. A segunda abordagem utiliza uma Rede Neural Convolutiva (*Convolutional Neural Network* – CNN). Ambas abordagens apresentaram resultados bastante similares, e por questão de tempo de execução, foi feita a opção do uso da abordagem de HOG com SVM.

Assim, abaixo é apresentada uma descrição resumida das principais etapas realizadas para a utilização da HOG e SVM:

- Pré-processamento das imagens de entrada com redimensionamento das mesmas;
- Cálculo dos gradientes horizontal e vertical da imagem, removendo informações não-essenciais;
- Divisão da imagem em células com o objetivo de alcançar uma representação mais compacta e tornar o HOG mais robusto a ruídos. O HOG é calculado para cada uma das células. Com o objetivo de estimar a direção de um gradiente dentro de uma região, faz-se a construção de um histograma entre os 64 valores das direções de gradientes e sua magnitude dentro de cada região;
- Normalização do bloco, que pode ser aplicado na imagem para torná-la invariante à iluminação, por exemplo. Nessa etapa, tem-se como resultado final um vetor de características que pode ser usado para treinar um classificador SVM.

Para encontrar o *bounding box* com cada face foi usado o algoritmo de detecção da associação da HOG e a SVM, disponível na biblioteca DLIB. Esse modelo considera a pontuação de quantos pontos de gradientes existentes em cada direção (para cima, para a direita, etc). Nesta abordagem utiliza-se a Estimativa de Ponto de Referência Facial, proposto por [Kazemi & Sullivan \(2014\)](#), cuja ideia principal é calcular 68 pontos de referências em faces. Em seguida, uma SVM é treinada para encontrar esses 68 pontos específicos. Este processo é realizado nas etapas ilustradas nas Figuras 5.5(c) e 5.5(d). A Figura 5.5(e) mostra uma amostra de 432 imagens de faces detectadas em uma pequena porção da imagem Gigapixel que corresponde à localização ao redor do palco do discurso do Presidente, onde ex-presidentes, ex-primeiras-damas, senadores, entre outros convidados, ficam na cerimônia.

5.4.2 Etapa 2: Emparelhamento Entre a Base de Imagens Construída Manualmente e as Faces da Imagem Gigapixel

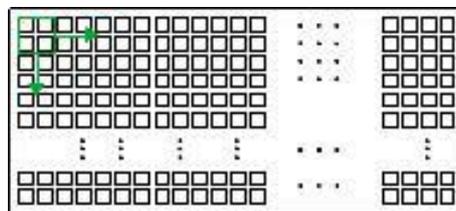
Nesta etapa, foi utilizado um conjunto de imagens de pessoas públicas que incluem autoridades e pessoas conhecidas, que estiveram presentes na cerimônia, considerando 3 poses de cada uma (frente, direita e esquerda). Então, foi construído o conjunto de dados de imagem mostrado na Seção 5.3.2. Semelhante à etapa anterior, também foi realizada uma etapa de detecção de face nas imagens da base de dados



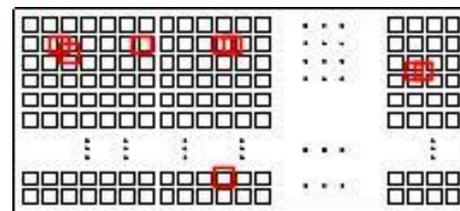
(a) Imagem Gigapixel



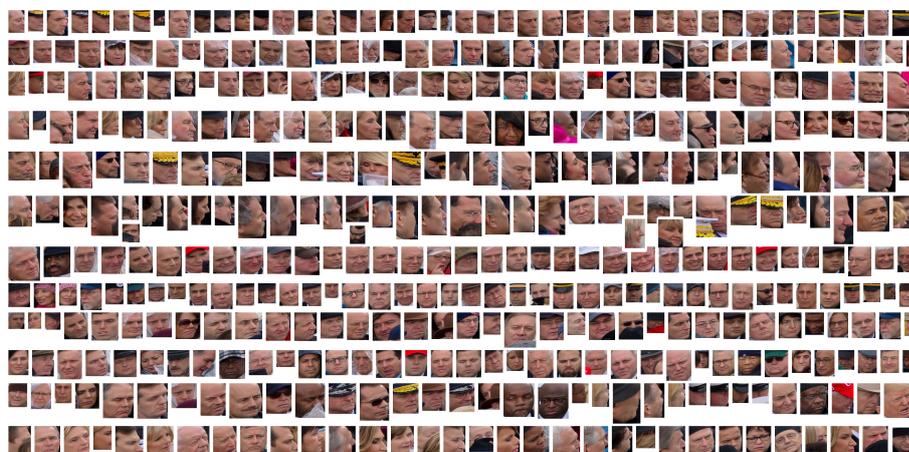
(b) Área de autoridades selecionada



(c) Detecção de face



(d) Junção das faces



(e) Base de dados final

Figura 5.5: Passos da detecção de faces na imagem Gigapixel.

manual através do DLIB, conforme mostrado na Figura 5.6a), e foi extraído um vetor de 128 características de marcação para cada uma. Para realizar o casamento de faces, foi calculada a distância euclidiana (Equação 5-1) entre o vetor de características de cada imagem a ser testada e todos os vetores de características de imagens de face extraídas de uma pequena porção da imagem Gigapixel considerada, conforme evidenciado na Figura 5.6c). Por fim, foi construído um *ranking* de distância usando a Equação 5-2 para ser usado como medida para concluir se uma face foi reconhecida corretamente ou não.

$$Distância_{i,p,j} = \sqrt{\left(\sum_{k=1}^n (fpessoa_i^p(k) - fgiga_j(k))\right)^2}, \quad (5-1)$$

onde $fpessoa_i$ e $fgiga_j$ são os conjuntos de características i e j na base de dados manual e da imagem Gigapixel, respectivamente, p é uma pose da pessoa (esquerda, direita, frontal), k é uma característica, n é a quantidade total de características ($n = 128$), e $Distância_{i,p,j}$ é a Distância Euclidiana entre as características.

$$Rank_i = \min(Distância_{i,j}), \quad (5-2)$$

onde $Rank_i$ é a menor distância da pessoa i em relação às imagens da base de dados Gigapixel.

5.5 Resultados Alcançados

Os experimentos foram realizados para encontrar pessoas na multidão considerando apenas o entorno da área em que o Presidente fez seu discurso, que é reservado ao novo presidente, vice-presidente e seus familiares e amigos, ex-presidentes, primeiras-damas, juízes da Suprema Corte, presidente da Câmara dos Deputados, líder partidário, autoridades especiais, convidados, entre outros.

