



Codificação e Fusão de Imagens Multifoco **Comparação entre imagens codificadas fundidas e** **fusão de imagens codificadas**

Marta Filipa Gomes Peraboa

Dissertação para obtenção do Grau de Mestre em
Engenharia Eletrotécnica e de Computadores
(2º ciclo de estudos)

Orientador: Prof. Doutor António Manuel Gonçalves Pinheiro

outubro de 2024.

Declaração de Integridade

Eu, Marta Filipa Gomes Peraboa, que abaixo assino, estudante com o número de inscrição M12909 de/o Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referência de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 11 /10/2024

Dedicatória

Dedico esta tese aos meus pais e aos meus avós, por me terem passado os valores e os princípios que me permitiram chegar até aqui. Por me terem ensinado que nenhum sonho é demasiado grande, que não se possa tornar realidade. E principalmente, por nunca terem duvidado de mim e me terem apoiado sempre incondicionalmente, em cada escolha que fiz ao longo do meu percurso até chegar aqui.

Gostava também de dedicar a minha tese ao meu irmão, que me tem mostrado, ao longo dos anos, que é possível alcançar os nossos objetivos de forma descontraída e me ensinou a olhar para os desafios de forma diferente.

A vocês que são os meus pilares.

Agradecimentos

Aos meus pais e irmão, por me terem apoiado desde o primeiro dia ao longo de toda esta jornada e nunca me terem deixado perder o foco. Agradeço a compreensão e a motivação que me deram ao longo deste trabalho.

À Catarina, que me tem visto crescer ao longo de muitos anos e que partilhou também comigo este percurso. A sua determinação e empenho fez-me querer ser melhor e aprender com ela todos os dias.

À minha Tia Rosa, que me conhece desde sempre, mas desde que comecei este percurso tem sido um dos meus pilares e me tem ajudado em tudo.

À Daniela e ao Marcus, por terem partilhado este percurso comigo desde o primeiro dia, e o terem tornado mais leve, por terem estado presentes nas minhas conquistas mas principalmente por me terem ajudado a superar as minhas dificuldades.

Finalmente, a todos os que de alguma forma contribuíram e marcaram este meu percurso e que têm um lugar especial no meu coração.

Resumo

A fusão de imagens é uma técnica da fotografia computacional que é essencial para obter imagens de alta qualidade. A aquisição destas imagens é um passo crucial nas aplicações de processamento de imagens.

Uma vez que, a profundidade de campo é limitada pelas lentes óticas usadas, tornando-se difícil obter uma imagem em que todos os objetos se encontrem focados. Para colmatar este problema é necessário usar uma técnica de baixo custo que ajude a obter uma imagem em que todos os planos estejam focados.

Este trabalho, tem assim como objetivo verificar se é possível usar imagens codificadas para fazer a fusão de imagens multifoco.

Sabe-se que o processo de fusão de imagens multifoco, tem sido aplicado em vários domínios diferentes como é o caso do processamento de imagens médicas, da fusão de micro-imagens, das redes de sensores visuais, da análise bioquímica, entre muitos outros. No entanto, têm várias limitações, que têm sido apresentadas como é o caso de erros de registo, efeitos de desfocagem e elevada sensibilidade ao ruído.

Estas imagens podem ser obtidas através de diferentes métodos de fusão de imagens multifoco. As quatro categorias principais são: o domínio espacial, o domínio da transformação, de *deep learning* e os métodos híbridos. Estes tentam colmatar algumas das limitações apresentadas pelos processos de fusão multifoco. Assim sendo, dentro de cada uma destas categorias existem várias subcategorias, que têm o mesmo propósito que a categoria a que pertencem, mas que utilizam metodologias diferentes de aquisição de imagens e de tratamento das mesmas.

Tipicamente são obtidas duas ou mais imagens do mesmo cenário com diferentes pontos de focagem o que permite através da fusão construir uma imagem mais rica em termos de focagem. Os dois métodos usados para a aquisição das imagens estudadas foram: codificar primeiro as imagens e de seguida fundi-las ou fundi-las e depois codificá-las.

Foi assim, selecionado um método de fusão que apresenta um bom desempenho, a ASR, que combina informações de duas imagens com diferentes pontos de focagem para gerar uma imagem de melhor qualidade. Dessa forma, as imagens geradas pelo algoritmo de fusão utilizando imagens codificadas, foram comparadas com aquelas que foram geradas pelo mesmo algoritmo, mas sem codificação.

Assim, as imagens usadas no processo de fusão vão ser codificadas usando diferentes codecs. Neste caso, os codecs utilizados são o JPEG, o JPEG 2000 e o JPEG XL.

Para avaliar a qualidade das imagens obtidas, utilizando os codecs escolhidos, foi usado o seguinte conjunto de métricas: o PSNR-HVS-M, o MS-SSIM e o FSIMc. Estas métricas fazem a comparação da imagem fundida (a partir das imagens originais) com as imagens que sofreram codificação e fusão ou fusão e codificação. São também feitas comparações das imagens codificadas com a imagem original da base de dados.

Os resultados obtidos foram apresentados através de tabelas, uma para cada métrica o

PSNR-HVS-M, o MS-SSIM e o FSIMc, onde são apresentados os valores da área entre as duas curvas obtidas. Uma das curvas é referente aos valores obtidos através das taxas de bits para as imagens codificadas e fundidas e a outra para as imagens fundidas com posterior codificação. Foi também elaborada uma tabela para analisar a Delta Bjøntegaard da média com o respectivo desvio padrão para as três métricas usadas. Analisou-se ainda, qual o custo em termos de taxa de bits para obter a mesma qualidade com imagens originais codificadas que se obtém com imagens originais sem codificação.

Através da análise dos codecs estudados, observou-se que a JPEG XL é o que tem maior capacidade de compressão. Isto foi comprovado através dos resultados obtidos através das métricas usadas. O objetivo deste trabalho era determinar se é possível comprimir as duas imagens originais em vez de comprimir a imagem fundida sem comprometer a qualidade da fusão. Os resultados obtidos mostram que tal é possível.

No entanto, uma vez que os codecs analisados teriam de codificar ambas as imagens, seria necessário o dobro do espaço de armazenamento (para armazenar duas imagens) em comparação com o armazenamento da imagem resultante da fusão. Contudo, é de notar que o armazenamento de duas imagens com capacidade de fusão permite novas fusões que podem resultar em imagens de maior qualidade.

Palavras-chave

Fusão de imagens, Multifoco, ASR, Compressão, Qualidade, Armazenamento

Índice

Dedicatória	v
Agradecimentos	vii
Resumo	ix
Índice	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
Lista de Acrónimos	xvii
1 Introdução	1
1.1 Objetivos	1
2 Estado da Arte	3
2.1 Imagem MFIF	3
2.2 Classificação de métodos de fusão de imagem	5
2.3 Métodos no domínio espacial	6
2.3.1 Métodos baseados em pixels	6
2.3.2 Métodos baseados em blocos	7
2.3.3 Métodos baseados na região	7
2.4 Métodos no domínio da transformada	8
2.4.1 Métodos baseados na decomposição multiescala (MSD)	8
2.4.2 Métodos baseados no domínio do gradiente (GD)	11
2.4.3 Métodos baseados em representação dispersa (SR)	12
2.4.4 Método baseado na transformada discreta do cosseno (DCT)	12
2.5 Métodos baseados em <i>deep learning</i>	12
2.5.1 Métodos baseados em classificação	13

2.5.2	Método baseado na regressão	13
2.6	Método híbrido	14
2.7	Fusão de Múltiplas Imagens	15
2.8	Codificação de Imagens	16
2.8.1	JPEG	16
2.8.2	JPEG 2000	17
2.8.3	JPEG XL	18
2.9	Métricas de desempenho	21
2.9.1	<i>PSNR-HVS-M</i>	21
2.9.2	<i>MS-SSIM</i>	22
2.9.3	<i>FSIMc</i>	22
3	Metodologia	25
3.1	Base de Dados	25
3.2	Procedimentos utilizados	27
3.3	Taxas de <i>bitrates</i> (BPP)	27
3.4	Exemplo dos procedimentos utilizados	28
4	Resultados	31
4.1	JPEG, JPEG 2000, JPEG XL	31
4.1.1	Codificação/descodificação com posterior fusão	31
4.1.2	Fusão com posterior codificação e descodificação	31
4.2	Análise de Resultados	41
4.2.1	Gráficos de análise das métricas (global)	41
5	Conclusões e trabalho futuro	53
	Bibliografia	55

Lista de Figuras

1.1	Esquema proposto e respectivas comparações	2
2.1	Processo da MFIF	4
2.2	Classificação das técnicas de MFIF, extraído de [1]	5
2.3	Visão geral da compressão e descompressão da JPEG, extraído de [2]	16
2.4	Arquitetura do sistema central(core) Kakadu, extraído de [3]	18
2.5	Cenário exemplo, extraído de [4]	19
2.6	Arquitetura do Codec, extraído de [4]	20
3.1	Imagens de referência da base de imagens escolhidas para o trabalho [5]	26
3.2	Imagens codificadas com posterior fusão	29
3.3	Imagens fundidas com posterior codificação	29
4.1	Codificação/descodificação usando o JPEG para a Imagem 086 com posterior fusão	32
4.2	Codificação/descodificação usando o JPEG para a Imagem 155 com posterior fusão	33
4.3	Codificação/descodificação usando o JPEG para a Imagem 621 com posterior fusão	34
4.4	Codificação/descodificação usando o JPEG 2000 para a Imagem 086 com posterior fusão	35
4.5	Codificação/descodificação usando o JPEG 2000 para a Imagem 155 com posterior fusão	36
4.6	Codificação/descodificação usando o JPEG 2000 para a Imagem 621 com posterior fusão	37
4.7	Codificação/descodificação usando o JPEG XL para a Imagem 086 com posterior fusão	38
4.8	Codificação/descodificação usando o JPEG XL para a Imagem 155 com posterior fusão	39
4.9	Codificação/descodificação usando o JPEG XL para a Imagem 621 com posterior fusão	40

4.10 Fusão da Imagem 086 com posterior codificação/descodificação	42
4.11 Fusão da Imagem 155 com posterior codificação/descodificação	43
4.12 Fusão da Imagem 621 com posterior codificação/descodificação	44

Lista de Tabelas

4.1	Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para a Imagem 086	45
4.2	Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para a Imagem 155	46
4.3	Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para a Imagem 086	47
4.5	Valores obtidos através das BD para o codec JPEG	47
4.4	Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para as restantes imagens	50
4.6	Valores obtidos através das BD para o codec JPEG 2000 (Kakadu)	50
4.7	Valores obtidos através das BD para o codec JPEG XL	51
4.8	Tabela das BD da Média \pm Desvio Padrão	51

ABC	<i>Artificial Bee Colony</i>
ASR	<i>Adaptive Sparse Representation</i>
BD	<i>Bjontegaard Delta</i>
CAB	<i>Content Adaptive Blurring</i>
CNN	<i>Convolutional Neural Network</i>
CSF	<i>Contrast Sensitivity Function</i>
CST	<i>Complex Shearlet</i>
DCT	<i>Discrete Cosine Transform</i>
DOF	<i>Depth of Field</i>
DRPL	<i>Deep Regression Pair Learning</i>
DWT	<i>Discrete Wavelet Transform</i>
EOL	<i>Energy of Laplacian</i>
FSIM	<i>Feature Similarity Index Measure</i>
FSD	<i>Filter-Subtract-Decimate</i>
FT	<i>Fourier Transform</i>
GA	<i>Generic Algorithm</i>
GD	<i>Gradient Domain</i>
GM	<i>Gradient Magnitude</i>
HVS	<i>Human Visual System</i>
IF	<i>Image Fusion</i>
IFCNN	<i>IF Scheme Based on Convolutional Neural Network</i>
JPEG	<i>Joint Photographic Experts Group</i>
LPT	<i>Laplacian Pyramid Transform</i>
LRR	<i>Low-Rank Representation</i>
LSF	<i>Local Spatial Frequency</i>
LUV	<i>LUV Color Space</i>
LWT	<i>Lifting Wavelet Transform</i>
MADCNN	<i>Multi-Scale Visual Attention Deep Convolutional Neural Network</i>
MFCRF	<i>Conditional Random Field Optimization</i>
MFIF	<i>Multi-Focus Image Fusion</i>
MGA	<i>Multi-Scale Geometric Analysis</i>
MLFCNN	<i>Multi-Level Features CNN</i>
MSD	<i>Multi-Scale Decomposition</i>
MS-SSIM	<i>Multi-Scale Structural Similarity Index Measure</i>
MPCNN	<i>Modified PCNN</i>
NSCT	<i>Non-Sub-Sampled Contourlet Transform</i>
OI-PCNN	<i>Motivated Pulse Information Coupled Neural Networks</i>
PCA	<i>Principal Component Analysis</i>
PCNN	<i>Pulse Coupled Neural Network</i>
PSNR	<i>Peak Signal-to-Noise Ratio</i>
QWT	<i>Quaternion Wavelet Transform</i>
RMSE	<i>Root Mean Square Error</i>

RMLP	<i>Region Mosaicking on Laplacian Pyramids</i>
SF	<i>Spatial Frequency</i>
SIFT	<i>Scale Invariant Feature Transform</i>
SNR	<i>Signal-to-Noise Ratio</i>
SR	<i>Sparse Representation</i>
SSIM	<i>Structural Similarity Index Measure</i>

Capítulo 1

Introdução

A Fusão de Imagem Multifoco, do inglês *Multi-focus image fusion* (MFIF) ([1]) é um método que combina duas ou mais imagens originais para obter uma única imagem focada, denominada imagem de referência, com melhor qualidade e com mais informações do que as imagens originais. Uma vez que, a profundidade de campo é limitada devido ao sistema de imagem, extrair todas as informações de uma única imagem é muito desafiador. Assim, duas ou mais imagens originais desfocadas são fundidas para obter uma imagem composta.

1.1 Objetivos

Este trabalho tem como objetivo verificar se é possível usar imagens codificadas para fazer a fusão de imagens multifoco. Para isso implementou-se o modelo descrito na Figura 1.1.

Como se pode observar vão-se codificar duas imagens com diferentes pontos focais. De seguida, o resultado da descodificação dessas duas imagens é fundida originando uma imagem multifoco. Essa imagem é comparada com a imagem resultante da fusão das originais.

Para permitir averiguar se este processo degrada a qualidade da imagem fundida, as imagens fundidas a partir do original também são codificadas e comparadas com a imagem fundida a partir dos originais.

Neste estudo são usados três codificadores de imagem: JPEG, JPEG 2000 e JPEG XL. A qualidade das imagens é analisada usando as métricas PSNR-HVS-M, MS-SSIM e FSIM cor.

Usa-se o método de representação esparsa adaptativa, do inglês *Adaptive Sparse Representation* ASR, para a fusão das imagens originais.

O resto da dissertação, está organizada da seguinte forma.

O Capítulo 2 tem um resumo dos métodos de MFIF que existem, explica também o conceito de fusão de múltiplas imagens, nomeadamente o algoritmo usado, e aprofunda os codificadores usados. O Capítulo 3 introduz os objetivos do trabalho, bem como a base de imagens que utilizada e as metodologias utilizadas. O Capítulo 4 apresenta as metodologias usadas neste trabalho. Finalmente, no Capítulo 5 encontram-se as conclusões e as perspectivas para um trabalho futuro.

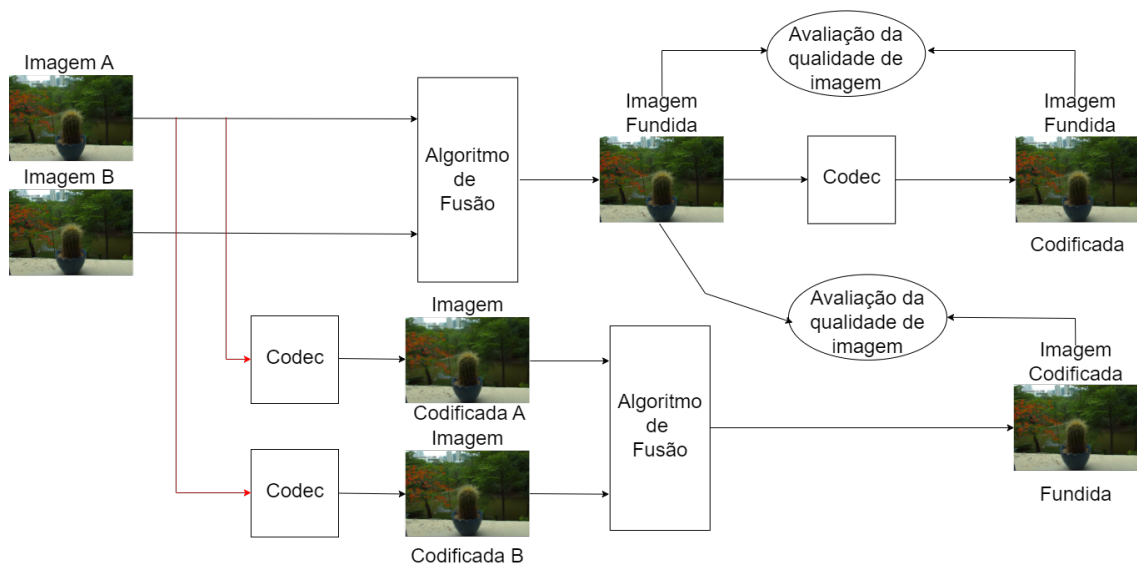


Figura 1.1: Esquema proposto e respetivas comparações

Capítulo 2

Estado da Arte

Este capítulo tem como objetivo explicar o conceito de fusão de múltiplas imagens, nomeadamente o algoritmo usado neste trabalho.

Foi feito um resumo dos métodos MFIF que existem. Estes são classificados em quatro categorias principais: no domínio espacial, no domínio da transformação, de *deep learning* e os métodos híbridos e serão discutidos juntamente com as suas vantagens e desvantagens, bem como os desafios que apresentam.

Serão aprofundados os codificadores da JPEG que são utilizados, sendo eles o próprio JPEG, o JPEG 2000 (versão Kakadu) e o JPEG XL bem como, as métricas que utilizadas para avaliar a similaridade entre a imagem de referência e as imagens que sofreram codificação. Após estas comparações, os resultados obtidos serão comparados através destas métricas e concluí-se qual dos métodos é melhor e o porquê com base no que cada métrica avalia.

2.1 Imagem MFIF

A aquisição de imagens é um passo importante nas aplicações de processamento de imagens. A profundidade de campo (do inglês *DOF-depth of field*) limitada pelas lentes óticas torna difícil captar uma imagem em que todos os objetos presentes estejam focados ([6]). Para obter uma profundidade de campo maior, esta deve ser reduzida. Para aumentar a profundidade de campo de uma lente e melhorar a qualidade da imagem, é necessária uma técnica de baixo custo que ajude a obter uma imagem em que cada objeto esteja perfeitamente focado, o que pode ser conseguido através da integração de várias imagens com diferentes planos de foco. A imagem resultante, que é obtida através da fusão de imagens em que todos os objetos estão focados, é designada por Fusão de Imagens Multi Foco (MFIF).

O processo MFIF é ilustrado na Figura 2.1 e como pode ver-se este processo combina várias imagens adquiridas para obter uma única imagem fundida, designada por imagem de referência, que tem mais informação e um aspeto mais realçado. Por outras palavras, é uma técnica de fusão que tem como objetivo, combinar imagens de planos focados semelhantes de um cenário, para obter um resultado de fusão eficaz, o que significa que a imagem de referência tem um plano focado diferente. O MFIF permite obter imagens com uma representação e um detalhe mais precisos de um cenário em comparação com a imagem original. Um aspeto importante desta abordagem é o desenvolvimento de medições

eficazes dos níveis de atividade para calcular a nitidez da imagem de origem.

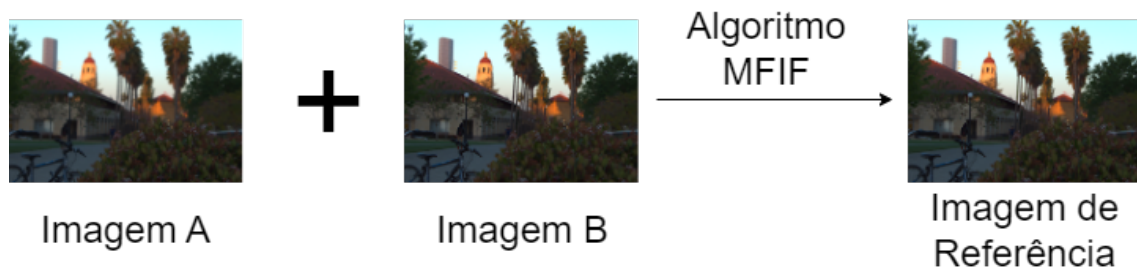


Figura 2.1: Processo da MFIF

No processo de fusão de imagens, do inglês *Image Fusion* (IF), um pré-requisito essencial é o registo das imagens, onde o principal objetivo deste registo é alinhar as imagens originais entre si. A partir da literatura existente [1], observou-se que as imagens originais utilizadas para o MFIF devem estar perfeitamente alinhadas, no entanto, conseguir isto é bastante desafiador em situações em tempo real.

Nos últimos anos, muitos esforços têm sido feitos por pesquisadores no campo do MFIF usando vários tipos de imagens multifoco, como imagens médicas [7], imagens microscópicas [8] e imagens reais com áreas desfocadas [9]. Ao longo dos anos, o número de artigos de investigação publicados no domínio da MFIF tem aumentado. Por conseguinte, pode concluir-se que a MFIF é um dos domínios mais recentes a considerar para a IF.

O principal objetivo do MFIF é obter todas as características e informações úteis das imagens adquiridas e desenvolver uma imagem nítida com DOF melhorado e todos os objetos focados. Por esta razão, o MFIF é referido como um ramo importante da IF, em que as imagens são fundidas de forma adequada para se adaptarem à máquina utilizada e à perceção humana. Estas imagens fundidas podem ser utilizadas para várias tarefas de processamento de imagem, como o reconhecimento de objetos, a extração de características e a segmentação [10]. A literatura mostra que o MFIF tem sido aplicado em vários domínios. Os exemplos incluem o processamento de imagens médicas [11], a fusão de micro-imagens [8], as redes de sensores visuais [9], a inspeção de patrulhas baseada na energia [12], a microscopia ótica [13] e a análise bioquímica [14].

A maioria dos métodos MFIF tem várias limitações, tais como, erros de registo, efeitos de desfocagem e elevada sensibilidade ao ruído. Em geral, as técnicas de MFIF podem ser classificadas nos domínios espacial e de transformação, mas surgiram várias técnicas que se adaptam simultaneamente às vantagens dos domínios espacial e de transformação [1]. Além disso, os métodos mais comuns atualmente utilizados para MFIF baseiam-se em *deep learning*, pelo que a classificação tradicional não é satisfatória para os requisitos existentes na literatura sobre MFIF.

Estes métodos são muito importantes devido ao *software* utilizado nos telemóveis. São também utilizados para gerar imagens de maior resolução, uma vez que a difusão é um conceito muito importante.

2.2 Classificação de métodos de fusão de imagem

De acordo com a literatura [7], a fusão de imagens (IF) ocorre normalmente a três níveis diferentes de informação:

1. Ao nível dos pixels, esta é a forma mais simples e a IF é efetuada ao nível mais baixo. A imagem única resultante é obtida considerando a média e a intensidade combinadas das duas imagens adquiridas [15]. São aplicadas funções matemáticas simples, como os valores médios e máximos, aos valores de pixels das imagens adquiridas para obter a imagem fundida [1]. As desvantagens deste tipo de fusão são a criação de um efeito de desfocagem e a suavização das arestas vivas da imagem fundida.
2. Ao nível da característica, é um intermédio entre os níveis de IF. Neste tipo de fusão, características como texturas e limites são extraídas da imagem adquirida e a fusão é realizada com base nessas características para atingir determinados critérios de seleção [1].
3. Ao nível da decisão, este tipo de técnica de IF funde informações a um nível elevado de abstração. Estas imagens são processadas separadamente para extrair informações das imagens adquiridas. Ao nível da decisão da IF, é obtida uma imagem fundida através da combinação de objetos da imagem original utilizando algoritmos de fusão adequados [16].

Com base nos métodos MFIF existentes, estes são classificados em quatro categorias: domínio espacial, domínio de transformação, *deep learning* e híbrido. A classificação do MFIF é apresentada na Figura 2.2.

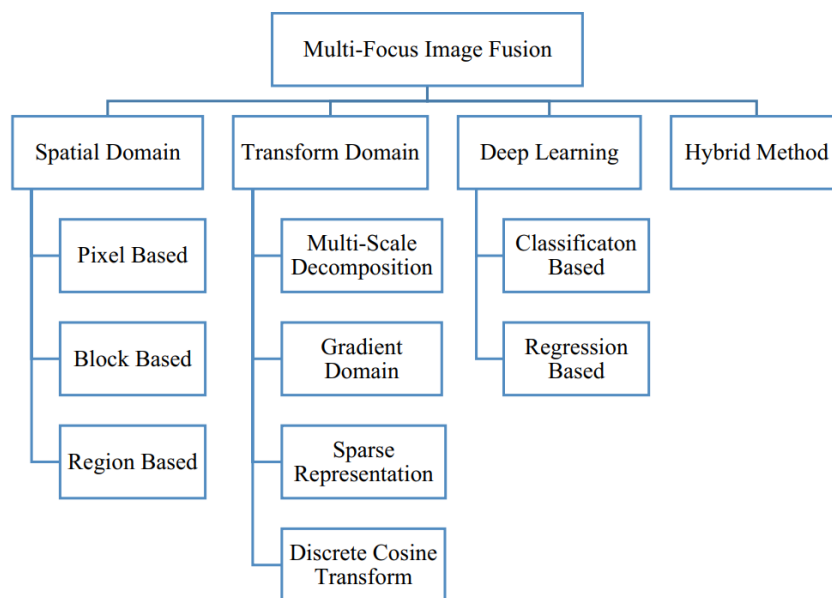


Figura 2.2: Classificação das técnicas de MFIF, extraído de [1]

2.3 Métodos no domínio espacial

No domínio espacial, este método funciona diretamente com as intensidades dos pixels [17]. Neste método, a fusão é efetuada diretamente nos pixels e não através da conversão para outra representação. Para obter o resultado pretendido no domínio espacial, os valores dos pixels das imagens adquiridas [18] são manipulados e os métodos dividem-se em três categorias: baseados em pixels, baseados no bloqueio e baseados em regiões [19]. O principal objetivo destes métodos é selecionar os pixels, blocos ou regiões com mais informação, respetivamente. As técnicas do domínio espacial são sensíveis ao ruído e têm certas limitações, como a desfocagem da imagem e a distorção espacial [16]. Os métodos típicos do domínio espacial sofrem de artefactos na fusão de imagens devido à identificação incorreta de regiões parciais [19]. Estes problemas podem ser resolvidos através de métodos no domínio da transformação.

2.3.1 Métodos baseados em pixels

A vantagem dos métodos baseados em pixels é que podem ser obtidos mapas de pesos exatos pixel a pixel, do inglês *pixel-wise weight maps*. Em muitos casos, o mapa de pesos é também identificado como um mapa de decisão porque o MFIF lida com pixels focados na imagem. Estes métodos dependem inteiramente da dimensão de foco para comparar a nitidez dos pixels presentes na imagem focada [20]. A principal desvantagem dos métodos tradicionais baseados em pixels é a sua elevada sensibilidade ao ruído e ao registo incorreto. Para melhorar a fiabilidade destes métodos, é necessário considerar uma série de vizinhanças espaciais na medição de foco. Vários métodos foram testados e melhorados ao longo dos anos. Em 2012 [21], foi apresentado um método de IF e de eliminação de ruído baseado em equações diferenciais parciais e na teoria da estimativa de erros. Este método foi eficaz para o melhoramento das arestas. O principal problema deste método era o facto de utilizar um método exaustivo de otimização de parâmetros.

Em 2015, o algoritmo MFIF baseado na transformada de recurso invariante da escala, do inglês *Scale Invariant Feature Transform* (SIFT) foi objeto de revisão numa revista [6]. Este método melhorou a qualidade das imagens adquiridas. O método produziu características de fusão melhoradas para regiões de movimento de objetos e limites de objetos. As principais desvantagens deste método são o facto de demorar mais tempo, ocupar mais memória e ser menos eficiente do ponto de vista computacional do que outros métodos. Na IF este método não é aplicável a exposições múltiplas ou à deteção remota.

Em 2018 [22], foi também introduzido um método para classificar os pixels mal identificados e são classificados em dois grupos - pixels agrupados e pixels dispersos. A filtragem probabilística e a correção de domínio foram utilizadas para classificar os pixels agrupados e dispersos, incorretamente identificados. A filtragem probabilística e de correção foi utilizada para classificar os pixels dispersos mal identificados e o método de correção do domínio foi utilizado para os pixels agrupados e mal identificados. Este método corrigiu

eficazmente os pixels mal identificados.

O método para MFIF baseado na desfocagem adaptativo do conteúdo, do inglês *Content Adaptive Blurring (CAB)* [23], ajudou a aumentar a profundidade de campo da imagem e produziu melhores resultados do que outros métodos. No mesmo ano [24], foi proposto um método de síntese de imagens multifoco baseado em caminhadas aleatórias, que ajudou a distinguir entre áreas focadas e desfocadas na imagem. Também ajudou a remover o ruído da imagem fundida e a reduzir o tempo de processamento. O método baseado na otimização do campo condicional aleatório, do inglês *Conditional Random Field (mfCRF)*[25] foi capaz de reter as características dos métodos multispectrais e do domínio espacial.

2.3.2 Métodos baseados em blocos

No caso dos métodos de blocos, as imagens originais são divididas em vários blocos de um determinado tamanho e o nível de atividade de cada bloco é medido. Estes sofrem do efeito de bloqueio, porque os blocos desfocados e focados estão presentes simultaneamente. O primeiro método de fusão de imagens baseado em blocos foi introduzido em 2001 [26]. Este método funde imagens utilizando frequências espaciais e tem sido útil para aplicações em tempo real. O método baseia-se na transformada de *wavelet*.

Além disso, foi apresentado um método [27] baseado em MPCNN para a fusão de imagens multifoco, que modifica as redes neurais acopladas a impulsos, do inglês *Pulse Coupled Neural Network (PCNN)* para reduzir o tempo de computação. Neste método, a energia de Laplacian, do inglês *Energy of Laplacian (EOL)* e a frequência espacial (SF) foram usadas como medidas de clareza.

Em 2015 [28], foi apresentado um método MFIF baseado no algoritmo de blocos e no algoritmo de evolução diferencial, em que as imagens foram divididas em blocos, e estes foram selecionados utilizando um critério de saliência para construir uma imagem fundida. Este método superou os métodos baseados no algoritmo genético e outros métodos MFIF.

Um método baseado no algoritmo da colônia artificial de abelhas (ABC) [29] encontra a solução ótima para a IF selecionando o bloco mais nítido da imagem original, e este método supera os métodos existentes.

2.3.3 Métodos baseados na região

No caso dos métodos baseados em regiões, estes foram desenvolvidos para segmentar as imagens originais de forma adaptativa. Os métodos baseados em regiões e em blocos espaciais têm uma estrutura semelhante, mas a principal diferença é que os métodos baseados em regiões medem o nível de atividade em regiões segmentadas de tamanho não uniforme. A desvantagem deste método é que sofre do efeito de bloqueio.

Em 2006 [30], foi introduzido um método IF baseado em PCNN, no qual a PCNN foi uti-

lizada para segmentar as imagens originais e as suas partes foram utilizadas para criar a imagem fundida. Este método é superior ao método de decomposição multi-escala e obtém melhores resultados com menos desfocagem, mesmo em imagens com registo impreciso e com movimento de objetos.

Em 2014 [31], uma abordagem de fusão usando a transformada *wavelet* de *quaternidade*, do inglês *quaternion wavelet transform* foi desenvolvida para MFIF, que superou muitos dos métodos existentes.

Além disso, foi apresentado um método MFIF baseado em regiões para imagens a cores no espaço de cores LUV [32]. Neste método, a técnica de segmentação por deslocamento do desvio médio foi usada para obter a sequência de imagens e verificou-se que este método funciona bem, embora a qualidade e a eficiência precisem de ser melhoradas.

Posteriormente [17], foi proposto o método de medição de focagem MFIF baseado na área da superfície da imagem original delimitada pelos pontos de intersecção. Este método permitiu distinguir entre áreas focadas e não focadas da imagem.

Além disso, em 2018 [33], foi desenvolvida uma técnica MFIF baseada na fusão de um algoritmo algébrico multi rede com o método de segmentação de bacias hidrográficas baseado em regiões. Isto ajudou a resolver os problemas que o MFIF mais frequentemente enfrenta, como o *ringing effect* e o baixo contraste. O método foi capaz de ter em conta o conteúdo de baixa e alta frequência para integrar duas ou mais imagens, mas não abordou o comportamento de avaliação desigual quando se combinam várias métricas de avaliação.

2.4 Métodos no domínio da transformada

No domínio da transformada, a imagem original é transformada noutra domínio de características em que pode ser fundida. A técnica de domínio da transformação é implementada através da transformação da imagem de origem no domínio da frequência e do cálculo da transformada de Fourier (FT). Uma vez obtida a FT da imagem de origem, a IF é completada e a FT inversa é efetuada para obter a imagem fundida. No entanto, os métodos no domínio da transformada requerem mais memória para execução e têm uma complexidade computacional mais elevada. Os métodos baseados no domínio da transformação são classificados como métodos de decomposição multiescala (MSD) do inglês *Multi-Scale Decomposition* (MSD), métodos baseados no domínio do gradiente do inglês *Gradient Domain* (GD), métodos baseados na representação dispersa, do inglês *Sparse Representation* (SR) e métodos baseados na DCT.

2.4.1 Métodos baseados na decomposição multiescala (MSD)

Os métodos baseados no MSD dominam o domínio das MFIF e são classificados em três categorias: métodos baseados em *wavelet*, em pirâmides e na análise geométrica multi-

escala, do inglês *Multi-Scale Geometric Analysis* (MGA). Estes métodos têm tido muito êxito no domínio das MFIF porque são adequados para representar características.

2.4.1.1 Métodos baseados na transformada da *wavelet*

Os métodos baseados na transformada de *wavelet* estão entre os métodos mais frequentemente mencionados na literatura. Isto deve-se ao facto de a transformada *wavelet* ser uma técnica de decomposição multi resolução e multiescala que permite focar todos os detalhes da imagem. É apresentado o método MFIF desenvolvido através da decomposição da integração *wavelet* com uma estratégia evolutiva [34]. Este método foi capaz de subtrair a região e preservar a informação das arestas usando altas frequências multiescala e informação multi resolução.

Em 2007 [35], foi introduzido um método IF baseado em regiões para a fusão de imagens multifoco utilizando transformadas *wavelet* baseadas na morfologia e algoritmos genéticos (GA). Este método faz a fusão ao nível do pixel e ao nível da característica para criar uma imagem fundida e nitidamente focada na região de interesse. Este método também é eficaz para imagens com registo incorreto. Mais tarde, foram comunicados métodos MFIF que utilizam transformadas de árvore dupla complexa, do inglês *Dual-tree complex wavelet transform*. A IF foi efetuada separadamente para componentes de baixa e alta frequência. Os problemas de baixo contraste e efeitos de bloqueio são resolvidos por este método.