Cada pessoa foi procurada usando uma pose frontal, esquerda e direita. No entanto, devido à posição da câmera, a pose da maioria das pessoas considerou face direita. Como o objetivo não é encontrar uma correspondência exata, mas, sim, reduzir o esforço de um observador, foram consideradas as 10 primeiras faces com base em um *ranking* de medida de distância. Os resultados são apresentados na Tabela 5.2, que resume os resultados por total, por pose de face e também agrupados por oclusão.

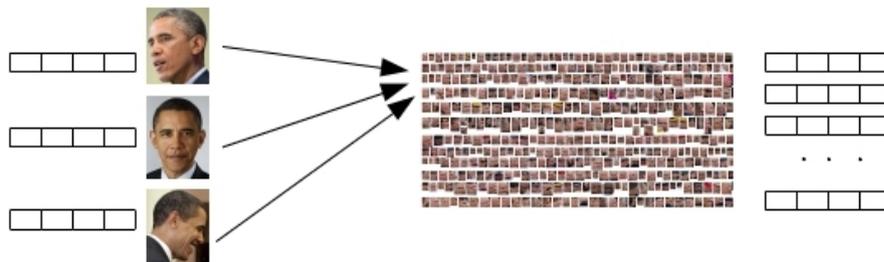
Pode-se observar através dos resultados que 69% foram encontrados, enquanto 31% foram considerados não encontrados. Embora o problema mais expressivo seja com a posição de costas, eles possuem apenas 4 pessoas e, também, a pose de costas



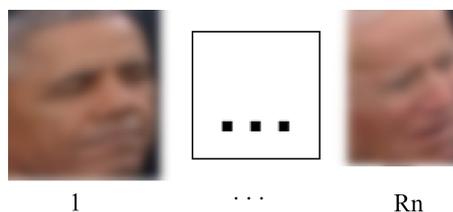
(a) Amostras



(b) Faces detectadas



(c) Emparelhamento de face



(d) Ranking

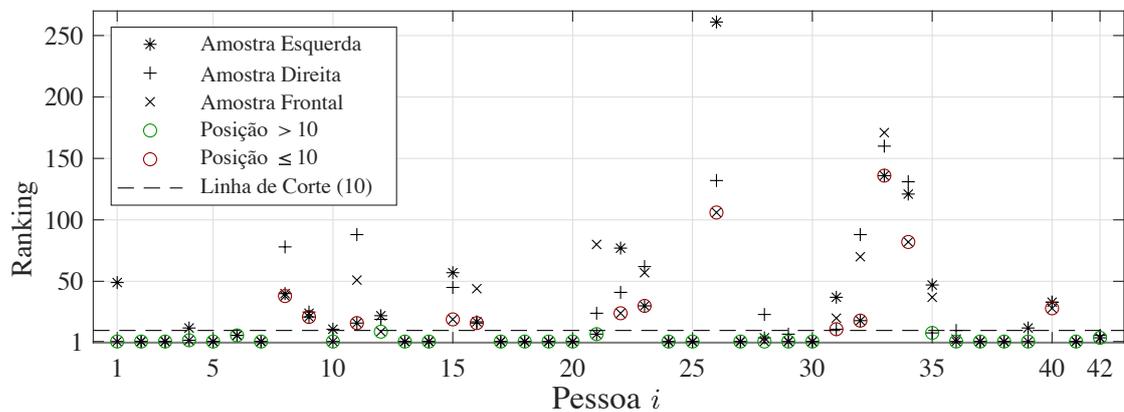
Figura 5.6: Emparelhamento entre a base de dados construída manualmente e as imagens de face da imagem Gigapixel.

Tabela 5.2: Resultado pelo *ranking* mínimo.

Ranking	<i>Ranking</i> por Pose de Face				Oclusão por Acessórios	Oclusão por Outras Faces	Nenhuma Oclusão	
	Total	Frontal	Direita	Esquerda				
≤ 10	69%	83%	64%	100%	25%	57%	44%	92%
> 10	31%	17%	36%	0%	75%	43%	56%	8%

não fornece pontos de referência suficientes para serem emparelhados, mas uma pessoa (25%) foi encontrada.

O segundo resultado mais expressivo é a pose da face direita, o qual contempla o maior grupo, por causa da audiência e posição da câmera. É possível observar que o motivo mais comum para não encontrar a face de alguém é a oclusão por acessório ou por outra face, pois apenas 8% sem oclusão não são encontrados. A Figura 5.7 ilustra os resultados experimentais.

Figura 5.7: *Ranking* de pessoas encontradas nos experimentos.

A pose da pessoa ranqueada abaixo e acima do top-10 (Linha de Corte de Classificação) é representada como círculos verdes e vermelhos, respectivamente. É possível observar que, embora haja um grupo de pessoas que não se classificou entre os dez primeiros, muitos deles estavam bem próximos. Além disso, muitas pessoas têm todas as poses classificadas em posições próximas. É importante notar que, enquanto $\approx 60\%$ das faces na base de dados Gigapixel são poses para a direita, apenas $\approx 41\%$ das faces emparelhadas são de fato poses para a direita, enquanto $\approx 33\%$ e $\approx 26\%$ emparelhadas são poses para esquerda e frontal, respectivamente. Isso mostra que poses diferentes são úteis na base de dados, mesmo que uma pose específica seja dominante na multidão. A Figura 5.8 ilustra um histograma do *ranking* mínimo por pessoa.

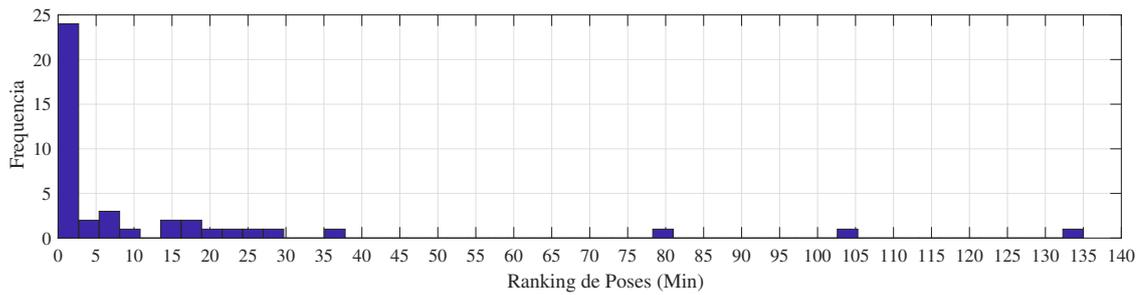


Figura 5.8: Histograma de *ranking* mínimo de poses de pessoas.