Também é apresentada uma abordagem MFIF que utiliza a transformada *wavelet* complexa de árvore dupla com um classificador *fuzzy* [36], que não é afetada pela deslocação. Tem uma boa resolução de frequência e é útil para resolver o problema da fraca direcionalidade enfrentado pelas transformadas *wavelet* discretas.

Em 2011 [37], foi proposta uma técnica de fusão MFIF baseada em DWT que atinge a máxima nitidez, usando um algoritmo baseado na energia para coeficientes de baixa e alta frequência.

Em 2012 [38], foi introduzida uma abordagem MFIF adaptativa baseada em *wavelets*, na qual as propriedades de distribuição dos coeficientes *wavelets* são utilizadas para avaliar a qualidade da imagem. A principal limitação deste método é que a redundância e as regiões complexas de *wavelet* não são implementadas. De seguida, foi apresentado um novo método MFIF que utiliza características de contraste locais do domínio da transformada de contorno multiescala sem subamostragem, do inglês *multi-scale in Non-Sub-Sampled Contourlet transform* (NSCT) [39]. As regras IF de alta frequência são realizadas na NSCT para obter uma elevada precisão. O método apresenta melhores resultados para imagens multifoco e sem ruído.

O método baseado na transformada *wavelet Quaternion*, do inglês *Quaternion wavelet transform* (QWT) foi apresentado em 2013 [40], o que permite uma fusão de imagens eficiente utilizando as vantagens do domínio da transformada, bem como do domínio espacial. O principal objetivo deste método é classificar imagens multifoco com regiões suaves

e não suaves. O método reduz os blocos de artefactos e proporciona uma elevada precisão e baixas taxas de erro na deteção de regiões focadas.

Em 2014 [10], a técnica MFIF baseada em transformadas do domínio *wavelet* foi introduzida e proporcionou um melhor desempenho do que as técnicas existentes. Esta técnica foi utilizada para analisar imagens médicas micro óticas e forneceu resultados de fusão eficientes. A análise de contraste e a técnica MFIF baseada na frequência espacial [41] e na transformada *wavelet* estacionária, do inglês *stationary wavelet transform* (SWT) são apresentadas e este método é utilizado para imagiologia médica, obtendo-se melhores resultados tanto em imagens de ressonância magnética como de tomografia computadorizada. A desvantagem deste método é que exigia muito tempo de execução.

Em 2018 [42], outro método baseado na transformada *wavelet* de elevação discreta, do inglês *Discrete Lifting Wavelet Transform* (LWT) foi relatado.

Este método é um método MFIF altamente eficiente com a ajuda de um filtro gaussiano que também fornece informações espectrais e espaciais da imagem. A escolha da forma de onda correta para um determinado fim é uma tarefa difícil. Outra limitação da transformada *wavelet* é o facto de ser computacionalmente dispendiosa.

Em 2019 [43], foi desenvolvido um método IF baseado na tecnologia de super resolução híbrida, que é útil para obter melhores resultados, como melhor reconhecimento visual, menos distorção e bordas mais nítidas.

2.4.1.2 Método baseado em pirâmide

Na literatura, vários autores relatam diferentes tipos de estudos utilizando pirâmides para MFIF. É apresentado um método de IF baseado na função Gaussiana da pirâmide de Laplace, do inglês *Laplacian Pyramid* (LP) [44]. Este método é muito eficiente tanto na transformação como na reconstrução da imagem. O principal problema deste método é a seletividade da orientação e da posição na SF. Além disso [45], foi apresentado um método baseado no gradiente de pirâmide para FI. Neste método, o conceito de filtragem direcional é utilizado para fundir as imagens multifoco originais. Os resultados deste método são superiores aos obtidos utilizando a pirâmide de Laplace. De seguida [46], foi estudado o método baseado na pirâmide filtrar-subtração-dizimar, do inglês *Filter-Subtract-Decimate* (FSD) [46].

Em 2018 [47], foi relatado um método MFIF baseado na Região de mosaico em Pirâmides de Laplace, do inglês *Region Mosaicking on Laplacian Pyramids* (RMLP). Este método converte a IF baseada em pirâmide de pixels em IF baseada em região. A avaliação mostra que o RMLP é menos propenso a erros e mais robusto do que os métodos MFIF existentes, enquanto o RMPL é menos sensível ao ruído e reduz a distorção da cor nas imagens fundidas. De seguida, são apresentadas técnicas MFIF baseadas na RMLP que são menos sensíveis ao ruído e reduzem a distorção da cor das imagens fundidas [48]. Seguidamente, foi apresentado o método da transformada da pirâmide de Laplace, do inglês *Laplacian Pyramid Transform* (LPT) [49]. Neste método, a LPT é combinada com a PCNN e a

frequência espacial local, do inglês *Local Spatial Frequency* (LSF). Foram efetuados testes em imagens a cores e a cinzento e as imagens combinadas obtidas com este método forneceram mais detalhes do que as obtidas com métodos de fusão convencionais.

2.4.1.3 Método baseado em MGA

Os métodos baseados na *Wavelet* (MGA) têm muitas vantagens em relação a outros métodos de fusão de imagens multifoco, mas a informação captada direcionalmente no espaço 2D é a sua maior limitação no campo da MFIF. Os métodos baseados em MGA são muito importantes para resolver este problema.

Em 2014 [50], foi estudada uma abordagem MFIF baseada numa combinação de superfícies e PCNN sintéticas. As imagens de origem são decompostas em sub bandas utilizando superfícies e combinadas com PCNN que utilizam a informação útil nestas sub bandas. Este método fornece melhores resultados visuais e objetivos, em comparação com os métodos baseados em PCA, DWT e Laplace.

Além disso [51], é apresentado outro método IF que combina filtros guiados e *shearlet* complexo, do inglês *complex shearlet* (CST). O CST é utilizado para decompor a imagem, enquanto os filtros guiados processam os componentes de baixa frequência da imagem de origem. Este método baseado em CST ajuda a preservar a informação contida na imagem de origem. O CST é utilizado para processar os componentes de baixa frequência da imagem de origem. Isto melhora a continuidade da imagem resultante e o método tem mostrado bons resultados em avaliações subjetivas e objetivas. Os parâmetros adaptativos do filtro de direção ainda têm de ser tidos em conta, uma vez que desempenham um papel muito importante na IF.

2.4.2 Métodos baseados no domínio do gradiente (GD)

Nos métodos baseados em GD, as imagens originais são representada como um gradiente e apenas este é fundido e utilizado para reconstruir a imagem. O método baseado na preservação da saliência em GD foi relatado em 2007 [52]. Neste método, é gerado um mapa da saliência da imagem original e o gradiente com o valor da saliência mais elevado é adequadamente direcionado para o gradiente alvo. No entanto, este método não é aplicável ao melhoramento local de imagens.

Em 2014 [53], foi estudado um método baseado em GD ponderado multiescala para resolver o problema do registo incorreto e da desfocagem anisotrópica. As medições foco deste método mostram estruturas de canto e saliência de arestas. Os artefactos de recálculo que ocorrem nos métodos GD que utilizam tensores estruturais convencionais são eliminados.

Em 2016 [54], foi desenvolvido um método de seleção máxima baseado em IF para a fusão de imagens a cores em que os canais de cromaticidade (relativos a valores de cor) e luminância (relativos à luz branca e preta) são considerados separadamente. Os gradientes

são estimados para cada direção da imagem fundida.

2.4.3 Métodos baseados em representação dispersa (SR)

Os métodos baseados em SR são considerados fiáveis devido às propriedades fisiológicas do sistema visual humano. Nos métodos baseados na SR, o sinal é representado sob a forma de vários átomos para um determinado dicionário. Os métodos SR surgiram como uma área importante no domínio da MFIF.

Em 2015 [55], foi introduzido o método MFIF baseado na SR adaptativa, a ASR, do inglês *Adaptive Sparse Representation*. Este método supera os métodos tradicionais baseados em SR tanto quantitativamente como qualitativamente. Em 2016 [56], foi apresentado um método MFIF baseado em SR para construir dicionários totalmente adaptáveis. Neste método, vários sub-dicionários adaptativos são combinados utilizando o algoritmo de decomposição do valor singular K, do inglês *K-singular value decomposition* (K-SVD). A regra de fusão multi-padrão ponderada máxima é utilizada para a reconstrução da imagem.

Em 2019 [57], foi introduzido um método MFIF baseado na combinação de soluções ótimas e de SR conjunta. Neste método, o K-SVD é utilizado para obter um dicionário adaptativo que é combinado com um dicionário fixo. Por fim, o SR é utilizado para remover componentes redundantes do dicionário combinado. Este método obtém bons resultados de fusão, mas é computacionalmente demorado.

2.4.4 Método baseado na transformada discreta do cosseno (DCT)

A DCT é uma técnica no domínio da transformação que condensa a informação útil da imagem de origem num pequeno número de componentes de baixa frequência. Na literatura, vários autores relataram vários tipos de estudos usando transformadas discretas do cosseno para MFIF.

Um esquema baseado em DCT [58], onde a variância da DCT é calculada para obter um método de fusão em tempo real para MFIF é relatado. Este esquema melhora a qualidade das imagens fundidas e reduz a complexidade deste método em aplicações em tempo real. No entanto, os pontos fracos deste método são os limites e as áreas desfocadas.

Em 2014 [59], foi relatado um método MFIF baseado em DCT muito simples e eficiente em termos energéticos para a fusão de imagens multifoco, a principal desvantagem do DCT é a sua complexidade.

2.5 Métodos baseados em *deep learning*

Nos últimos anos, o *deep learning* tem suscitado grande interesse no domínio da MFIF. O *deep learning* caracteriza-se por um modelo que tenta aprender descrições hierárquicas

das características de uma imagem a vários níveis de abstração.

Os métodos baseados em *deep learning* podem ser classificados como métodos baseados na regressão. Alguns dos métodos baseados em redes neurais mencionados na literatura são descritos de seguida.

2.5.1 Métodos baseados em classificação

Uma abordagem baseada numa rede neuronal convolucional (CNN) para MFIF foi introduzida pela primeira vez em 2017 [11]. Neste método, foi criado um mapeamento direto entre a imagem original e o mapa de focagem, ao treinar o modelo CNN, as regras de fusão, bem como as medições do nível de atividade foram co-criadas, o que ajudou a eliminar os problemas enfrentados pelos métodos de fusão simultânea.

Em 2018 [60], um método MFIF baseado em p-CCN para fusão de imagens multifoco foi apresentado e treinado usando um conjunto de imagens de treino avançado. O método tem sido útil para melhorar a complexidade computacional enfrentada por várias técnicas MFIF. A principal desvantagem deste método é a necessidade de reduzir o tempo de execução. Além disso, consome mais memória em aplicações em tempo real. A implementação do método proposto para imagens multiexposição foi um grande desafio.

O método proposto baseia-se na deteção pixel a pixel e em áreas de foco [61]. Neste método, em vez de se aprenderem trajetórias da imagem, aprende-se toda a imagem. Foi aplicado em imagens em tons de cinzento e a cores e superou os métodos existentes, tanto em termos qualitativos como quantitativos.

Também nesse ano [62], foi comunicada uma abordagem MFIF baseada na integração de mapas de decisão modificados e de redes neurais de conexão de informação de impulso motivado (OI-PCNN), do inglês *Motivated Pulse Information Coupled Neural Networks* (OI-PCNN). A OI-PCNN é utilizada para fundir imagens e são utilizadas operações morfológicas para tomar a decisão final. Além disso [63], é apresentada uma CNN de características multinível, do inglês *Multi-Level Features CNN* (MLFCNN) baseada no método IF para imagens multifoco. Neste método, as características da camada anterior são transferidas para a camada seguinte. Os mapas iniciais são obtidos utilizando a MLFCNN de treino e são aplicadas operações morfológicas a estes mapas. No passo final, os mapas de decisão são utilizados para reconstruir a imagem fundida utilizando o método da soma dos pesos.

2.5.2 Método baseado na regressão

Os métodos baseados na regressão surgiram como um método importante no domínio das MFIF.

Em 2019 [64], foi introduzida uma rede dupla em forma de “U” para MFIF, e este método de fusão utilizava componentes fundidos, como a crominância e a luminância do espaço de cor YCbCr. A ideia subjacente a este método era desenvolver um método ba-

seado na rede em forma de “U”. No mesmo ano [65], foi apresentado um método MFIF baseado em redes neuronais convolucionais profundas de atenção visual multiescala, do inglês *method based on multi-scale visual attention deep convolutional neural network* (MADCNN) para resolver o problema da distinção exata entre regiões focadas e desfocadas da imagem original.

Em 2020 [66], foi introduzido um método IF baseado em CNN (IFCNN), no qual as características da imagem original são extraídas utilizando duas camadas convolucionais e, de seguida, estas características extraídas são combinadas para obter características fundidas. Por fim, estas características são reconstruídas para obter uma imagem final que contém informações mais precisas do que a imagem original. Além disso [67], foram desenvolvidos modelos de *deep learning* por regressão em pares, do inglês *Deep Regression Pair Learning* (DRPL) para MFIF. No modelo CNN tradicional de ponta a ponta, a imagem original é dividida em pequenos fragmentos cuja informação é divergente. No modelo apresentado, em vez de fragmentos, toda a imagem é convertida numa máscara binária que é depois fundida.

2.6 Método híbrido

No domínio das MFIF, os métodos do domínio espacial e do domínio da transformação têm as suas próprias vantagens e desvantagens. Os métodos de base híbrida combinam as vantagens dos métodos do domínio espacial e do domínio da transformação para obter métodos de IF mais eficientes que podem ultrapassar os problemas enfrentados pelos métodos tradicionais. Uma vez que o *deep learning* surgiu como um dos ramos mais rápidos da MFIF, vários métodos do domínio da transformação e do domínio espacial foram combinados com *deep learning* para obter melhores resultados de fusão.

Em 2016 [68], foram apresentadas duas tecnologias MFIF baseadas na distância multidirecional e na distância de vizinhança multi escala. Foram utilizados dois esquemas atualizados para melhorar o desempenho da fusão, mas a principal desvantagem desta abordagem é que não pode ser aplicada a sensores de IF remotos, de IF infravermelhos/visíveis e de IF médicos.

Em 2018 [69], um método de fusão de imagens com ruído em várias escalas que utiliza a representação de baixo nível, do inglês *Low-Rank Representation* (LRR) e funde componentes de baixa frequência através da manipulação de frequências espaciais, e também componentes de alta frequência utilizando coeficientes LRR. Como resultado, obtém-se o melhor desempenho de fusão para imagens com ruído e o método proposto supera outros métodos existentes, alcançando valores máximos de RMSE, PSNR e SSIM para diferentes tipos de ruído, como sal e pimenta, o ruído gaussiano e o ruído de Poisson. Para obter melhores resultados, foi desenvolvida a representação esparsa robusta combinada com PCNN adaptativa para MFIF [70]. Este método, tem um desempenho superior ao de ou-

tros métodos existentes, mas é necessário melhorar a eficiência. No domínio da NSCT, é apresentado um esquema MFIF baseado na segmentação ao nível da região focada e na PCNN [71]. Nesta técnica, a imagem fundida retém os pixels mais brilhantes em comparação com a imagem original e também fornece informações sobre as margens e os detalhes das regiões desfocadas.

Posteriormente, foi desenvolvido o esquema MFIF [22], que integra a técnica de domínio da transformada para além da técnica espacial. Este esquema combina as características úteis da DWT e da observação do olho humano.

2.7 Fusão de Múltiplas Imagens

A ASR ([72]) é apresentada para o processamento simultâneo de fusão de imagens e remoção de ruído. Com uma poderosa técnica de modelação de sinal, a representação dispersa (SR) tem sido utilizada com sucesso em muitas aplicações de processamento de imagem, como a remoção de ruído e a fusão. As aplicações tradicionais baseadas na SR exigem sempre uma elevada redundância no dicionário para satisfazer os requisitos de reconstrução do sinal, uma vez que a estrutura difere significativamente entre diferentes trajetos de imagem. Isto leva a elevados custos computacionais, bem como a artefactos visuais.

Em vez de aprender um único dicionário redundante, o modelo ASR aprende um subconjunto mais compacto de dicionários a partir de várias partes das imagens de alta qualidade pré-classificadas em várias categorias correspondentes à informação do gradiente. No processo de fusão e redução de ruído, um dos subdicionários é selecionado de forma adaptativa para um determinado conjunto de partes da imagem original.

Os resultados experimentais em conjuntos de imagens multifoco e imagens multimodais revelam que o método baseado em ASR supera o método convencional baseado em SR em termos de qualidade visual e avaliação objetiva.

As técnicas de fusão de imagens têm como objetivo produzir uma imagem composta através da integração de informação complementar de várias imagens de origem do mesmo cenário [73]. A imagem original pode ser obtida a partir de um sensor de imagem cujos parâmetros óticos podem ser ajustados para diferentes tipos de sensores. A imagem final pode ser mais adequada para a perceção humana e para o processamento por máquinas do que uma imagem de origem única.

O estudo dos problemas de ruído na fusão de imagens é particularmente importante porque as imagens são frequentemente corrompidas por ruído durante a sua aquisição ou transmissão. No modelo ASR, a fusão de imagens e a remoção de ruído são oferecidas simultaneamente.

Para a maioria das tarefas de visão computacional, como a deteção e identificação de objetos, é desejável utilizar imagens focadas como entrada em vez de imagens desfocadas. No entanto, devido a um DOF limitado da câmara, é muitas vezes difícil captar diretamente uma imagem perfeitamente focada. Surge então, a fusão de imagens multifoco,

esta técnica de combina duas ou mais imagens originais parcialmente focadas numa imagem totalmente focada, é de grande importância no domínio da visão computacional e do processamento de imagens e tem recebido muita atenção nos últimos anos.