A maioria das imagens está na classificação zero do *ranking* e um *ranking* mínimo de 10 é capaz de obter um *cluster* de imagens. Se o *ranking* for aumentado para 12, 13, 14, 15 e assim por diante, é possível obter mais uma ou duas imagens em cada incremento. No entanto, isso aumentaria o número de imagens a serem inspecionadas por um observador. Também fica claro que algumas imagens estão longe do limite definido (10), por exemplo, existem imagens com um *ranking* mínimo entre 35 a 135.

A Figura 5.9 mostra alguns resultados interessantes alcançados nos experimentos. Por exemplo, na Figura 5.9(a), uma pessoa é uma das faces ranqueadas em primeiro lugar para todas as três poses, mas grande parte de sua face está oclusa por outra face na imagem Gigapixel. Outro caso mais complexo, mostrado na Figura 5.9(b), é de uma pessoa que corresponde a apenas uma das três poses, embora sua face esteja quase completamente oclusa por outra pessoa. Além disso, um caso desafiador devido ao fato de a pessoa estar em posição lateralizada ou até mesmo de costas é apresentado na Figura 5.9(c), onde a pessoa fica nas dez primeiras imagens no *ranking* de faces, reduzindo bastante a quantidade de faces a serem analisadas por um especialista.

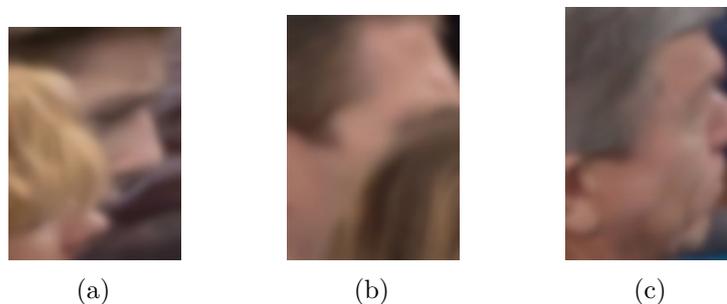


Figura 5.9: Amostras de faces em posições laterais e oclusas reconhecidas pela abordagem proposta.

5.6 Considerações Finais do Capítulo

Este capítulo apresentou um modelo para auxiliar um observador humano a encontrar pessoas em grandes imagens de multidões, reduzindo o espaço de busca de várias imagens para um *ranking* de dez imagens relacionadas a uma determinada pessoa, conforme desenvolvido por [Ferreira et al. \(2020b\)](#). Foi desenvolvida a detecção de faces em uma imagem Gigapixel apoiada em uma abordagem de varredura com agrupamento denso de pessoas e, em seguida, realizado um reconhecimento de face considerando um conjunto de três poses diferentes para cada pessoa em uma base de dados construída manualmente. Vários desafios foram abordados no processo, como a oclusão da face.

O objetivo não era encontrar uma correspondência exata, mas, sim, reduzir o esforço de busca do observador ou especialista, de modo que se a distância medida classificar uma pessoa entre as dez primeiras imagens, a pessoa seja considerada encontrada. A abordagem mostrou resultados promissores, pois foram alcançadas taxas de reconhecimento de 69% no total. Ao considerar imagens com classificação superior a 10 para as três poses, 92% delas apresentam oclusões. O método proposto tem grande potencial para ajudar um observador humano a encontrar pessoas em multidões, principalmente em imagens desordenadas, por meio de um espaço de busca reduzido.

O próximo capítulo abordará as considerações finais e trabalhos futuros, bem como apresentará os trabalhos derivados do estudo e desenvolvimento desta tese.

Conclusões, Trabalhos Futuros e Contribuições

Sumário

6.1	Conclusões	79
6.2	Trabalhos Futuros	80
6.3	Contribuições, Publicações Derivadas desta Tese e Colaborações de Pesquisa	81
6.3.1	Trabalho Publicado em Periódico Internacional	81
6.3.2	Trabalhos Publicados em Conferências ⁴	81
6.3.3	Trabalhos Publicados como cooperação, mas não relacionados com esta tese	82

6.1 Conclusões

O presente trabalho desenvolveu uma abordagem para lidar com imagens Gigapixel, as quais apresentam algumas questões como o grande volume de dados, bem como sua distribuição espacial complexa e seu processo de formação da imagem total. Nesse sentido, a hipótese considerada foi a de que usar fragmentos ou fatias da imagem Gigapixel, além de sua manipulação em resoluções menores, pode viabilizar o uso de algoritmos e técnicas tradicionais que foram projetados para serem utilizados em imagens não-Gigapixel.

Buscando entender o contexto trabalhado e também atingir seus objetivos, o presente trabalho realizou, inicialmente, uma revisão literária abordando trabalhos relacionados a imagens Gigapixel, além dos que trataram do desenvolvimento de estratégias e modelos para a detecção de pedestres e de faces em imagens convencionais e, também, em imagens Gigapixel, enfocando o crescente uso de redes neurais convolucionais.

O trabalho apresentou, ainda, uma contextualização a respeito da geração de giga imagens por processo de varredura e algumas técnicas de visualização de imagens panorâmicas, bem como discutiu problemas inerentes como a duplicidade de artefatos que se movimentam em uma cena, a separação dos pedaços de um mesmo artefato em localizações distintas, os desalinhamentos vertical e horizontal no processo de combinação de imagens, dentre outros.

Outro ponto importante trazido pelo trabalho foi a realização de um estudo e consequente desenvolvimento de um método de análise de imagens Gigapixel por varredura com o objetivo de detectar pedestres apoiado em uma análise multirresolução. Nesta etapa foram considerados quatro algoritmos de redução de resolução na detecção de pedestres e faces, tornando possível observar que essa busca constitui uma tarefa desafiadora quando não há o conhecimento prévio de detalhes da imagem Gigapixel.

Além disso, também foi proposto um modelo de varredura em agrupamento denso de pessoas, tendo sido verificados resultados promissores, no sentido de reduzir o espaço de busca do observador para um *ranking* de dez possibilidades, ao invés de um universo muito maior a ser analisado.

Dessa forma, de maneira resumida, o trabalho apresentou como contribuições o desenvolvimento de um método para lidar com as imagens Gigapixel no sentido de conseguir realizar um processo de varredura nas mesmas que possibilite encontrar pessoas e faces, considerando a possibilidade da existência ou não de um agrupamento denso das mesmas.

Também foi realizada a análise em diferentes níveis de resolução, o que constituiu uma grande contribuição alcançada, visto que foi possível verificar que as características relevantes no processo de detecção de pessoas e faces continuam presentes em resoluções diferentes da resolução original.