2.8 Codificação de Imagens

A norma JPEG estabelece que o codec define a forma como uma imagem é comprimida num fluxo de bytes e descomprimida novamente numa imagem, mas não define o formato do ficheiro utilizado para conter esse fluxo. Os três codecs que foram utilizados para desenvolver este trabalho são, portanto, descritos em pormenor abaixo, sendo eles o JPEG, o JPEG 2000 (kakadu) e o JPEG XL.

2.8.1 JPEG

Uma imagem digital a cores é um conjunto de pixels, cada um dos quais representa um vetor de cor tridimensional (3D) ([2]). Os elementos do vetor definem a cor dos pixels em relação a um espaço de cor selecionado, como RGB ou YCbCr. O JPEG, do inglês *Joint Photographic Experts Group* é frequentemente utilizado como norma para a compressão de imagens digitais a cores. A compressão JPEG utiliza a quantização da Transformada Discreta de Coseno (DCT) e três coeficientes de imagem no plano de cor e é apresentada esquematicamente na Figura 2.3. No entanto, as várias definições para a compressão e descompressão JPEG não estão normalizadas.

As seguintes definições JPEG podem ser selecionadas pelo utilizador ou pelo processador de imagem:

1. Um espaço de cor utilizado para comprimir imagens em três planos de cor, independentemente uns dos outros;
2. A compressão utiliza a subamostragem para cada cor de cada plano e a compressão utiliza a interpolação de fase;
3. As tabelas de quantização são utilizadas para comprimir cada plano de cor.

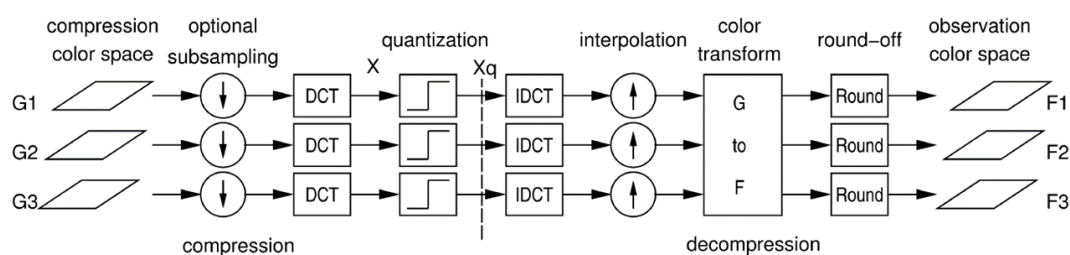


Figura 2.3: Visão geral da compressão e descompressão da JPEG, extraído de [2]

A percepção das cores é a sensação que ocorre quando a luz estimula a retina humana. Esta definição é bastante redundante, uma vez que existem apenas três tipos de recetores na retina humana.

Estamos acostumados a pensar nas imagens como bidimensionais, mas na verdade elas têm várias outras dimensões, como a cor, o tempo (movimento) e profundidade (tridimensional e pares estéreo) podem acrescentar dimensões adicionais para a informação de uma imagem, e se for considerada uma imagem como quebrada, a questão de dimensionalidade torna-se ainda mais complexa ([74]). Por definição, a JPEG diz respeito principalmente a imagens que têm duas dimensões espaciais. Contêm informação em escala de cinzentos e de cores e não têm dependência temporal.

2.8.2 JPEG 2000

Uma implementação eficiente e fiel do standard JPEG 2000 é uma tarefa desafiante. Por esta razão, foi desenvolvido o sistema de programação Kakadu. Este fornece uma base muito ampla e sólida para a construção de aplicações usando o JPEG 2000 e poupa anos de desenvolvimento e esforço [75].

O programa Kakadu oferece assim um conjunto completo de ferramentas aos utilizadores que exigem um desempenho sem paralelo na compressão e manipulação de todos os tipos de imagens, incluindo codificação de vídeo, gestão de imagens de satélite, arquivo de imagens e imagens médicas [76].

Os subsistemas mais importantes da arquitetura de núcleo do Kakadu são apresentados na Figura 2.4 [3]. As linhas tracejadas no diagrama mostram os objetos de interface utilizados para a interação com a aplicação interna. Na maioria dos casos, os objetos de interface não têm espaço de armazenamento e servem principalmente para separar os subsistemas uns dos outros e da aplicação. A direção das setas nestas linhas não implica o fluxo de dados, mas sim a prestação de serviços. O diagrama mostra apenas uma pequena parte da interação. O Kakadu fornece suporte para o formato de ficheiro JPEG 2000 e fornece uma série de objetos e subsistemas baseados no núcleo que interagem com o servidor. Em particular, fornece suporte destinado a facilitar a ligação dos programadores à interface da aplicação Kakadu.

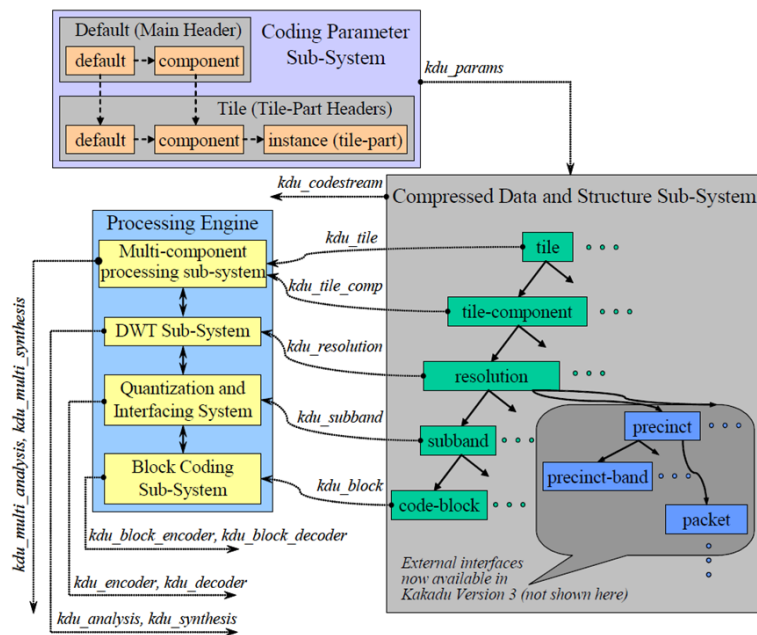


Figura 2.4: Arquitetura do sistema central(core) Kakadu, extaido de [3]

Na JPEG 2000, os parâmetros de codificação são definidos em intervalos progressivamente mais estreitos ao longo da hierarquia de especificações. A um nível mais elevado, a maioria dos parâmetros de codificação são normalmente definidos no cabeçalho principal, isto aplica-se a todos os componentes de todas as partes da imagem, exceto se for explicitamente substituído.

O fluxo de código do cabeçalho principal do JPEG 2000 pode também conter substituições de certos componentes para os parâmetros de codificação por defeito. O cabeçalho de bloco pode substituir a especificação padrão de componentes individuais ou apenas componentes de bloco individuais.

Finalmente, os parâmetros de codificação permitem que a informação seja atualizada de forma incremental em diferentes partes do bloco. Esta estrutura e os mecanismos adequados de herança de parâmetros refletem-se no subsistema de parâmetros de codificação. É importante notar que o Kakadu fornece uma interface mais consistente para os parâmetros de codificação, uma vez que o JPEG 2000 tem muitas irregularidades na sua sintaxe e o Kakadu omite-as do utilizador ou programador de aplicações.

2.8.3 JPEG XL

O JPEG XL é uma norma de codificação de imagem que abrange a captura, criação, armazenamento e distribuição de imagens fotográficas [4]. Este codec proporciona uma compreensão eficaz e fiável das características de transição, com elevada qualidade visual. O sistema de codificação de imagem JPEG XL é rico em funcionalidades em comparação com outros codecs e foi concebido para satisfazer as necessidades da fotografia profissional, arte digital, impressão, aplicações médicas e científicas, bem como distribuição de

imagens. O JPEG XL suporta uma vasta gama de cores, uma elevada gama dinâmica e imagens com elevada profundidade de bits e inclui funcionalidades como animação, camadas e imagens reduzidas.

O JPEG XL oferece uma melhoria significativa na qualidade da imagem quando comparado com os JPEGs mais antigos. Em comparação com as alternativas modernas, o JPEG XL tem uma especificação mais curta e uma implementação mais simples, mas oferece um desempenho de compressão igual ou superior. Foi concebido para proporcionar uma codificação e decodificação computacionalmente eficientes em dispositivos móveis, utilizando uma implementação de programa sem aceleração de *hardware* adicional.

Além disso, o JPEG XL inclui várias características para ajudar na migração de formatos de codificação JPEG antigos. Os formatos JPEG existentes podem ser convertidos sem perdas para ficheiros JPEG XL, resultando numa redução significativa do tamanho, como mostra a Figura 2.5. Estas imagens podem ser reconstruídas dentro do mesmo ficheiro JPEG, assegurando a compatibilidade com as aplicações existentes. Tanto a transcodificação como a reconstrução são computacionalmente eficientes.

Uma vez que o servidor pode armazenar um único ficheiro JPEG XL para suportar tanto clientes JPEG como JPEG XL, é fornecido um caminho de migração sem problemas de plataformas JPEG antigas para plataformas JPEG XL modernas.

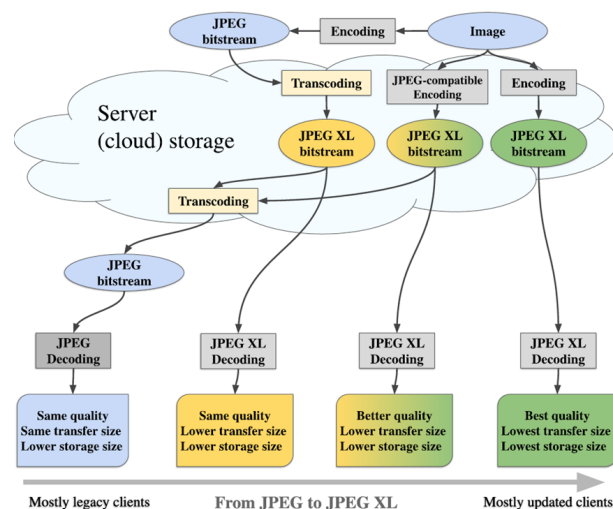


Figura 2.5: Cenário exemplo, extraído de [4]

O codec JPEG XL foi concebido para satisfazer três critérios de aplicação principais: fidelidade da imagem original (muito próxima da percepção humana), velocidade de codificação e decodificação e rácio de compressão, que varia normalmente entre 20:1 e 50:1.

Relativamente à eficiência da compressão, os resultados básicos são os seguintes: a transcodificação JPEG sem perdas reduz o tamanho do JPEG em cerca de 16-22%. A compressão sem perdas de pixels é cerca de 35% mais pequena que o PNG otimizado, para imagens de 8 bits e ainda melhor para imagens com maiores profundidades de bits (com por exemplo 10 ou 12 bits). Comparado com outros codecs, o JPEG XL oferece uma compressão superior com uma qualidade de imagem inalterada.

2.8.3.1 Arquitetura do Codec

O JPEG XL tem dois modos de codificação, que podem ser vistos no diagrama de codificação na Figura 2.6.

O VarDCT (verde) é uma codificação fotográfica sem perdas centrada no sistema visual humano e pode também ser utilizada para a transcodificação sem perdas de imagens JPEG existentes.

O Modular (vermelho) é uma codificação matemática sem perdas ou com perdas genéricas. O mesmo que DC na JPEG, mas os blocos só são utilizados para imagens 8x8, os campos de controlo (por exemplo, para correlação e filtragem de cores) e as características da imagem são extraídas. A Transcodificação do fluxo de bits JPEG sem perdas (reconstrução não só dos dados de imagem, mas também dos ficheiros com bits exatos).

Opcionalmente, o fluxo de bits dos dados de reconstrução JPEG pode ser guardado (cor amarela). A imagem original comprimida com a JPEG XL é codificada com um fluxo de código com ou sem perdas e descodificada numa imagem que é idêntica ou percentualmente semelhante à imagem final. O fluxo de bits de entrada da JPEG contém um fluxo de código que utiliza um subconjunto de meios de codificação do modo VarDCT (para representar dados da imagem) e um fluxo de código da JPEG para reconstrução de dados, que pode ser descodificado numa reconstrução da JPEG idêntica do fluxo de dados JPEG XL.

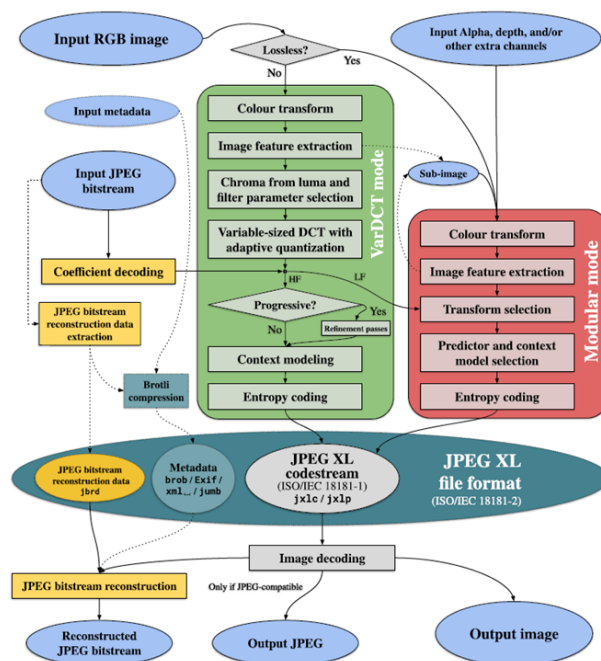


Figura 2.6: Arquitetura do Codec, extraído de [4]

2.9 Métricas de desempenho

Atualmente, a codificação é uma tecnologia indispensável na nossa sociedade e é utilizada milhares de milhões de vezes por dia por uma parte significativa da população mundial [77]. Isto inclui não só fotografias pessoais amplamente partilhadas nas redes sociais, mas também aplicações e serviços profissionais, como bancos de imagens e fluxos de vídeo (por exemplo, capas de filmes). A maioria das fotografias pessoais é adquirida em dispositivos móveis, as imagens ocupam uma porção significativa do espaço de armazenamento e são a principal razão para a compra de novos dispositivos.

Nos últimos anos, tem-se registado um aumento significativo do número de pedidos de análise de aplicações de imagiologia automatizadas [78]. Muitas destas aplicações requerem um processamento de imagem muito rápido e muitas vezes exigem modificações nos algoritmos da aplicação para reduzir o tempo de execução.

Na maioria dos casos, essas modificações conduzem a um certo erro nos resultados obtidos. Para determinar se uma determinada aplicação é adequada, a qualidade dos resultados deve ser cuidadosamente analisada, comparando as imagens obtidas (distorcida) e as desejadas (não distorcidas), de modo a chegar a um compromisso entre a velocidade do algoritmo e a qualidade dos resultados.

Muitas métricas de distorção utilizadas para avaliar a semelhança entre duas imagens pertencem ao grupo das métricas baseadas em pixels, que medem a diferença entre os pixels correspondentes a cada posição nas imagens sem distorção e distorcidas. As métricas de distorção mais utilizadas no processamento de imagens são a relação sinal/ruído, do inglês *Signal to Noise Ratio* (SNR) e a relação sinal/ruído de pico, *Peak Signal to Noise Ratio* (PSNR).

De seguida, são descritos em pormenor três métricas, nomeadamente o PSNR-HVS-M, o MS-SSIM (semelhança estrutural multiescala, do inglês *Multiscale structural similarity*) e o FSIMc (índice de semelhança de características, do inglês *feature similarity index*).

2.9.1 PSNR-HVS-M

O PSNR é uma métrica usada para avaliar a relação do sinal de pico para o ruído, mas apresenta um mau desempenho na distinção estrutural do conteúdo nas imagens [79]. Isto ocorre porque são aplicados vários tipos de degradação à mesma imagem.

Foram então desenvolvidas outras métricas baseadas no PSNR, sendo elas o PSNR-HVS e PSNR-HVS-M, que tentam colmatar o mau desempenho do PSNR.

O PSNR-HVS-M ([80], [81]), é uma versão do PSNR que utiliza um modelo de qualidade simples e eficaz, baseado no sistema visual humano, do inglês *human visual system* (HVS) que utiliza funções DCT básicas. Este modelo funciona com um bloco de pixels 8x8 e calcula a distorção máxima que não é visível devido ao coeficiente de máscara, do inglês *coefficient masking*. A métrica PSNR-HVS-M proposta tem em conta o modelo proposto e a função de sensibilidade ao contraste, do inglês *contrast sensitivity function* (CSF).

2.9.2 MS-SSIM

O MS-SSIM [82] é uma das métricas mais conhecidas que calculam valores de qualidade relativa entre uma imagem de referência e uma imagem distorcida, comparando pormenores como a resolução e proporcionando um elevado desempenho para codecs baseados na aprendizagem. O MS-SSIM é mais flexível do que os métodos de escala única, como o SSIM, porque incorpora alterações na resolução da imagem e nas condições de visualização da imagem. Além disso, esta métrica introduz uma abordagem baseada na síntese de imagens para calibrar um parâmetro que pondera a importância relativa entre diferentes escalas.

Assim, o algoritmo MS-SSIM [83] surgiu para melhorar o desempenho da avaliação em concordância com a percepção humana, sendo a sua equação:

$$\text{MS-SSIM}(A, B) = [l_M(A, B)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(A, B)]^{\beta_j} [s_j(A, B)]^{\gamma_j} \quad (2.1)$$

Onde o M é a escala máxima da imagem, para a escala j , a semelhança de luminosidade, do contraste e da estrutura de A e B são denotados por $l_j(A, B)$, $c_j(A, B)$ e $s_j(A, B)$, respetivamente. Quanto maior for este valor, melhor será a qualidade da imagem, a nível estrutural.