Por fim, destaca-se que o estudo realizado e o modelo proposto possui espaço para aperfeiçoamentos que podem ser implantados conforme o desenvolvimento dos trabalhos futuros relacionados na seção seguinte.

6.2 Trabalhos Futuros

Em seguida, são apresentadas propostas de trabalhos a serem desenvolvidos:

- Analisar a eficácia do modelo proposto em outras imagens Gigapixel com uma variação maior de cenários;
- Testar o modelo proposto com o uso de redes neurais convolucionais como a NASNet, a qual alcançou o nível de estado da arte em resultados, conforme abordado no trabalho de [Zoph et al. \(2018\)](#);

- Avaliar o impacto da utilização de super-resolução em imagens Gigapixel nos resultados alcançados, objetivando a melhoria de sua qualidade antes da utilização do modelo proposto;
- Verificar a influência de outros algoritmos de detecção de face no modelo proposto.

6.3 Contribuições, Publicações Derivadas desta Tese e Colaborações de Pesquisa

O presente trabalho trouxe contribuições relevantes como o desenvolvimento de um método de varredura em imagens Gigapixel com o objetivo de encontrar pessoas e faces, considerando giga imagens com agrupamento denso ou não das mesmas. Além disso, foi desenvolvida uma análise em diferentes níveis de resolução, com o objetivo de verificar se as características relevantes se mantiveram em resoluções diferentes da resolução original e buscando confirmar a hipótese de que é possível viabilizar o uso, em imagens Gigapixel, de técnicas e algoritmos projetados para imagens não-Gigapixel.

Além disso, é importante ressaltar que, com o estudo e o desenvolvimento do trabalho realizado nesta tese, os resultados obtidos foram submetidos e publicados nas conferências e revista detalhadas nas Subseções 6.3.1 e 6.3.2. Eles se encontram apresentados pelo tipo de publicação e organizados em ordem cronológica decrescente. Também houve a cooperação em trabalhos não relacionados com a tese, mas desenvolvidos pelo grupo de pesquisa do qual a autora da tese é parte integrante. Tais trabalhos estão relacionados na Subseção 6.3.3.

6.3.1 Trabalho Publicado em Periódico Internacional

- Ferreira, Cristiane B. R.; Soares, Fabrizzio A. A. M. N.; Pedrini, Helio; Bruce, Neil; Ferreira, William D.; Da Cruz, Gelson. (2020). **A Study of Dimensionality Reduction Impact on an Approach to People Detection in Gigapixel Images**. In: *Canadian Journal of Electrical and Computer Engineering, CJECE*, 43(August), 122–128 (Qualis A3) ([Ferreira et al., 2020a](#))

6.3.2 Trabalhos Publicados em Conferências⁴

- Internacionais

- Ferreira, Cristiane B. R.; Soares, Fabrízio A. A. M. N.; Pedrini, Helio; de Souza Alencar, Wanderley; Ferreira, William D.; Carvalho, Thyago P.; Sousa, Naiane. **Where’s Wally: A Gigapixel Image Study for Face Recognition in Crowds.** *In: 15th International Symposium on Visual Computing, ISVC 2020, Virtual Event, October 5-7, 2020*, vol. 12510, p. 386-397, Springer. Lecture Notes in Computer Science. (Qualis A4) ([Ferreira et al., 2020b](#))
- Ferreira, Cristiane B. R.; Soares, Fabrízio A. A. M. N.; Pedrini, Helio; Bruce, Neil; Ferreira, William D.; Da Cruz, Gelson. **Multiresolution Analysis on Searching for People in Gigapixel Images.** *In 31st Canadian Conference on Electrical Computer & Computer Engineering, CCECE 2018, Québec, Canada, May 13-16, 2018*, p. 1-4, IEEE 2018. (Qualis A3) ([Ferreira et al., 2018](#))
- Nacional
 - Ferreira, Cristiane B. R.; Soares, Fabrízio A. A. M.N.; Martins, Wellington S.; Costa, Ronaldo M., Ferreira; William D., Carvalho; Thyago P.; Parreira, Rafael T.; Junior, Gelson C. **Parallel CUDA Based Implementation of Gaussian Pyramid Image Reduction.** *WVC 2016 – XII Workshop de Visão Computacional, Campo Grande, MS, 2016*. (Qualis B4) ([Ferreira et al., 2016](#))

6.3.3 Trabalhos Publicados como cooperação, mas não relacionados com esta tese

- Nascimento, T. H. ; Ferreira, C. B. R.; Soares, F. **Interaction with Smartwatches Using Gesture Recognition: A Systematic Literature Review.** *In: IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 2020*, p. 1661-1666, IEEE 2020. (Qualis A2) ([Horbylon Nascimento et al., 2020](#))
- Ferreira, W. D.; Ferreira, C. B. R; Da Cruz Júnior, G.; Soares, F. **A Review of Digital Image Forensics.** *In: Computers & Electrical Engineering*, vol. 85, p. 1-9. (Qualis A1) ([Ferreira et al., 2020c](#))
- Carvalho, T.; Costa, C.; Mombach, J.; Ferreira, C.; Fernandes, D.; Soares, F. **PETA-System – A Piano Expression Teaching Aid System.** *In: IEEE Canadian Conference on Electrical Computer & Computer Engineering, CCECE 2020, Ago 30 - Sep 2, Ontario, 2020*, p. 1-5, IEEE 2020. (Qualis A3) ([Carvalho et al., 2020](#))

- Mombach, J.; Ferreira, C.; Felix, J.; Salvini, R.; Soares, F. **Mirrored and Rotated Letters in Children Spellings: An Automatic Analysis Approach.** *In: IEEE Canadian Conference on Electrical Computer & Computer Engineering, CCECE 2020, Ago 30 - Sep 2, Ontario, 2020*, p. 1-4, IEEE 2020. (Qualis A3) ([Mombach et al., 2020](#))
- Kai, P. M.; Soares, F. A. A. M. N.; Oliveira, L. L. G.; Laureano, G. T.; Ferreira, C. B. R; Carvalho, T. P. **Classificação de Doenças Foliaves em Culturas de Soja com Transformadas Wavelet de Haar.** *In: 13th Brazilian Symposium on Intelligent Automation (SBAI), Out 1-4, Porto Alegre, Brasil, 2017*, p. 115-120, SBAI 2017. (Qualis B4) ([Kai et al., 2017](#))

ANEXO A

**Termos de Direitos de Cópia de
Publicações Utilizadas Nesta Tese**

A.1 Artigo: Gigapixel Computational Imaging (Cossairt & Miao, 2011)



Home | Help | Live Chat | Sign in | Create Account



Gigapixel Computational Imaging

Conference Proceedings:
2011 IEEE International Conference on Computational Photography (ICCP)