2.9.3 FSIMc

O FSIMc [84] é uma métrica de semelhança de características baseada no cálculo de duas características de baixo nível que desempenham papéis complementares na caracterização da qualidade da imagem. A congruência de fase, do inglês *phase congruency* (PC) é uma característica sem dimensão que descreve a importância da estrutura local, enquanto a magnitude dos gradientes da imagem, do inglês *gradient magnitude* (GM) descreve a informação de contraste. Assim o algoritmo FSIM [84] é calculado através da seguinte equação:

$$\text{FSIM} = \frac{\sum_{\mathbf{x} \in \Omega} S_L(\mathbf{x}) \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x} \in \Omega} PC_m(\mathbf{x})}$$

Onde, $S_L(\mathbf{x})$ representa o mapa de similaridade local e o $PC_m(\mathbf{x})$ representa a congruência de fase.

A métrica FSIMc foi baseada na métrica FSIM, acrescentando o parâmetro $S_C(\mathbf{x})$, obtendo a seguinte equação:

$$\text{FSIM}_C = \frac{\sum_{\mathbf{x} \in \Omega} S_L(\mathbf{x}) \cdot [S_C(\mathbf{x})]^\lambda \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x} \in \Omega} PC_m(\mathbf{x})}$$

Onde, $S_C(\mathbf{x})$ representa a similaridade de cromaticidade, do inglês *chrominance similarity* e $\lambda > 0$ é o parâmetro que permite ajustar a importância das componentes cromáticas. Quanto mais elevados forem os valores desta métrica, maior será a qualidade da imagem avaliada.

Capítulo 3

Metodologia

Neste capítulo, descreve-se a base de imagens utilizada e destacam-se as imagens de referência da base de imagens utilizadas no estudo. Os procedimentos utilizados são depois descritos em pormenor, desde a preparação da imagem até à aplicação do método de fusão. A taxa de bits é explicada em relação aos cálculos e às análises pertinentes. Por fim, é apresentado um exemplo prático para ilustrar as etapas efetuadas, fornecendo uma visão clara das etapas realizadas ao longo do procedimento prático.

3.1 Base de Dados

Foi utilizada a base de dados de imagens MFIF Real [5] que contém imagens multifoco.

1. Tipos de imagens:

- Contém imagens multifoco realistas.
- Contém vários cenários incluindo: edifícios, plantas, centros comerciais, praças e outras paisagens.

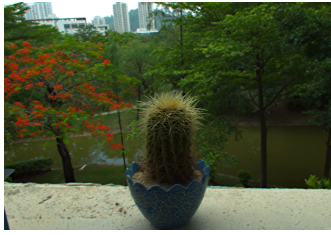
2. Características da base de imagens:

- Apresenta 710 pares de imagens diferentes.
- Cada par de imagens tem duas imagens (A e B) com pontos diferentes de focagem. Tem também uma imagem de referência (denominada de imagem de referência da base de imagens), apesar desta não ser necessária mas, conveniente para a avaliação pretendida.
- Estas imagens não sofreram codificação.

3. Imagens selecionadas:

- Dos 710 pares de imagens foram selecionados 18 pares que abrangessem o maior número possível de cenários e diferentes características das imagens.
- Um conjunto maior de imagens, seria redundante.

Este conjunto de imagens foi criado com uma câmara *lighy field* chamda Lytro illum, que utiliza um campo de luz para criar as imagens. A Figura 3.1 mostra as 18 imagens de referência da base de imagens selecionadas.



(a) Imagem Original (001)



(b) Imagem Original (028)



(c) Imagem Original (086)



(d) Imagem Original (091)



(e) Imagem Original (101)



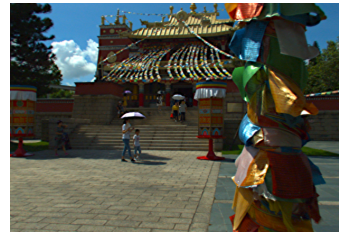
(f) Imagem Original (122)



(g) Imagem Original (138)



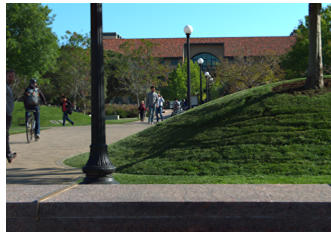
(h) Imagem Original (155)



(i) Imagem Original (186)



(j) Imagem Original (224)



(k) Imagem Original (271)



(l) Imagem Original (276)



(m) Imagem Original (291)



(n) Imagem Original (427)



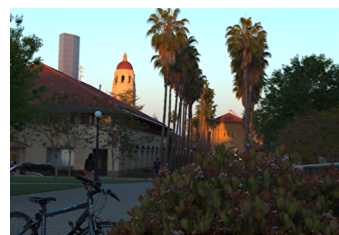
(o) Imagem Original (546)



(p) Imagem Original (621)



(q) Imagem Original (648)



(r) Imagem Original (694)

Figura 3.1: Imagens de referência da base de imagens escolhidas para o trabalho [5]

3.2 Procedimentos utilizados

A Figura 1.1 esquematiza os processos implementados, nomeadamente o de codificação/-descodificação com posterior fusão (Processo 1) e o de fusão com posterior codificação/-descodificação (Processo 2).

1. Processo 1:

- Às imagens A e B aplica-se o algoritmo de fusão, obtendo uma imagem fundida. Esta imagem vai ser a imagem de referência usada para as comparações, à qual foi denominada de imagem de referência.
- Codificação/descodificação da imagem fundida, obtendo a imagem fundida comprimida codec.

2. Processo 2 (setas a vermelho):

- Comprimir as imagem A e a imagem B separadamente, obtendo o que no esquema é denominado de codec A e codec B.
- Fundir a imagem codec A e codec B através do algoritmo de fusão, obtendo a imagem codificada fundida.

3. Comparação com a imagem de referência:

- Ambas as imagens obtidas pelos dois processos (uma obtida pela compressão e posterior fusão e a outra pela compressão com posterior fusão), são comparadas através de métricas objetivas com a imagem de referência. Para além disto, as imagens obtidas também são comparadas com a imagem de referência da base de imagens.
- As métricas utilizadas são o PSNR-HVS-M, MS-SSIM e o FSIMc.

Para a realização deste procedimento os codecs utilizados para a compressão foram o JPEG, o JPEG 2000 e o JPEG XL, e o algoritmo de fusão usado foi a ASR.

3.3 Taxas de *bitrates* (BPP)

Para cada codificador, foram usadas cinco taxas alvo de bits diferentes, sendo elas: 0,30; 0,56; 1,00; 1,78 e 3,00.

Por definição o conjunto de taxas de bits desejadas para avaliações subjetivas é um subconjunto das taxas de bits desejadas para avaliações objetivas e depende da complexidade espacial das imagens de teste [85]. Assim sendo, as taxas de bits mencionadas contabilizam o número total de bits no arquivo ou nos arquivos codificados, para o descodificador

que pode reconstruir uma versão com perdas de toda a imagem. A taxa de bits é o número de bits por pixel (bpp), definido através da seguinte expressão:

$$\text{BPP} = \frac{\text{N_TOT_BITS}}{\text{N_TOT_PIXELS}}$$

Onde, o N_TOT_BITS é o número de bits para a representação compacta da imagem e N_TOT_PIXELS é o número de pixels na imagem reconstruída.

Isto permite comparar degradações diferentes da mesma imagem com a imagem de referência, para avaliar até que ponto é que o olho humano detetava que a imagem tinha sofrido codificação/descodificação.

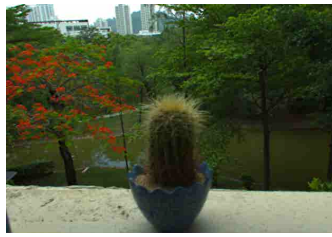
3.4 Exemplo dos procedimentos utilizados

Como se pode observar na figura 3.2, estão alguns exemplos das imagens que sofreram primeiro compressão e depois fusão, sendo que na primeira linha foi usado o codificador JPEG, na segunda linha o codificador JPEG 2000 e na terceira linha o codificador JPEG XL, bem como as que sofreram primeiro fusão e depois codificação, na figura 3.3, em que na primeira linha se encontra a imagem A e B e a imagem fundida em terceiro lugar, e na linha abaixo encontram-se os mesmos codificadores acima mencionados pela mesma ordem, respetivamente.

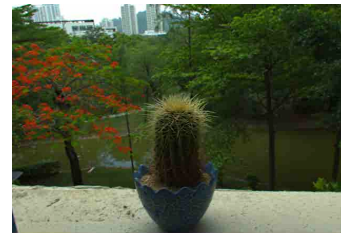
Depois de concluídos estes processos, foram aplicadas as métricas PSNR, MS-SSIM e FSIMc para determinar qual das imagens tem melhor qualidade quando comparada com a imagem de referência.



(a) *Imagem A codec (001)*



(b) *Imagem B codec (001)*



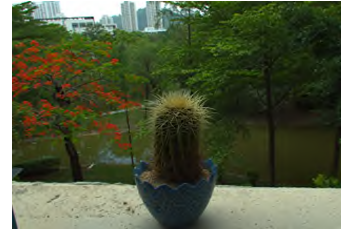
(c) *Imagem Fundida JPEG (001)*



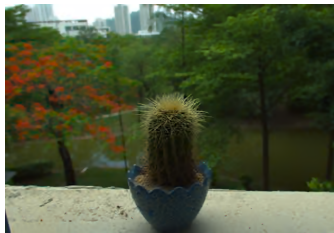
(d) *Imagem A codec (001)*



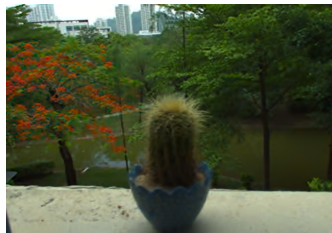
(e) *Imagem B codec (001)*



(f) *Imagem Fundida JPEG2000 (001)*



(g) *Imagem A codec (001)*



(h) *Imagem B codec (001)*

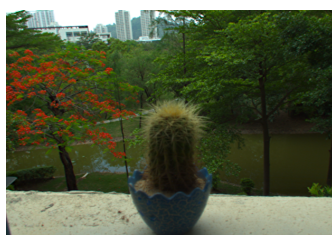


(i) *Imagem Fundida JPEGXL (001)*

Figura 3.2: Imagens codificadas com posterior fusão



(a) *Imagem A codec*



(b) *Imagem B codec*



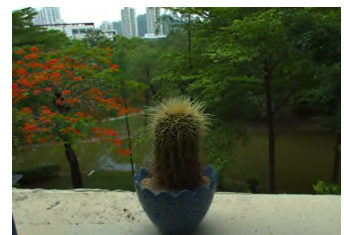
(c) *Imagem Fundida*



(d) *Imagem Fundida JPEG*



(e) *Imagem Fundida JPEG 2000*



(f) *Imagem Fundida JPEGXL*

Figura 3.3: Imagens fundidas com posterior codificação

Capítulo 4

Resultados

Este capítulo irá servir para reportar os resultados obtidos com este trabalho, começando pelas imagens que sofreram codificação e decodificação antes da fusão e de seguida as que sofreram codificação e decodificação depois da fusão.

4.1 JPEG, JPEG 2000, JPEG XL

Como já foi referido, foi utilizado o codificador JPEG, neste caso o *software* de referência utilizado foi ISO/IEC 10918-1 | ITU-T Rec. T.81 [85], para um intervalo com cinco taxas de bits.

Assim como no codec JPEG, para o codec JPEG 2000 (kakadu) o *software* usado foi o kakadu versão 8.4.1 [86]. Também foi usado um intervalo com cinco taxas de bits, para cada codec, com se pode observar na Figura 4.4, 4.5 e 4.6.

Tal como no codec JPEG, e no JPEG 2000 (kakadu), para o codec JPEG XL foi usado um intervalo com cinco taxas de bits, para cada codec como se pode observar nas Figuras 4.1, 4.2 e 4.3, *software* utilizado foi a versão 0.10.3 [87], 4.7, 4.8 e 4.9 que apresentam as imagens 086, 155 e 621 para os codificadores mencionados e para as taxas de bits referidas.

4.1.1 Codificação/descodificação com posterior fusão

Primeiramente, é apresentada a imagem de referência da base de imagens e ao seu lado a imagem de referência fundida usando o método ASR. Abaixo, encontra-se em cada linha a Imagem A e B, e a respetiva imagem fundida para as taxas de bits 0,30; 0,56; 1,00 e 3,00. As imagens para a taxa de bits correspondente a 1,78 não são mostradas uma vez que não se conseguem observar diferenças quando comparadas com a taxa de bits de 3,00. Apenas as imagens da terceira coluna (imagem fundida), para cada taxa de bits será utilizada para avaliar a sua qualidade.

4.1.2 Fusão com posterior codificação e decodificação

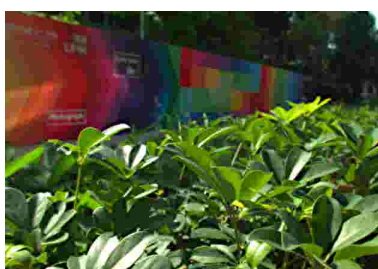
Neste caso, em primeiro lugar foi feita a fusão das imagens A e B (primeira linha) e de seguida foram usados os codecs para a mesma Imagem Fundida. Assim sendo, na primeira linha encontra-se a imagem de referência da base de imagens. Na linha seguinte o processo de fusão da Imagem A e B e a Imagem Fundida, e nas linhas abaixo encontram-se



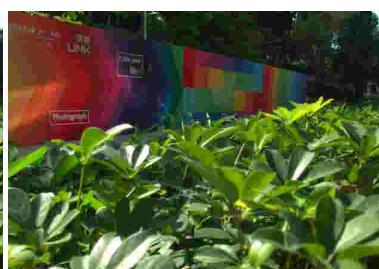
(a) Imagem de referência da base de imagens



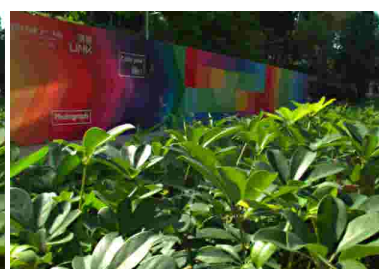
(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.1: Codificação/descodificação usando o JPEG para a Imagem 086 com posterior fusão



(a) Imagem de referência da base de imagens

(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.2: Codificação/descodificação usando o JPEG para a Imagem 155 com posterior fusão



(a) Imagem de referência da base de imagens



(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.3: Codificação/descodificação usando o JPEG para a Imagem 621 com posterior fusão



(a) Imagem de referência gt

(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)

(d) Imagem B (taxa de bits 0.3)

(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)

(g) Imagem B (taxa de bits 0.56)

(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)

(j) Imagem B (taxa de bits 1.0)

(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)

(m) Imagem B (taxa de bits 3.0)

(n) Imagem fundida (taxa de bits 3.0)

Figura 4.4: Codificação/descodificação usando o JPEG 2000 para a Imagem 086 com posterior fusão



(a) Imagem de referência gt



(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.5: Codificação/descodificação usando o JPEG 2000 para a Imagem 155 com posterior fusão



(a) Imagem de referência gt

(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)

(d) Imagem B (taxa de bits 0.3)

(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)

(g) Imagem B (taxa de bits 0.56)

(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)

(j) Imagem B (taxa de bits 1.0)

(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)

(m) Imagem B (taxa de bits 3.0)

(n) Imagem fundida (taxa de bits 3.0)

Figura 4.6: Codificação/descodificação usando o JPEG 2000 para a Imagem 621 com posterior fusão



(a) Imagem de referência da base de imagens



(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.7: Codificação/descodificação usando o JPEG XL para a Imagem o86 com posterior fusão



(a) Imagem de referência da base de imagens

(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.8: Codificação/descodificação usando o JPEG XL para a Imagem 155 com posterior fusão



(a) Imagem de referência da base de imagens

(b) Imagem de referência



(c) Imagem A (taxa de bits 0.3)



(d) Imagem B (taxa de bits 0.3)



(e) Imagem fundida (taxa de bits 0.3)



(f) Imagem A (taxa de bits 0.56)



(g) Imagem B (taxa de bits 0.56)



(h) Imagem fundida (taxa de bits 0.56)



(i) Imagem A (taxa de bits 1.0)



(j) Imagem B (taxa de bits 1.0)



(k) Imagem fundida (taxa de bits 1.0)



(l) Imagem A (taxa de bits 3.0)



(m) Imagem B (taxa de bits 3.0)



(n) Imagem fundida (taxa de bits 3.0)

Figura 4.9: Codificação/descodificação usando o JPEG XL para a Imagem 621 com posterior fusão

os codecs JPEG, JPEG 2000 (kakaду) e JPEG XL, em cada coluna respetivamente, para as taxas de bits 0,30; 0,56; 1,00 e 3,00 em cada linha.

A imagens para a taxa de bits correspondente a 1,78 não são mostradas uma vez que não se consegue observar diferenças quando comparadas com a taxa de bits de 3,00. Como se pode observar na Figura 4.10, 4.11 e 4.12.

4.2 Análise de Resultados

De seguida, foram usadas as métricas PSNR, MS-SSIM e FSIMc, para fazer a comparação das imagens obtidas (com codificação/descodificação com posterior fusão ou com fusão e posterior codificação/descodificação) com a imagem de referência e com a imagem de referência da base de imagens. Abaixo encontram-se esses gráficos.

A primeira linha corresponde à métrica PSNR, a linha seguinte corresponde à métrica MS-SSIM e terceira linha corresponde à métrica FSIM, cada coluna refere-se respetivamente aos codecs estudados, pela seguinte ordem JPEG, JPEG 2000 e JPEG XL.