Author: Oliver S. Cossairt
Publisher: IEEE
Date: April 2011

Copyright © 2011, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis online.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK
CLOSE WINDOW

© 2023 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)
| [For California Residents](#) | [Terms and Conditions](#) Comments? We would like to hear from you. E-mail us at customer-care@copyright.com

A.2 Artigo: Multiresolution Analysis on Searching for People in Gigapixel Images (Ferreira et al., 2018)

Home Help ▾ Live Chat Sign in Create Account



Requesting permission to reuse content from an IEEE publication

Multiresolution Analysis on Searching for People in Gigapixel Images

Conference Proceedings:
2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)
Author: Cristiane B. R. Ferreira
Publisher: IEEE
Date: May 2018

Copyright © 2018, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW

© 2023 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy
| For California Residents | Terms and Conditions Comments? We would like to hear from you. E-mail us at customer-care@copyright.com

A.3 Artigo: A Study of Dimensionality Reduction Impact on an Approach to People Detection in Gigapixel Images (Ferreira et al., 2020a)

Home Help Live Chat Sign in Create Account



Requesting permission to reuse content from an IEEE publication

A Study of Dimensionality Reduction Impact on an Approach to People Detection in Gigapixel Images

Author: Cristiane B. R. Ferreira
Publication: Canadian Journal of Electrical and Computer Engineering
Publisher: IEEE
Date: Summer 2020

Copyright © 2020, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK

CLOSE WINDOW

© 2023 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)
| [For California Residents](#) | [Terms and Conditions](#) Comments? We would like to hear from you. E-mail us at customercare@copyright.com

A.4 Artigo: Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds (Ferreira et al., 2020b)

13/03/2023, 15:57

Rightslink® by Copyright Clearance Center



- [Home](#)
- [Help](#)
- [Live Chat](#)
- [Cristiane Ferreira](#)

Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds

Author: Cristiane B. R. Ferreira, Helio Pedrini, Wanderley de Souza Alencar et al
 Publication: Springer eBook
 Publisher: Springer Nature
 Date: Jan 1, 2020

Copyright © 2020, Springer Nature Switzerland AG

Order Completed

Thank you for your order.

This Agreement between Cristiane Ferreira ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number: 5507181327247 [Printable Details](#)

License date: Mar 13, 2023

<p>Licensed Content</p> <table border="0" style="width: 100%;"> <tr><td>Licensed Content Publisher</td><td>Springer Nature</td></tr> <tr><td>Licensed Content Publication</td><td>Springer eBook</td></tr> <tr><td>Licensed Content Title</td><td>Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds</td></tr> <tr><td>Licensed Content Author</td><td>Cristiane B. R. Ferreira, Helio Pedrini, Wanderley de Souza Alencar et al</td></tr> <tr><td>Licensed Content Date</td><td>Jan 1, 2020</td></tr> </table>	Licensed Content Publisher	Springer Nature	Licensed Content Publication	Springer eBook	Licensed Content Title	Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds	Licensed Content Author	Cristiane B. R. Ferreira, Helio Pedrini, Wanderley de Souza Alencar et al	Licensed Content Date	Jan 1, 2020	<p>Order Details</p> <table border="0" style="width: 100%;"> <tr><td>Type of Use</td><td>Thesis/Dissertation</td></tr> <tr><td>Requestor type</td><td>academic/university or research institute</td></tr> <tr><td>Format</td><td>electronic</td></tr> <tr><td>Portion</td><td>figures/tables/illustrations</td></tr> <tr><td>Number of figures/tables/illustrations</td><td>20</td></tr> <tr><td>Will you be translating?</td><td>yes, without original language</td></tr> <tr><td>Number of languages</td><td>1</td></tr> <tr><td>Circulation/distribution</td><td>50000 or greater</td></tr> <tr><td>Author of this Springer Nature content</td><td>yes</td></tr> </table>	Type of Use	Thesis/Dissertation	Requestor type	academic/university or research institute	Format	electronic	Portion	figures/tables/illustrations	Number of figures/tables/illustrations	20	Will you be translating?	yes, without original language	Number of languages	1	Circulation/distribution	50000 or greater	Author of this Springer Nature content	yes
Licensed Content Publisher	Springer Nature																												
Licensed Content Publication	Springer eBook																												
Licensed Content Title	Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds																												
Licensed Content Author	Cristiane B. R. Ferreira, Helio Pedrini, Wanderley de Souza Alencar et al																												
Licensed Content Date	Jan 1, 2020																												
Type of Use	Thesis/Dissertation																												
Requestor type	academic/university or research institute																												
Format	electronic																												
Portion	figures/tables/illustrations																												
Number of figures/tables/illustrations	20																												
Will you be translating?	yes, without original language																												
Number of languages	1																												
Circulation/distribution	50000 or greater																												
Author of this Springer Nature content	yes																												
<p>About Your Work</p> <table border="0" style="width: 100%;"> <tr><td>Title</td><td>Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds</td></tr> <tr><td>Institution name</td><td>UFG</td></tr> <tr><td>Expected presentation date</td><td>May 2023</td></tr> </table>	Title	Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds	Institution name	UFG	Expected presentation date	May 2023	<p>Additional Data</p> <table border="0" style="width: 100%;"> <tr><td>Order reference number</td><td>1</td></tr> <tr><td>Portions</td><td>Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Table 1, Table 2</td></tr> <tr><td>Specific Languages</td><td>Portuguese</td></tr> </table>	Order reference number	1	Portions	Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Table 1, Table 2	Specific Languages	Portuguese																
Title	Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds																												
Institution name	UFG																												
Expected presentation date	May 2023																												
Order reference number	1																												
Portions	Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Table 1, Table 2																												
Specific Languages	Portuguese																												
<p>Requestor Location</p> <table border="0" style="width: 100%;"> <tr><td colspan="2">Cristiane Ferreira Rua T-29, n. 686, ap. 2502, St Bueno</td></tr> <tr><td>Requestor Location</td><td>Goiânia, Goiás 74150-200 Brazil Attn: Cristiane Ferreira</td></tr> </table>	Cristiane Ferreira Rua T-29, n. 686, ap. 2502, St Bueno		Requestor Location	Goiânia, Goiás 74150-200 Brazil Attn: Cristiane Ferreira	<p>Tax Details</p> <table border="0" style="width: 100%;"> <tr><td>Customer VAT ID</td><td>BR78899206104</td></tr> </table>	Customer VAT ID	BR78899206104																						
Cristiane Ferreira Rua T-29, n. 686, ap. 2502, St Bueno																													
Requestor Location	Goiânia, Goiás 74150-200 Brazil Attn: Cristiane Ferreira																												
Customer VAT ID	BR78899206104																												
<p>Price</p> <table border="0" style="width: 100%;"> <tr><td>Total</td><td>0.00 USD</td></tr> </table>	Total	0.00 USD	<p>Total: 0.00 USD</p>																										
Total	0.00 USD																												

[CLOSE WINDOW](#) [ORDER MORE](#)

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Mar 13, 2023

This Agreement between Cristiane Ferreira ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5507181327247
License date	Mar 13, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds
Licensed Content Author	Cristiane B. R. Ferreira, Helio Pedrini, Wanderley de Souza Alencar et al
Licensed Content Date	Jan 1, 2020
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	20
Will you be translating?	yes, without original language
Number of languages	1

Circulation/distribution	50000 or greater
Author of this Springer Nature content	yes
Title	Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds
Institution name	UFG
Expected presentation date	May 2023
Order reference number	1
Portions	Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Table 1, Table 2
Specific Languages	Portuguese
Requestor Location	Cristiane Ferreira Rua T-29, n. 686, ap. 2502, St Bueno Goiânia, Goiás 74150-200 Brazil Attn: Cristiane Ferreira
Customer VAT ID	BR78899206104
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time

as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at <https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/>.
4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.
4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: '*Reproduced with permission from Springer Nature*'.

8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH

DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United Nations Convention on Contracts for the International Sale of Goods) or to Germany's choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

Questions? For questions on Copyright Clearance Center accounts or website issues please contact springernaturesupport@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit <https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution>

Other Conditions:

Version 1.4 - Dec 2022

Questions? customercare@copyright.com.

Bibliografia

- Abbadi, Nidhal K EL, Al Hassani, Safaa Alwan, & Abdulkhaleq, Ali Hussein. 2021. A review over panoramic image stitching techniques. Page 012115 of: Journal of Physics: Conference Series, vol. 1999. IOP Publishing.
- Arthur, Kwamina E. 2017 (acessado em 3 de Agosto de 2020). How CNN Captured the Gigapixel Image of Trump's Inauguration. <http://www.afd-techtalk.com/gigapixel-trump/>.
- Ben-Ezra, Moshe. 2010. High resolution large format tile-scan camera: Design, calibration, and extended depth of field. Pages 1–8 of: 2010 IEEE International Conference on Computational Photography (ICCP). IEEE.
- Brunetti, Antonio, Buongiorno, Domenico, Trotta, Gianpaolo Francesco, & Bevilacqua, Vitoantonio. 2018. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing, **300**(7), 17–33.
- Burt, Peter J. 1981. Fast Filter Transforms for Image Processing*. Computer Graphics and Image Processing, **16**, 20–51.
- Burt, Peter J, & Adelson, Edward H. 1983a. The Laplacian Pyramid as a Compact Image Code. IEEE Transactions on Communications.
- Burt, Peter J, & Adelson, Edward H. 1983b. A Multiresolution Spline With Application to Image Mosaics. ACM Transactions on Graphics, **2**, 217–236.
- Cao, Xian Bin, Qiao, Hong, & Keane, John. 2008. A low-cost pedestrian-detection system with a single optical camera. IEEE Transactions on Intelligent Transportation Systems, **9**(3), 58–67.
- Cao, Xianbin, Qiao, Hong, Wang, Fei-Yue, & Zhang, Xinzheng. 2005. LNCS 3495 - Application of Cooperative Co-evolution in Pedestrian Detection Systems. vol. 3495.

- Cao, Zhijie, Yan, Renyou, Huang, Yiyong, & Shi, Zhiru. 2019. Gigapixel-level image crowd counting using csrnet. Pages 426–428 of: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE.
- Carvalho, Thyago, Costa, Carlos, Mombach, Jaline, Ferreira, Cristiane, Fernandes, Deborah, & Soares, Fabrizzio. 2020. PETA-System – A Piano Expression Teaching Aid System. Pages 1–5 of: IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).
- Clarke, Alasdair D. F., Elsner, Micha, & Rohde, Hannah. 2013. Where’s Wally: the influence of visual salience on referring expression generation. Frontiers in Psychology, **4**.
- Cossairt, Oliver S, & Miao, Daniel. 2011. Gigapixel Computational Imaging. Pages 1–8 of: 2011 IEEE International Conference on Computational Photography (ICCP).
- Cruz, Marcos, & González-Villa, Javier. 2021. Unbiased population size estimation on still gigapixel images. Sociological Methods & Research, **50**(2), 627–648.
- Dalal, Navneet, & Triggs, Bill. 2005. Histograms of oriented gradients for human detection. Pages 886–893 of: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. I.
- Dollar, Piotr. 2018. Piotr’s Computer Vision Matlab Toolbox (PMT).
- Dollar, Piotr, Tu, Zhuowen, Perona, Pietro, & Belongie, Serge. 2009. Integral Channel Features.
- Dollar, Piotr, Belongie, Serge, & Perona, Pietro. 2010a. The fastest pedestrian detector in the west. British Machine Vision Association, BMVA.
- Dollar, Piotr, Wojek, Christian, Schiele, Bernt, & Perona, Pietro. 2010b. Pedestrian Detection: A Benchmark. Institute of Electrical and Electronics Engineers (IEEE).
- Dollar, Piotr, Wojek, Christian, Schiele, Bernt, & Perona, Pietro. 2012. Pedestrian Detection: An Evaluation of the State of the Art. IEEE Transactions on Pattern Analysis and Machine Intelligence, **34**(4), 743–761.
- Dollar, Piotr, Appel, Ron, Belongie, Serge, & Perona, Pietro. 2014. Fast Feature Pyramids for Object Detection.
- Fan, Jiahao, Liu, Huabin, Yang, Wenjie, See, John, Zhang, Aixin, & Lin, Weiyao. 2022. Speed Up Object Detection on Gigapixel-Level Images With Patch Arrangement.