A curva azul corresponde à comparação das imagens codificadas com posterior fusão com a imagem de referência da base de imagens (denominada de *ground true*). A curva laranja corresponde à comparação das imagens codificadas com posterior fusão com a imagem de referência (denominada de *generated image*). A curva amarela corresponde à comparação das imagens fundidas com posterior codificação com a imagem de referência da base de imagens (denominada de *ground true*). Por fim, a curva lilás corresponde à comparação das imagens fundidas com posterior codificação com a imagem de referência (denominada de *generated image*). Toda a descrição apresentada é diretamente aplicável aos gráficos analisados a seguir. Através dos gráficos, é possível verificar que as curvas laranja e lilás, têm um melhor desempenho que as curvas azul e amarela. É possível inferir também, que o método que permite maior nível de qualidade é aquele em que primeiro há codificação/descodificação e só depois é que ocorre a fusão da imagem.

4.2.1 Gráficos de análise das métricas (global)

Nesta secção são apresentados os gráficos das métricas estudadas para as restantes imagens que foram analisadas também. No entanto, estas não são analisadas tão pormenorizadamente, uma vez que os três exemplos acima ilustram esse mesmo procedimento. Assim sendo, foram elaborados gráficos em que para cada imagem, os três codecs são apresentados em simultâneo. Cada linha vai corresponder a uma imagem, pela ordem em que são apresentadas na Figura 3.1, exceto as que foram analisadas em pormenor acima. Cada coluna vai corresponder a uma métrica, sendo a primeira o PSNR, a segunda o MS-SSIM e a terceira o FSIM.



(a) Imagem de referência da base de imagens



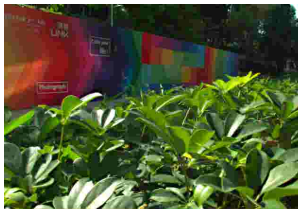
(b) Imagem A



(c) Imagem B



(d) Imagem Fundida



(e) JPEG taxa de bits 0.30



(f) JPEG 2000 taxa de bits 0.30



(g) JPEG XL taxa de bits 0.30



(h) JPEG taxa de bits 0.56



(i) JPEG 2000 taxa de bits 0.56



(j) JPEG XL taxa de bits 0.56



(k) JPEG taxa de bits 1.00



(l) JPEG 2000 taxa de bits 1.00



(m) JPEG XL taxa de bits 1.00



(n) JPEG taxa de bits 3.0



(o) JPEG 2000 taxa de bits 3.0



(p) JPEG XL taxa de bits 3.0

Figura 4.10: Fusão da Imagem 086 com posterior codificação/descodificação



(a) Imagem de referência da base de imagens



(b) Imagem A



(c) Imagem B



(d) Imagem Fundida



(e) JPEG taxa de bits 0.30



(f) JPEG 2000 taxa de bits 0.30



(g) JPEG XL taxa de bits 0.30



(h) JPEG taxa de bits 0.56



(i) JPEG 2000 taxa de bits 0.56



(j) JPEG XL taxa de bits 0.56



(k) JPEG taxa de bits 1.00



(l) JPEG 2000 taxa de bits 1.00



(m) JPEG XL taxa de bits 1.00



(n) JPEG taxa de bitse 3.0



(o) JPEG 2000 taxa de bits 3.0



(p) JPEG XL taxa de bits 3.0

Figura 4.11: Fusão da Imagem 155 com posterior codificação/descodificação



(a) *Imagem de referência da base de imagens*



(b) *Imagem A*



(c) *Imagem B*



(d) *Imagem Fundida*



(e) *JPEG taxa de bits 0.30*



(f) *JPEG 2000 taxa de bits 0.30*



(g) *JPEG XL taxa de bits 0.30*



(h) *JPEG taxa de bits 0.56*



(i) *JPEG 2000 taxa de bits 0.56*



(j) *JPEG XL taxa de bits 0.56*



(k) *JPEG taxa de bits 1.00*



(l) *JPEG 2000 taxa de bits 1.00*



(m) *JPEG XL taxa de bits 1.00*



(n) *JPEG taxa de bits 3.0*



(o) *JPEG 2000 taxa de bits 3.0*



(p) *JPEG XL taxa de bits 3.0*

Figura 4.12: Fusão da Imagem 621 com posterior codificação/descodificação

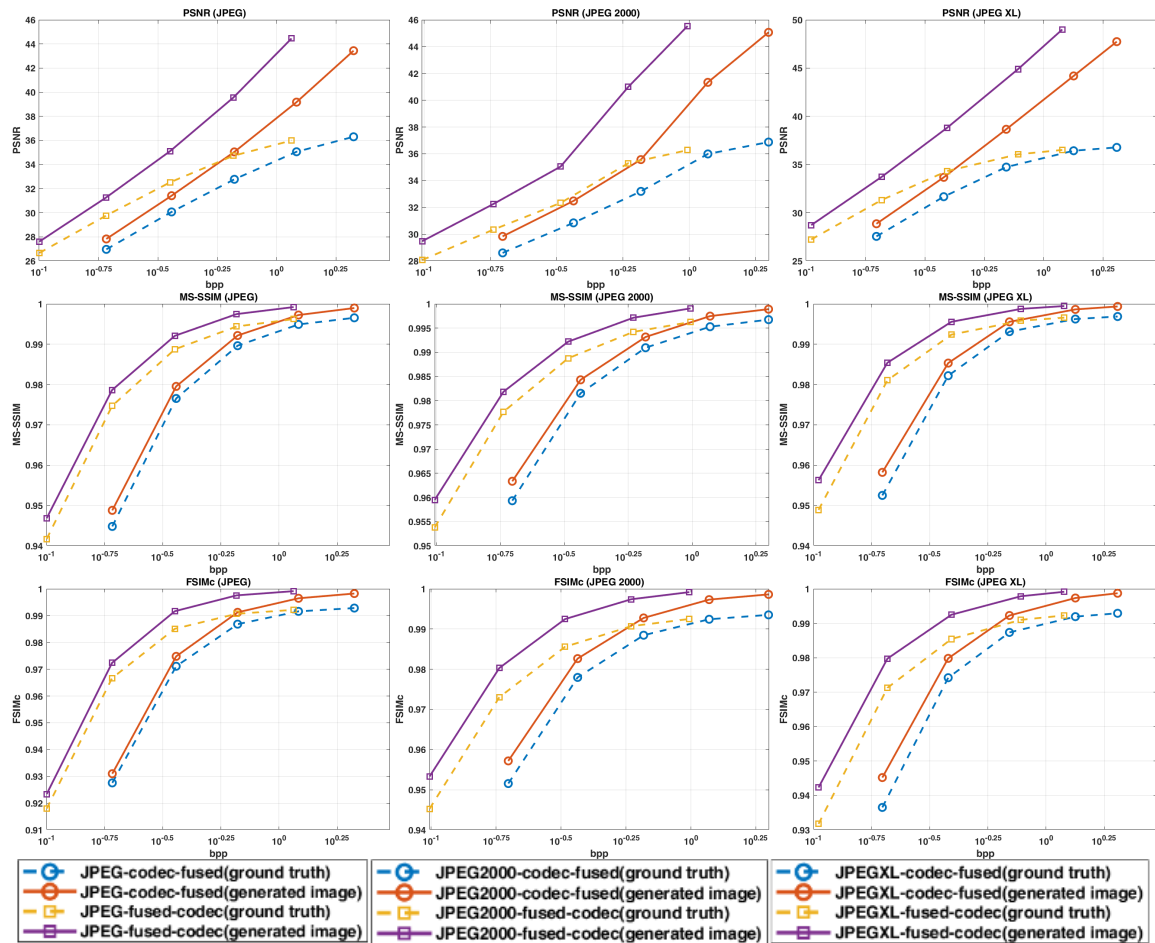


Tabela 4.1: Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para a Imagem o86

A curva azul representa o codec JPEG e compara as imagens codificadas, seguidas de fusão, com a imagem de referência (chamada de *generated image*). A curva laranja também corresponde ao codec JPEG, mas compara as imagens fundidas, seguidas de codificação/-descodificação, com a mesma imagem de referência. A curva amarela refere-se ao codec JPEG 2000 (Kakadu) e compara as imagens codificadas, seguidas de fusão, com a imagem de referência. A curva lilás também se refere ao JPEG 2000, mas compara as imagens fundidas, seguidas de codificação/-descodificação, com a referência. A curva verde representa o codec JPEG XL, que compara as imagens codificadas, seguidas de fusão, com a imagem de referência. Finalmente, a curva turquesa corresponde ao JPEG XL, e compara as imagens fundidas, seguidas de codificação, com a imagem de referência.

Através dos gráficos, é possível verificar que as curvas verde e turquesa, continuam a ter um melhor desempenho que todas as outras curvas. Inferindo-se que de forma consistente O JPEG XL que sofre primeiro codificação e depois fusão quando comparado com a imagem gerada, tem sempre um desempenho melhor e o JPEG que sofre primeiro codificação e depois fusão tem sempre o desempenho pior. O desempenho dos outros codecs e varia consoante o gráfico em análise. É possível inferir também, que o método com melhores resultados é aquele em que primeiro ocorre a fusão da imagem e só depois é que

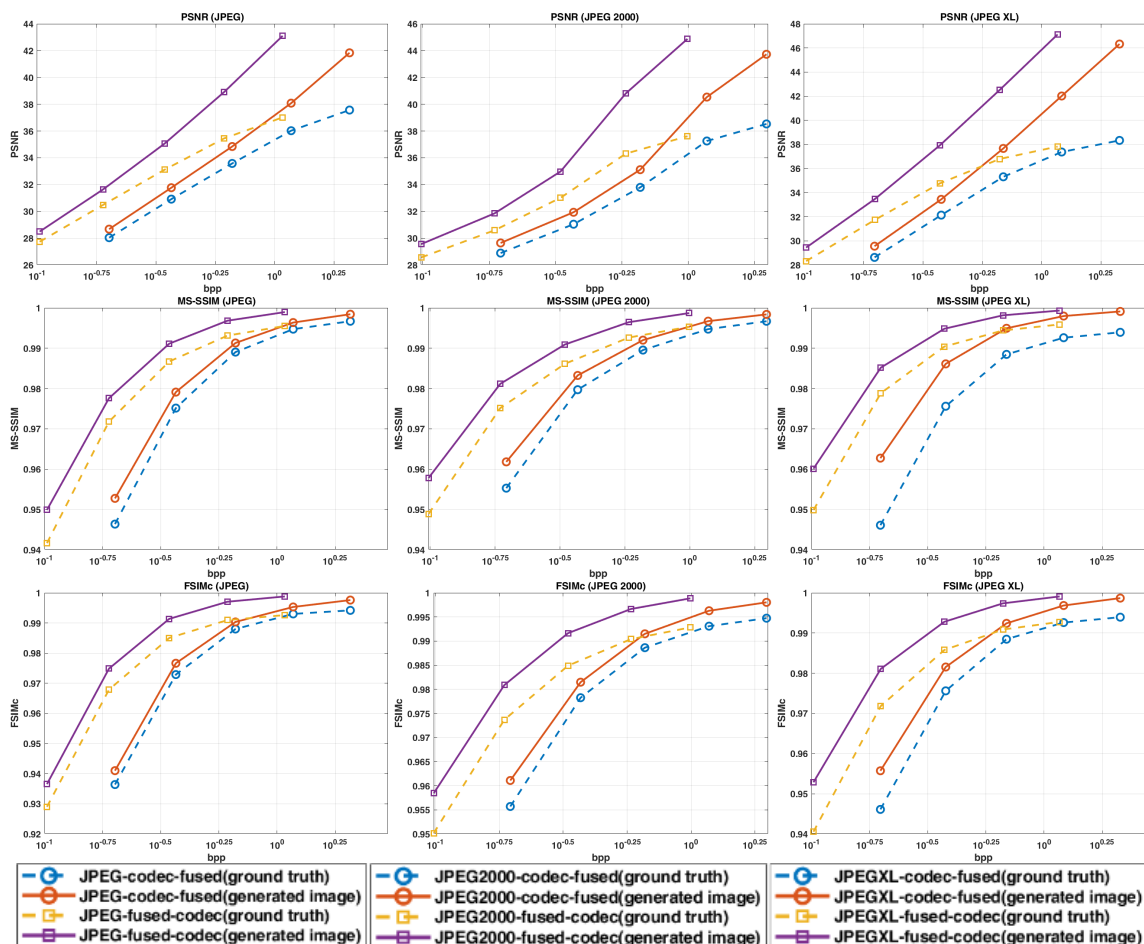


Tabela 4.2: Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para a Imagem 155

há codificação/descodificação, para o codec JPEG XL de um modo geral.

Para tornar mais fácil a análise destes gráficos, foi criada uma tabela para cada codificador, que calcula a área entre as duas curvas denominada de Delta de Bjøntegaard, do inglês *Bjøntegaard Delta* BD. É possível, concluir através da análise das tabelas abaixo que o facto de os valores serem negativos significa que a segunda curva que é analisada em relação à curva de referência (a primeira), se encontra abaixo desta, daí o valor ser negativo. Estas inconsistências para as métricas MS-SSIM e FSIMc, ocorrem porque o BD foi desenvolvido para avaliar a métrica do PSNR.

Nas três tabelas, 4.5, 4.6 e 4.7, há números a negrito que representam os valores desviantes e, por esta razão, os números a negrito (das imagens 122, 224, 276, 546 e 648), não são tidos em conta para o cálculo das médias e para a posterior elaboração da tabela dos valores médios com o desvio padrão 4.8.

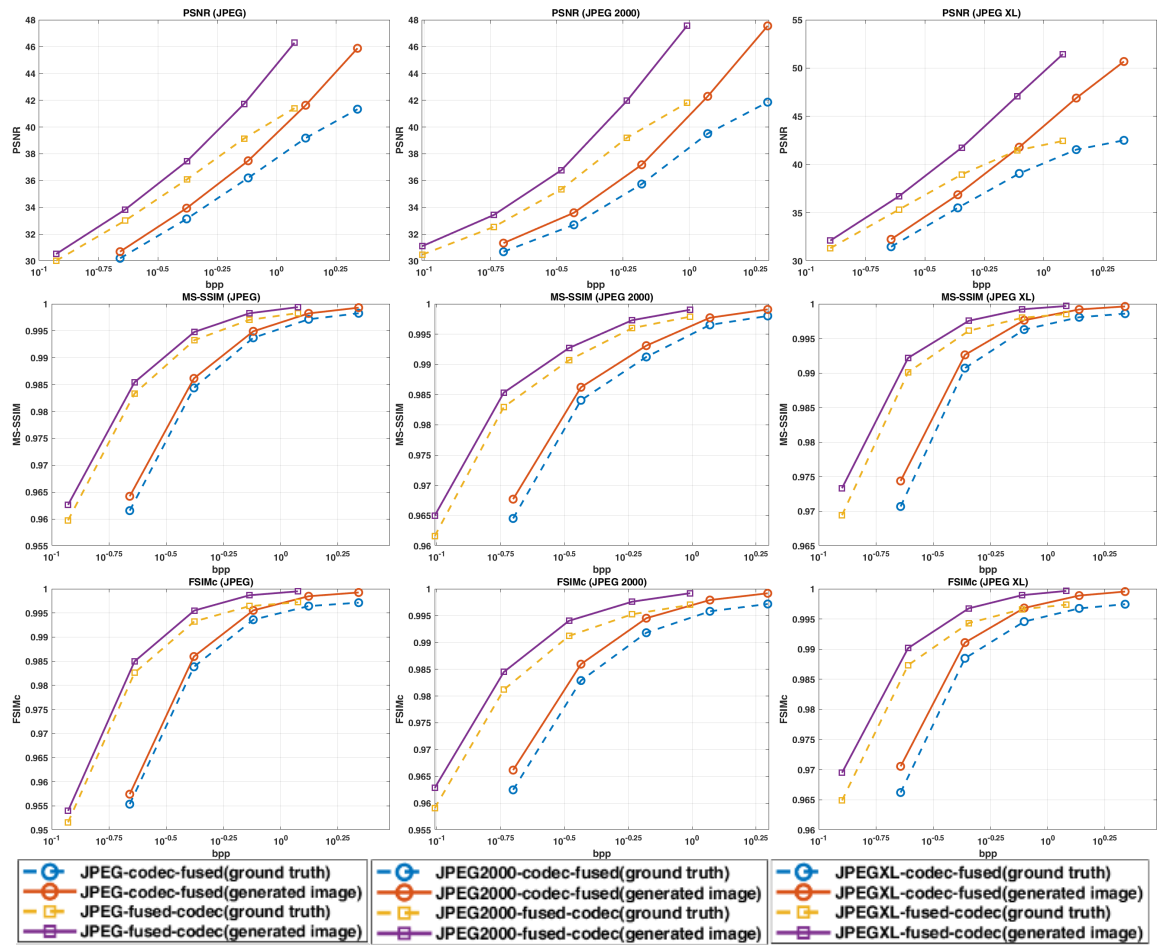
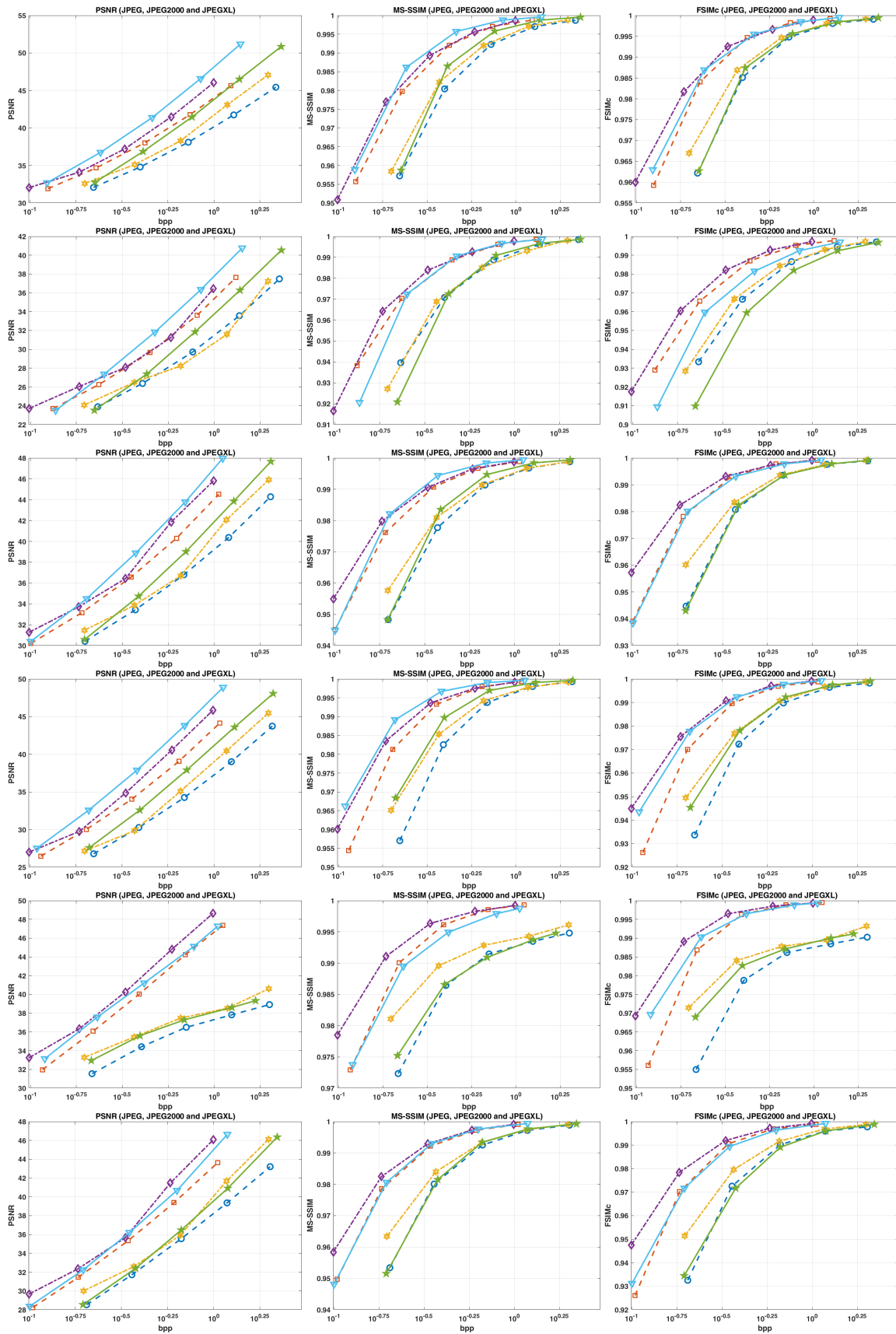
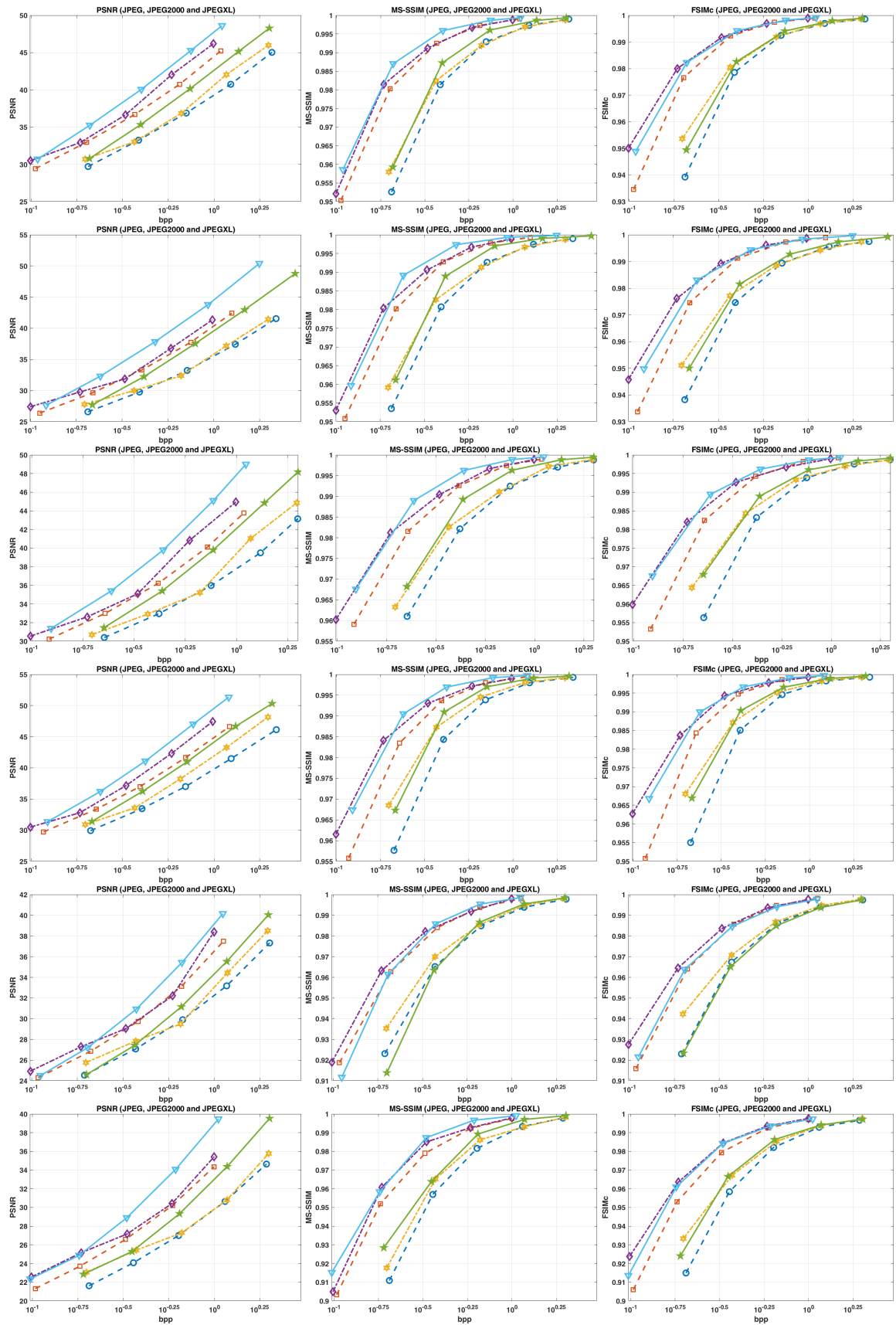


Tabela 4.3: Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para a Imagem o86

Imagem	JPEG					
	dsnr PSNR	rate PSNR (%)	dsnr MS-SSIM	rate MS-SSIM (%)	dsnr FSIMc	rate FSIMc (%)
001	-3,20	-0,01	-0,01	71,30	-0,01	68,66
028	-3,16	-0,01	-0,01	71,02	-0,01	68,51
086	-4,24	-0,01	-0,01	84,84	-0,01	76,21
091	-3,77	-3,77	-0,01	88,89	-0,01	93,65
101	-4,91	-0,01	-0,01	89,25	-0,02	82,24
122	-6,68	-0,01	-0,01	83,25	-0,02	-14,42
138	-4,20	-0,01	-0,01	87,36	-0,02	88,07
155	-4,16	-0,01	-0,01	90,46	-0,01	87,42
186	-4,11	-0,01	-0,01	89,02	-0,01	90,09
224	-3,78	-0,01	-0,01	71,53	-0,01	63,30
271	-3,74	-0,01	-0,01	78,90	-0,01	77,19
276	-4,13	-0,01	-0,01	78,27	-0,01	56,43
291	-2,93	-0,02	-0,02	71,56	-0,02	71,23
427	-3,41	-0,02	-0,02	84,78	-0,02	85,47
546	-4,17	-0,01	-0,01	90,62	-0,01	120,01
621	-3,95	-0,01	-0,01	79,28	-0,01	74,62
648	-8,72	-0,01	-0,01	191,21	-0,03	-64,64
694	-5,09	-0,01	-0,01	79,93	-0,01	53,71

Tabela 4.5: Valores obtidos através das BD para o codec JPEG





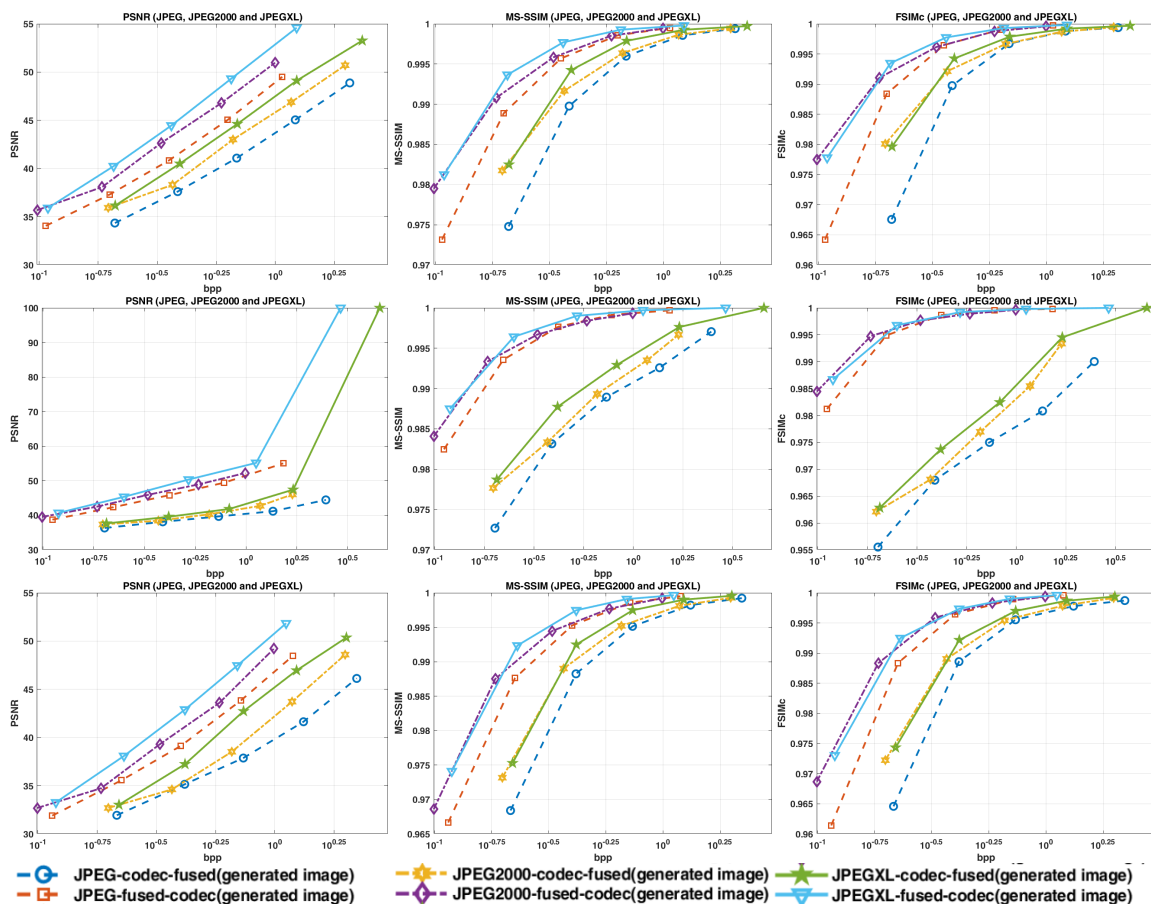


Tabela 4.4: Gráficos de comparação das métricas em função da ordem em que os codecs foram usados para as restantes imagens

JPEG 2000 (Kakadu)						
Imagem	dsnr PSNR	rate PSNR (%)	dsnr MS-SSIM	rate MS-SSIM (%)	dsnr FSIMc	rate FSIMc (%)
001	-3,29	-0,01	-0,01	78,49	-0,01	87,90
028	-3,04	-0,02	-0,02	77,59	-0,02	79,83
086	-4,61	-0,01	-0,01	90,66	-0,01	95,43
091	-4,40	-4,40	-0,01	93,32	-0,01	92,56
101	-5,88	-0,01	-0,01	84,26	-0,01	89,46
122	-6,56	-0,01	-0,01	84,33	-0,01	4,75
138	-4,90	-0,01	-0,01	83,86	-0,01	87,75
155	-4,89	-0,01	-0,01	89,35	-0,01	95,47
186	-4,94	-0,01	-0,01	83,92	-0,01	88,39
224	-3,79	-0,01	-0,01	80,12	-0,01	88,69
271	-4,33	-0,01	-0,01	85,59	-0,01	87,32
276	-4,52	-0,01	-0,01	82,44	-0,01	90,35
291	-2,65	-0,01	-0,01	73,72	-0,01	74,26
427	-3,16	-0,02	-0,02	78,35	-0,02	84,71
546	-4,61	-0,004	-0,004	85,62	-0,004	91,91
621	-4,74	-0,01	-0,01	89,53	-0,01	90,00
648	-8,32	-0,01	-0,01	351,17	-0,03	526,63
694	-5,44	-0,01	-0,01	81,80	-0,01	63,99

Tabela 4.6: Valores obtidos através das BD para o codec JPEG 2000 (Kakadu)

JPEG XL						
Imagem	dsnr PSNR	rate PSNR (%)	dsnr MS-SSIM	rate MS-SSIM (%)	dsnr FSIMc	rate FSIMc (%)
001	-3,94	-0,01	-0,01	90,36	-0,01	91,58
028	-3,75	-0,01	-0,01	75,39	-0,02	70,32
086	-5,02	-0,01	-0,01	60,67	-0,01	71,24
091	-4,53	-4,53	-0,01	84,54	-0,01	91,62
101	-5,84	-0,01	-0,01	87,98	-0,01	84,88
122	-6,05	-0,01	-0,01	91,39	-0,01	23,10
138	-4,54	-0,01	-0,01	82,80	-0,02	81,08
155	-4,77	-0,01	-0,01	84,80	-0,01	79,61
186	-4,80	-0,01	-0,01	96,88	-0,01	91,58
224	-4,68	-0,01	-0,01	-0,52	-0,01	39,88
271	-4,60	-0,01	-0,01	82,10	-0,01	78,56
276	-4,82	-0,01	-0,01	115,21	-0,01	85,83
291	-3,74	-0,02	-0,02	80,93	-0,02	78,82
427	-4,46	-0,02	-0,02	78,45	-0,02	82,97
546	-4,87	-0,004	-0,004	99,74	-0,004	108,92
621	-4,86	-0,01	-0,01	94,22	-0,01	92,56
648	-13,68	-0,01	-0,01	-36,25	-0,02	-77,95
694	-5,43	-0,01	-0,01	76,49	-0,01	47,97

Tabela 4.7: Valores obtidos através das BD para o codec JPEG XL

Após realizar o cálculo da variância para as tabelas 4.5, 4.6 e 4.7, as tabelas abaixo representam esses mesmos valores calculados.

Métrica	Bjontegaard	JPEG	JPEG 2000 (Kakadu)	JPEG XL
	Média±Desvio Padrão			
dsnr PSNR		-4,35 ± 1,34	-4,67 ± 1,31	-5,24 ± 2,13
rate PSNR (%)		-0,22 ± 0,91	-0,25 ± 1,00	-0,01 ± 1,07
dsnr MS-SSIM		-0,01 ± 3,39×10 ⁻³	-0,01 ± 3,66×10 ⁻³	-0,01 ± 4,07×10 ⁻³
rate MS-SSIM (%)		81,78 ± 6,98	83,70 ± 5,00	84,45 ± 9,48
dsnr FSIMc		-0,01 ± 4,90×10 ⁻³	-0,01 ± 4,96×10 ⁻³	-0,01 ± 4,30×10 ⁻³
rate FSIMc (%)		78,55 ± 12,13	86,75 ± 7,82	76,52 ± 19,77

Tabela 4.8: Tabela das BD da Média±Desvio Padrão

Por fim, esta tabela 4.8 permite-nos analisar as BD da Média±Desvio Padrão das três métricas usadas neste estudo para avaliar os três codecs escolhidos para fazer a codificação/descodificação das imagens, permitindo-nos perceber que para o dsnr PSNR o codec que apresenta melhor desempenho é o JPEG XL, no caso da rate PSNR (%) é o JPEG 2000 (kakadu) que apresenta o melhor desempenho. Para o dsnr MS-SSIM o codec que apresenta melhor desempenho é o JPEG XL, no caso da rate MS-SSIM (%) também é o JPEG XL que apresenta o melhor desempenho. Por fim, no caso do dsnr FSIMc o codec que apresenta melhor desempenho é o JPEG 2000 (Kakadu), no caso da rate MS-SSIM (%) também é o JPEG 2000 (Kakadu) que apresenta o melhor desempenho.

Capítulo 5

Conclusões e trabalho futuro

O objetivo deste trabalho é determinar se é possível comprimir as duas imagens originais em vez de comprimir a imagem fundida sem comprometer a qualidade da fusão. Os resultados obtidos mostram que tal é possível. No entanto, uma vez que os codecs analisados teriam de codificar ambas as imagens, seria necessário o dobro do espaço de armazenamento (para armazenar duas imagens) em comparação com o armazenamento da imagem resultante da fusão. Contudo, é de notar que o armazenamento de duas imagens com capacidade de fusão permite novas fusões que podem resultar em imagens de maior qualidade.

Relativamente aos codecs estudados, observou-se que a JPEG XL é o que tem maior capacidade de compressão. Isto foi comprovado através dos resultados obtidos através das métricas usadas e da análise das curvas obtidas nos gráficos.

Num trabalho futuro, pretende-se realizar um teste subjetivo para analisar a qualidade das imagens obtidas e, posteriormente, escrever um artigo que reporte esses resultados. Uma vez que as duas imagens são muito semelhantes, a taxa de bits pode ser significativamente reduzida através da investigação de um modelo de codificação que permita a codificação conjunta das duas imagens e elimine as redundâncias comuns.