- Pages 4653–4661 of: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Feng, Yuantao, Yu, Shiqi, Peng, Hanyang, Li, Yan Ran, & Zhang, Jianguo. 2022. Detect Faces Efficiently: A Survey and Evaluations. IEEE Transactions on Biometrics, Behavior, and Identity Science, **4**(1), 1–18.
- Ferreira, Cristiane B. R., Soares, Fabrizzio A. A. M. N., Martins, Wellington S., Costa, Ronaldo M., Ferreira, William D., Carvalho, Thyago P., Parreira, Rafael T., & Junior, Gelson C. 2016 (October). Parallel CUDA Based Implementation of Gaussian Pyramid Image Reduction. In: 12th Brazilian Workshop on Computer Vision (WVC). UCDB, Campo Grande, MS, Brazil.
- Ferreira, Cristiane B. R., Soares, Fabrizzio, Pedrini, Helio, Bruce, Neil, Ferreira, William D., & Da Cruz Junior, Gelson. 2018. Multiresolution Analysis on Searching for People in Gigapixel Images. Pages 1–4 of: 2018 IEEE Canadian Conference on Electrical Computer Engineering (CCECE).
- Ferreira, Cristiane B. R., Soares, Fabrizzio A. A. M. N., Pedrini, Helio, Bruce, Neil, Ferreira, William D., & Da Cruz, Gelson. 2020a. A Study of Dimensionality Reduction Impact on an Approach to People Detection in Gigapixel Images. Canadian Journal of Electrical and Computer Engineering, **43**(August), 122–128.
- Ferreira, Cristiane B. R., Pedrini, Helio, de Souza Alencar, Wanderley, Ferreira, William D., Carvalho, Thyago Peres, Sousa, Naiane, & Soares, Fabrizzio. 2020b (October). Where's Wally: A Gigapixel Image Study for Face Recognition in Crowds. Pages 386–397 of: Advances in Visual Computing: 15th International Symposium (ISVC), San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II. Springer.
- Ferreira, William D., Ferreira, Cristiane B. R., da Cruz Júnior, Gelson, & Soares, Fabrizzio. 2020c. A Review of Digital Image Forensics. Computers & Electrical Engineering, **85**(July), 106685.
- Gall, Dider Le. 1991. MPEG: A Video Compression Standard for Multimedia Applications. Communications of the ACM, **34**.
- Gonzalez, Rafael C., & Woods, Richard E. 2008. Digital image processing. Upper Saddle River, N.J.: Prentice Hall.
- Gueguem, Lionel, Pesaresi, Martino, & Soille, Pierre. 2011. An interactive image mining tool handling gigapixel images. Pages 1581–1584 of: IEEE International Geoscience and Remote Sensing Symposium. IEEE.

- Harris, Chris, & Stephens, Mike. 1988. A combined corner and edge detector.
- Horbylon Nascimento, Thamer, Ferreira, Cristiane B. R., Rodrigues, Wellington G., & Soares, Fabrizzio. 2020. Interaction with Smartwatches Using Gesture Recognition: A Systematic Literature Review. Pages 1661–1666 of: IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC).
- Hosang, Jan, Benenson, Rodrigo, Dollar, Piotr, & Schiele, Bernt. 2015. What makes for effective detection proposals? 2.
- Hu, P., & Ramanan, D. 2017. Finding Tiny Faces. Pages 1522–1530 of: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society.
- Kai, Priscila M., Soares, Fabrizzio A. A. M. N., Oliveira, Leandro L. G., Laureano, Gustavo T., Ferreira, Cristiane B. R., & Carvalho, Thyago P. 2017. Classificação de Doenças Foliares em Culturas de Soja com Transformadas Wavelet de Haar. Pages 115–120 of: 13th Brazilian Symposium on Intelligent Automation (SBAI).
- Kazemi, Vahid, & Sullivan, Josephine. 2014. One millisecond face alignment with an ensemble of regression trees. Pages 1867–1874 of: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society.
- Kihara, N., Nakamura, K., Saito, E., & Kambara, M. 1982. The Electronic Still Camera a New Concept in Photography. IEEE Transactions on Consumer Electronics, **CE-28**, 325–331.
- Leininger, Brian, Edwards, Jonathan, Antoniades, John, Chester, David, Haas, Dan, Liu, Eric, Stevens, Mark, Gershfield, Charlie, Braun, Mike, Targove, James D., Wein, Steve, Brewer, Paul, Madden, Donald G., & Shafique, Khurram Hassan. 2008. Autonomous Real-time Ground Ubiquitous Surveillance - Imaging System (ARGUS-IS). Page 69810H of: Defense Transformation and Net-Centric Systems 2008, vol. 6981. SPIE.
- Levenberg, Kenneth. 1944. A Method for the Solution of Certain Non-linear Problems in Least Squares. Vol. 2.
- Li, Haoxiang, Lin, Zhe, Shen, Xiaohui, Brandt, Jonathan, & Hua, Gang. 2015a. A convolutional neural network cascade for face detection. Pages 5325–5334 of: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07-12-June-2015. IEEE Computer Society.

- Li, Jianguo, & Zhang, Yimin. 2013. Learning SURF cascade for fast and accurate object detection. Pages 3468–3475 of: IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- Li, Lingling, Guo, Xiaohui, Wang, Yan, Ma, Jingjing, Jiao, Licheng, Liu, Fang, & Liu, Xu. 2022. Region NMS-based deep network for gigapixel level pedestrian detection with two-step cropping. Neurocomputing, **468**(1), 482–491.
- Li, Stan Z, Yang, Bin, Yan, Junjie, & Lei, Zhen. 2015b. Convolutional Channel Features For Pedestrian, Face and Edge Detection Convolutional Channel Features. Computer Science.
- Lienhart, Rainer, Kuranov, Alexander, & Pisarevsky, Vadim. 2003. Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. Pages 297–304 of: Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25. Springer.
- Lin, Tsung Yi, Goyal, Priya, Girshick, Ross, He, Kaiming, & Dollar, Piotr. 2020. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, **42**(2), 318–327.
- Lowe, David G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, **60**, 91–110.
- Luo, Shi, Li, Xiongfei, Zhu, Rui, & Zhang, Xiaoli. 2019. SFA: Small Faces Attention Face Detector. IEEE Access, **7**, 171609–171620.
- Lyu, Wei, Zhou', Zhong, Chen, Lang, & Zhou, Vi. 2019a. A survey on image and video stitching. Virtual Reality & Intelligent Hardware, **1**, 55–83.
- Lyu, Wentao, Ding, Peng, Zhang, Yingliang, Chen, Anpei, Wu, Minye, Yin, Shu, & Yu, Jingyi. 2019b. Refocusable gigapixel Panoramas for immersive VR experiences. IEEE Transactions on Visualization and Computer Graphics, **27**(3), 2028–2040.
- Mao, Jiayuan, Xiao, Tete, & Jiang, Yuning. 2017. What Can Help Pedestrian Detection? Pages 3127–3136 of: IEEE Conference on Computer Vision and Pattern Recognition.
- Marks, Daniel L, & Brady, David J. 2010. Gigagon: a monocentric lens design imaging 40 gigapixels. Page ITuC2 of: Imaging Systems. Optica Publishing Group.
- Mo, Wanghao, Ke, Yan, Huo, Qingao, Cao, Ruyi, Song, Guochang, & Zhang, Wendong. 2022. ADN: ATSS-based Deep Network for Pedestrian Detection