Além disso, também é um objetivo fazer uma comparação entre os algoritmos de fusão utilizados para verificar qual deles é melhor e o motivo dessa superioridade.

Bibliografia

- [1] S. Bhat and D. Koundal, “Multi-focus image fusion techniques: a survey,” *Artificial Intelligence Review*, vol. 54, pp. 1–53, 02 2021. xiii, 1, 4, 5
- [2] R. Neelamani, R. de Queiroz, Z. Fan, S. Dash, and R. Baraniuk, “JPEG compression history estimation for color images,” *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1365–1378, 2006. xiii, 16
- [3] D. S. Taubman, “Kakadu survey documentation (last updated for version 5.0),” 2005. xiii, 17, 18
- [4] J. P. E. G. (JPEG), “Jpeg xl: A new standard for image compression,” tech. rep., 2021. Whitepaper. xiii, 18, 19, 20
- [5] J. Zhang, Q. Liao, S. Liu, H. Ma, W. Yang, and J.-H. Xue, “Real-mff: A large realistic multi-focus image dataset with ground truth,” *Pattern Recognition Letters*, vol. 138, pp. 370–377, 2020. xiii, 25, 26
- [6] Y. Liu, S. Liu, and Z. Wang, “Multi-focus image fusion with dense SIFT,” *Information Fusion*, vol. 23, pp. 139–155, 2015. 3, 6
- [7] Z. Wang and Y. Ma, “Medical image fusion using m-pcnn,” *Information Fusion*, vol. 9, no. 2, pp. 176–185, 2008. 4, 5
- [8] J. Zhi-guo, H. Dong-bing, C. Jin, and Z. Xiao-kuan, “A wavelet based algorithm for multi-focus micro-image fusion,” in *Third International Conference on Image and Graphics (ICIG’04)*, pp. 176–179, 2004. 4
- [9] K. Sujatha and D. Shalini Punithavathani, “Optimized ensemble decision-based multi-focus imagefusion using binary genetic grey-wolf optimizer in camera sensor networks,” *Multimedia Tools and Applications*, vol. 77, pp. 1735–1759, Jan 2018. 4
- [10] Y. Yang, W. Zheng, and S. Huang, “Effective multifocus image fusion based on hvs and bp neural network,” *Journal of Signal Processing*, February 6 2014. Citations: 7. 4, 10
- [11] H. Li, H. Qiu, Z. Yu, and B. Li, “Multifocus image fusion via fixed window technique of multiscale images and non-local means filtering,” *Signal Processing*, vol. 138, pp. 71–85, 2017. 4, 13
- [12] Z. Chen, D. Wang, S. Gong, and F. Zhao, “Application of multi-focus image fusion in visual power patrol inspection,” in *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 1688–1692, 2017. 4

- [13] Y. Song, M. Li, Q. Li, and L. Sun, "A new wavelet based multi-focus image fusion scheme and its application on optical microscopy," in *2006 IEEE International Conference on Robotics and Biomimetics*, pp. 401–405, 2006. 4
- [14] R. Van de Plas, J. Yang, J. Spraggins, and R. M. Caprioli, "Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping," *Nature Methods*, vol. 12, pp. 366–372, Apr 2015. 4
- [15] R. P. R. Paper, H. Kaur, E. Jyoti, and M. T. Student, "Analytical comparison of various image fusion techniques," 2015. 5
- [16] R. Garg, P. Gupta, and H. Kaur, "Survey on multi-focus image fusion algorithms," in *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*, pp. 1–5, 2014. 5, 6
- [17] M. Nejati, S. Samavi, N. Karimi, S. Reza Soroushmehr, S. Shirani, I. Roosta, and K. Najarian, "Surface area-based focus criterion for multi-focus image fusion," *Information Fusion*, vol. 36, pp. 284–295, 2017. 6, 8
- [18] K. Kanagaraj, P. Arumuga, and K. Arulmozhi, "Optimal decomposition level of discrete, stationary and dual tree complex wavelet transform for pixel based fusion of multi-focused images," *SERBIAN JOURNAL OF ELECTRICAL ENGINEERING*, vol. 7, pp. 81–93, 04 2010. 6
- [19] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: A survey of the state of the art," *Information Fusion*, vol. 64, pp. 71–91, 2020. 6
- [20] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Information Fusion*, vol. 25, pp. 72–84, 2015. 6
- [21] C. Ludusan and O. Laviolle, "Multifocus image fusion and denoising: A variational approach," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1388–1396, 2012. 6
- [22] X. Xia, Y. Yao, L. Yin, S. Wu, H. Li, and Z. Yang, "Multi-focus image fusion based on probability filtering and region correction," *Signal Processing*, vol. 153, pp. 71–82, 2018. 6, 15
- [23] M. S. Farid, A. Mahmood, and S. A. Al-Maadeed, "Multi-focus image fusion using content adaptive blurring," *Information Fusion*, vol. 45, pp. 96–112, 2019. 7
- [24] J. Ma, Z. Zhou, B. Wang, L. Miao, and H. Zong, "Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps," *Neurocomputing*, vol. 335, pp. 9–20, 2019. 7
- [25] O. Bouzos, I. Andreadis, and N. Mitianoudis, "Conditional random field model for robust multi-focus image fusion," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5636–5648, 2019. 7

- [26] S. Li, J. T. Kwok, and Y. Wang, "Combination of images with diverse focuses using the spatial frequency," *Information Fusion*, vol. 2, no. 3, pp. 169–176, 2001. 7
- [27] D. Agrawal and S. Jyoti, "Multifocus image fusion using modified pulse coupled neural network for improved image quality," *Image Processing, IET*, vol. 4, pp. 443 – 451, 01 2011. 7
- [28] V. Aslantas and R. Kurban, "Fusion of multi-focus images using differential evolution algorithm," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8861–8870, 2010. 7
- [29] A. Banharnsakun, "Multi-focus image fusion using best-so-far abc strategies," *Neural Computing and Applications*, vol. 31, pp. 2025–2040, Jul 2019. 7
- [30] M. Li, W. Cai, and Z. Tan, "A region-based multi-sensor image fusion scheme using pulse-coupled neural network," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1948–1956, 2006. 7
- [31] Y. Liu, J. Jin, Q. Wang, Y. Shen, and X. Dong, "Region level based multi-focus image fusion using quaternion wavelet and normalized cut," *Signal Processing*, vol. 97, pp. 9–30, 2014. 8
- [32] X. Hao, H. Zhao, and J. Liu, "Multifocus color image sequence fusion based on mean shift segmentation," *Appl. Opt.*, vol. 54, pp. 8982–8989, Oct 2015. 8
- [33] Y. Huang, W. Li, M. Gao, and Z. Liu, "Algebraic multi-grid based multi-focus image fusion using watershed algorithm," *IEEE Access*, vol. 6, pp. 47082–47091, 2018. 8
- [34] L. Ming, W. Yan, and W. Shunjun, "Multi-focus image fusion based on wavelet decomposition and evolutionary strategy," in *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003*, vol. 2, pp. 951–955 Vol.2, 2003. 9
- [35] Y. Lu, X. Feng, J. Zhang, R. Wang, K. Zheng, and J. Kong, "A multi-focus image fusion based on wavelet and region detection," in *EUROCON 2007 - The International Conference on "Computer as a Tool"*, pp. 294–298, 2007. 9
- [36] J. Saeedi, K. Faez, and S. Mozaffari, "Multi-focus image fusion based on fuzzy and wavelet transform," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (E. Bayro-Corrochano and J.-O. Eklundh, eds.), (Berlin, Heidelberg), pp. 970–977, Springer Berlin Heidelberg, 2009. 9
- [37] Y. Yang, "A novel dwt based multi-focus image fusion method," *Procedia Engineering*, vol. 24, pp. 177–181, 2011. International Conference on Advances in Engineering 2011. 9

- [38] J. Tian and L. Chen, “Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure,” *Signal Processing*, vol. 92, no. 9, pp. 2137–2146, 2012. 9
- [39] Y. Chai, H. Li, and X. Zhang, “Multifocus image fusion based on features contrast of multiscale products in nonsubsampled contourlet transform domain,” *Optik*, vol. 123, no. 7, pp. 569–581, 2012. 9
- [40] Y. Liu, J. Jin, Q. Wang, Y. Shen, and X. Dong, “Novel focus region detection method for multifocus image fusion using quaternion wavelet,” *Journal of Image Processing*, May 29 2013. Author Affiliations: Various institutions. 9
- [41] M. Patil, “Multi focus image fusion based on spatial frequency and contrast based analysis under stationary wavelet transform domain,” *International Journal of Scientific Engineering Research, ISSN 2229-5518*, vol. 7, p. 225..230, 05 2016. 10
- [42] T. Singh and P. Manchanda, “Multi-focus image fusion using modified gaussian filter and discrete lifting wavelet transform,” *International Journal of Electronics Engineering*, vol. 10, no. 2, pp. 201–209, 2018. June 2018-Dec 2018, UGC Recommended Journal, Guru Nanak Dev University, Amritsar, Punjab, India, 143001. 10
- [43] S. Aymaz and C. Köse, “A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion,” *Information Fusion*, vol. 45, pp. 113–127, 2019. 10
- [44] P. J. Burt and E. H. Adelson, “Merging images through pattern decomposition,” *Journal of Image Processing*, December 19 1985. Author Affiliations: Various institutions. 10
- [45] P. Burt and R. Kolczynski, “Enhanced image capture through fusion,” in *1993 (4th) International Conference on Computer Vision*, pp. 173–182, 1993. 10
- [46] L. Bogoni and M. Hansen, “Pattern-selective color image fusion,” *Pattern Recognition*, vol. 34, no. 8, pp. 1515–1526, 2001. 10
- [47] L. Kou, L. Zhang, K. Zhang, J. Sun, Q. Han, and Z. Jin, “A multi-focus image fusion method via region mosaicking on laplacian pyramids,” *PLOS ONE*, vol. 13, p. e0191085, May 17 2018. 10
- [48] J. Sun, Q. Han, L. Kou, L. Zhang, K. Zhang, and Z. Jin, “Multi-focus image fusion algorithm based on laplacian pyramids,” *J. Opt. Soc. Am. A*, vol. 35, pp. 480–490, Mar 2018. 10
- [49] X. Jin, J. Hou, R. Nie, S. Yao, D. Zhou, Q. Jiang, and K. He, “A lightweight scheme for multi-focus image fusion,” *Multimedia Tools and Applications*, vol. 77, pp. 23501–23527, Sep 2018. 10

- [50] B. Zhang, C. Zhang, L. Yuanyuan, W. Jianshuai, and L. He, "Multi-focus image fusion algorithm based on compound pcnn in surfacelet domain," *Optik*, vol. 125, no. 1, pp. 296–300, 2014. 11
- [51] S. Liu, M. Shi, Z. Zhu, and J. Zhao, "Image fusion based on complex-shearlet domain with guided filtering," *Multidimensional Systems and Signal Processing*, vol. 28, pp. 207–224, Jan 2017. 11
- [52] R. Hong, C. Wang, Y. Ge, M. Wang, X. Wu, and R. Zhang, "Saliency preserving multi-focus image fusion," in *2007 IEEE International Conference on Multimedia and Expo*, pp. 1663–1666, 2007. 11
- [53] Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus images," *Information Fusion*, vol. 20, pp. 60–72, 2014. 11
- [54] S. Paul, I. S. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain," *Journal of Circuits, Systems and Computers*, vol. 25, no. 10, p. 1650123, 2016. 11
- [55] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Processing*, vol. 9, pp. 347–357, May 1 2015. 12
- [56] H. Yin, Y. Li, Y. Chai, Z. Liu, and Z. Zhu, "A novel sparse-representation-based multi-focus image fusion approach," *Neurocomputing*, vol. 216, pp. 216–229, 2016. 12
- [57] X. Ma, S. Hu, S. Liu, J. Fang, and S. Xu, "Multi-focus image fusion based on joint sparse representation and optimum theory," *Signal Processing: Image Communication*, vol. 78, pp. 125–134, 2019. 12
- [58] M. B. A. Haghghat, A. Aghagolzadeh, and H. Seyedarabi, "Real-time fusion of multi-focus images for visual sensor networks," in *2010 6th Iranian Conference on Machine Vision and Image Processing*, pp. 1–6, 2010. 12
- [59] Y. A. V. Phamila and R. Amutha, "Discrete cosine transform based fusion of multi-focus images for visual sensor networks," *Signal Processing*, vol. 95, pp. 161–170, 2014. 12
- [60] H. Tang, B. Xiao, W. Li, and G. Wang, "Pixel convolutional neural network for multi-focus image fusion," *Information Sciences*, vol. 433-434, pp. 125–141, 2018. 13
- [61] X. Guo, R. Nie, J. Cao, D. Zhou, and W. Qian, "Fully convolutional network-based multifocus image fusion," *Neural Computation*, vol. 30, pp. 1–26, 06 2018. 13
- [62] C. Du and S. Gao, "Multi-focus image fusion algorithm based on pulse coupled neural networks and modified decision map," *Optik*, vol. 157, pp. 1003–1015, 2018. 13
- [63] Y. Yang, Z. Nie, S. Huang, P. Lin, and J. Wu, "Multilevel features convolutional neural network for multifocus image fusion," *IEEE Transactions on Computational Imaging*, vol. 5, no. 2, pp. 262–273, 2019. 13

- [64] H. Li, R. Nie, J. Cao, X. Guo, D. Zhou, and K. He, “Multi-focus image fusion using u-shaped networks with a hybrid objective,” *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9755–9765, 2019. 13
- [65] R. Lai, Y. Li, J. Guan, and A. Xiong, “Multi-scale visual attention deep convolutional neural network for multi-focus image fusion,” *IEEE Access*, vol. 7, pp. 114385–114399, 2019. 14
- [66] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, “Ifcnn: A general image fusion framework based on convolutional neural network,” *Information Fusion*, vol. 54, pp. 99–118, 2020. 14
- [67] J. Li, X. Guo, G. Lu, B. Zhang, Y. Xu, F. Wu, and D. Zhang, “Drpl: Deep regression pair learning for multi-focus image fusion,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4816–4831, 2020. 14
- [68] H. Li, X. Liu, Z. Yu, and Y. Zhang, “Performance improvement scheme of multifocus image fusion derived by difference images,” *Signal Processing*, vol. 128, pp. 474–493, 2016. 14
- [69] H. Li, X.-J. Wu, and T. Durrani, “Multi-focus noisy image fusion using low-rank representation,” *arXiv preprint arXiv:1804.09429*, April 25 2018. Last revised: January 29, 2022. 14
- [70] Y. Yang, M. Yang, S. Huang, M. Ding, and J. Sun, “Robust sparse representation combined with adaptive pcnn for multifocus image fusion,” *IEEE Access*, vol. 6, pp. 20138–20151, 2018. 14
- [71] K. He, D. Zhou, X. Zhang, R. Nie, and X. Jin, “Multi-focus image fusion combining focus-region-level partition and pulse-coupled neural network,” *Soft Computing*, vol. 23, pp. 4685–4699, Jul 2019. 15
- [72] Y. Liu and Z. Wang, “Simultaneous image fusion and denoising with adaptive sparse representation,” *IET Image Process.*, vol. 9, pp. 347–357, 2015. 15
- [73] J. Zhang, Q. Liao, S. Liu, H. Ma, W. Yang, and J.-H. Xue, “Real-mff: A large realistic multi-focus image dataset with ground truth,” *Pattern Recognition Letters*, vol. 138, pp. 370–377, 2020. 15
- [74] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. USA: Kluwer Academic Publishers, 1st ed., 1992. 17
- [75] K. Software, “Kakadu software documentation and downloads.” <https://kakadusoftware.com/documentation-downloads/>, 2024. Accessed: 2024-10-01. 17
- [76] K. Software, “Kakadu software: High performance jpeg2000 software toolkit.” <https://kakadusoftware.com/>, 2024. Accessed: 2024-10-01. 17

- [77] J. Ascenso, E. Alshina, and T. Ebrahimi, “The jpeg ai standard: Providing efficient human and machine visual data consumption,” *IEEE MultiMedia*, vol. 30, no. 1, pp. 100–111, 2023. 21
- [78] M. Prieto and A. Allen, “A similarity metric for edge images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1265–1273, 2003. 21
- [79] A. Horé and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in *2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, 2010. 21
- [80] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, “On between-coefficient contrast masking of dct basis functions,” in *Proceedings of the International Conference on Image Processing*, 2023. 21
- [81] “Psnr-hvs-m download page.” 21
- [82] Z. Wang, E. Simoncelli, and A. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, pp. 1398–1402 Vol.2, 2003. 22
- [83] D. M. Rouse and S. S. Hemami, “Understanding and simplifying the structural similarity metric,” in *2008 15th IEEE International Conference on Image Processing*, pp. 1188–1191, 2008. 22
- [84] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011. 22
- [85] ISO/IEC JTC 1/SC 29/WG 1, “Jpeg ai common training and test conditions,” Tech. Rep. ISO/IEC JTC 1/SC29/WG1 N100421, 98th JPEG Meeting, Sydney, 16-20 January 2023, Lausanne, Switzerland, 2023. Approved for distribution. Project: JPEG AI (ISO/IEC 6048). 27, 31
- [86] K. S. P. Ltd., “Welcome to kakadu software: The world’s leading jpeg2000 developer toolkit.” <https://kakadusoftware.com/> and <https://kakadusoftware.com/documentation-downloads/downloads/>, 2024. Accessed: 2024-10-01. 31
- [87] T. libjxl Contributors, “libjxl: Jpeg xl image codec.” <https://github.com/libjxl/libjxl>, 2024. Accessed: 2024-10-01. 31