- on High-Resolution Images. Pages 454–458 of: 2022 International Conference on Networking and Network Applications (NaNA). IEEE.
- Mo, Wanghao, Zhang, Wendong, Wei, Hongyang, Cao, Ruyi, Ke, Yan, & Luo, Yiwen. 2023. PVDet: Towards pedestrian and vehicle detection on gigapixel-level images. Engineering Applications of Artificial Intelligence, **118**, 105705.
- Mombach, Jaline, Ferreira, Cristiane, Felix, Juliana, Salvini, Rogerio, & Soares, Fabrizio. 2020. Mirrored and Rotated Letters in Children Spellings: An Automatic Analysis Approach. Pages 1–4 of: IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).
- NASA, JPL-Caltech, Cornell, & ASU. 2019. Opportunity Legacy Pan.
- Nichols, Mary H., Ruyle, George B., & Nourbakhsh, Illah R. 2009. Very-high-resolution panoramic photography to improve conventional rangeland monitoring. Rangeland Ecology and Management, **62**, 579–582.
- Nourbakhsh, Illah R. 2010. GigaPixels for science. IEEE Robotics and Automation Magazine, **17**(6), 101–104.
- Papageorgiou, Constantine P., Oren, Michael, & Poggio, Tomaso. 1998. General framework for object detection. Pages 555–562 of: IEEE International Conference on Computer Vision. IEEE.
- Parulski, K., & Rabbani, M. 2000. The Continuing Evolution of Digital Cameras and Digital Photography Systems. vol. 1. IEEE.
- Port Authority of New York and New Jersey. 2018 (Apr.). 2017 Annual Airport Traffic Report. Retrieved: 2019-02-21.
- Powell, Mark W., Rossi, Ryan A., & Shams, Khawaja. 2010. A scalable image processing framework for gigapixel mars and other celestial body images.
- Preedy, Victor R. 2012. Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease. Springer Science & Business Media.
- Rasouli, Amir, Kotseruba, Iuliia, & Tsotsos, John K. 2017. Agreeing to cross: How drivers and pedestrians communicate. Pages 264–269 of: IEEE Intelligent Vehicles Symposium (IV).
- Richardson, Iain E. 2011. The H. 264 advanced video compression standard. John Wiley & Sons.

- Rosten, Edward, & Drummond, Tom. 2006. *Machine Learning for High-Speed Corner Detection*. Springer Berlin Heidelberg.
- Sargent, Randy, Bartley, Chris, Dille, Paul, Keller, Jeff, & Nourbakhsh, Illah. 2010. *Timelapse GigaPan: Capturing, Sharing, and Exploring Timelapse Gigapixel Imagery*.
- Schott, Ron. 2009 (1). [GigaPan High Resolution Images](#).
- Seidl, M, & Breiteneder, C. 2011. *Detection and Classification of Petroglyphs in Gigapixel Images-Preliminary Results*. Pages 45–48 of: 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST.
- Smith, M. Alex. 2010. *A year in an urban forest: Dairy Bush GigaPan 2009-2010*.
- StackExchange. 2017 (acessado em 15 de Março de 2022). [How did CNN take that Gigapixel Photo for Trump's Inauguration?](https://photo.stackexchange.com/questions/86489/how-did-cnn-take-that-gigapixel-photo-for-trumps-inauguration)
<https://photo.stackexchange.com/questions/86489/how-did-cnn-take-that-gigapixel-photo-for-trumps-inauguration>.
- Val, Mauricio. 2013 (7). [GigaPan High Resolution Images](#).
- Viola, Paul, & Jones, Michael. 2001. *Rapid object detection using a boosted cascade of simple features*. In: [IEEE Computer Society Conference on Computer Vision and Pattern Recognition](#), vol. 1.
- Viola, Paul, & Jones, Michael J. 2004. *Robust Real-Time Face Detection*. [International Journal of Computer Vision](#), **57**, 137–154.
- Viola, Paul, Jones, Michael J, & Snow, Daniel. 2005. *Detecting Pedestrians Using Patterns of Motion and Appearance*. [International Journal of Computer Vision](#), **63**, 153–161.
- Wang, Chunlei, Feng, Wenquan, Liu, Binghao, Ling, Xinyang, & Yang, Yifan. 2023. *Exploiting the Potential of Overlapping Cropping for Real-World Pedestrian and Vehicle Detection with Gigapixel-Level Images*. [Applied Sciences](#), **13**(6), 3637.
- Wang, Xueyang, Zhang, Xiya, Zhu, Yinheng, Guo, Yuchen, Yuan, Xiaoyun, Xiang, Liuyu, Wang, Zerun, Ding, Guiguang, Brady, David, Dai, Qionghai, et al. 2020a. *Panda: A gigapixel-level human-centric video dataset*. Pages 3268–3278 of: [Proceedings of the IEEE/CVF conference on computer vision and pattern recognition](#).

- Wang, Yi, Yang, Youlong, & Zhao, Xi. 2020b. Object detection using clustering algorithm adaptive searching regions in aerial images. Pages 651–664 of: [European Conference on Computer Vision](#). Springer.
- Wilburn, Bennett, Joshi, Neel, Vaish, Vaibhav, Talvala, Eino-Ville, Antunez, Emilio, Barth, Adam, Adams, Andrew, Horowitz, Mark, & Levoy, Marc. 2005. High performance imaging using large camera arrays. Pages 765–776 of: [ACM SIGGRAPH 2005 Papers](#).
- Zhang, Haodi, Pop, Danut Ovidiu, Rogozan, Alexandrina, & Bensrhair, Abdelaziz. 2021. Accelerate High Resolution Image Pedestrian Detection with Non-Pedestrian Area Estimation. [IEEE Access](#), **9**, 8625–8636.
- Zhang, Kaipeng, Zhang, Zhanpeng, Li, Zhifeng, & Qiao, Yu. 2016. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. [IEEE Signal Processing Letters](#), **23**(10), 1499–1503.
- Zhang, Lun, Chu, Rufeng, Xiang, Shiming, Liao, Shengcai, & Li, Stan Z. 2007. Face detection based on multi-block LBP Representation. [Advances in Biometrics](#), 11–18.
- Zoph, Barret, Vasudevan, Vijay, Shlens, Jonathon, & Le, Quoc V. 2018 (7). Learning Transferable Architectures for Scalable Image Recognition. In: [IEEE Conference on Computer Vision and Pattern Recognition \(CVPR\)](#).