



## **Fusão de Imagem para Detecção de Pessoas Camufladas**

**Nádia Alexandra Ferreira Bento**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Eletrotécnica e de Computadores**

Orientador: Doutor José Manuel Bioucas-Dias  
Coorientador: Doutor José Silvestre Serra da Silva

### **Júri**

Presidente: (nome completo e categoria)

Orientador: (nome completo e categoria)

Vogais: (nome completo e categoria)

**Outubro de 2015**

Aos meus pais, irmã e amigos pela  
compreensão, paciência e tolerância.

# Agradecimentos

A presente dissertação é fruto de uma cooperação, e nesse sentido não poderia deixar de expressar os meus mais sinceros agradecimentos a todos que de alguma forma contribuíram para a sua realização e também para o seu enriquecimento, em particular às seguintes pessoas e instituições:

À minha família, em particular aos meus pais e irmã, pela paciência, apoio e palavras de incentivo ao longo destes seis anos na Academia Militar. Sem a sua dedicação e amor não teria sido possível ultrapassar os mais diversos obstáculos que foram surgindo no meu percurso;

Ao meu Orientador, Doutor José Bioucas-Dias, e ao meu Coorientador, Doutor José Silvestre Serra da Silva, pela sua total disponibilidade, dedicação, incentivo e preocupação, nunca recusando a prestação de ajuda. Sem as suas orientações e conselhos não teria sido possível realizar este trabalho com o rigor que o próprio exige. Quero ainda agradecer todas as correções efetuadas durante a escrita da dissertação;

Aos Oficiais que exerceram a função de Diretor de Curso da Arma de Transmissões, ao longo no meu percurso na Academia Militar, pelo constante acompanhamento, preocupação e dedicação prestados em auxílio de todos os alunos de Transmissões;

À Academia Militar pelo contributo que teve na minha formação e no meu desenvolvimento enquanto pessoa e militar;

A todos os Oficiais que permitiram a realização do meu trabalho de campo, tanto no exercício em Beja como em Santa Margarida, demonstrando uma total disponibilidade e cooperação;

A todos os alunos da Academia Militar que prontamente se disponibilizaram para a recolha de imagens nos exercícios de campo;

Por último, mas não menos importante, aos meus camaradas da Academia Militar, especialmente à Ana Mesquita, ao Joel Dias, ao Jorge Pimenta, ao André Silva, ao Jonathan Guimarães, ao David Campos, à Miriana Gonçalves, à Marta Frade, à Catarina Correia, à Ana Fernandes e ao Tiago Silva pela amizade, camaradagem, espírito de sacrifício, motivação, ânimo e apoio em todos os momentos ao longo destes anos. O meu mais sincero Obrigado.

## Resumo

A utilização de imagens térmicas constitui um benefício para as Forças Armadas. Devido às suas inúmeras vantagens, têm um grande número de aplicações, entre as quais a detecção de pessoas camufladas. Para melhores resultados, a informação visível pode ser combinada com a informação térmica. A fusão de informação gerada por múltiplas fontes de imagem, permite um maior detalhe, resultando num maior grau de segurança.

No presente estudo foram implementados três métodos de fusão de imagem ao nível do pixel baseados em: análise em componentes principais, pirâmide laplaciana e transformada wavelet discreta. As imagens utilizadas têm dimensão de  $320 \times 240$  pixéis, cada uma é constituída por uma imagem da banda do visível e uma imagem térmica, e foram obtidas com uma câmara FLIR. Foi realizada uma análise qualitativa de onde se concluiu que o método que obtém melhores resultados é aquele que utiliza as wavelets, seguido do que utiliza a pirâmide laplaciana. De seguida, foi efetuada uma análise quantitativa usando como métricas de desempenho o desvio padrão, a entropia, a frequência espacial, a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural. Os valores adquiridos sustentam as conclusões retiradas da análise qualitativa.

Conclui-se que a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural são as métricas adequadas para medir a qualidade da fusão de imagem, pois têm em consideração a relação existente entre a imagem fundida e as imagens fonte.

**Palavras-chave:** Fusão de imagem ao nível do pixel, wavelets, pirâmide laplaciana, análise das componentes principais, métrica de desempenho, segurança.

## Abstract

The use of thermal imaging is a benefit for the Armed Forces. Due to their several advantages, they have a large number of applications, including the detection of camouflaged people. For better results, the visible information can be merged with the thermal information. The information generated by fusing multiple input images, allows a greater detail, resulting in a greater degree of security.

The present study implemented three-pixel level image fusion methods based on: principal components analysis, laplacian pyramid and discrete wavelet transform. The images used have  $320 \times 240$  pixels, each one consisting of a visible image and a thermal image, and were obtained with a FLIR camera. A qualitative analysis concluded that the method which performs better is the one that uses wavelets, followed by the laplacian pyramid. Then it was made a quantitative analysis using performance metrics as the standard deviation, entropy, spatial frequency, mutual information, fusion quality index and structural similarity index. The values obtained support the conclusions drawn from the qualitative analysis.

We conclude that the mutual information, fusion quality index and structural similarity index are the appropriate metrics to measure the quality of image fusion as they take into account the relationship between the fused image and the input images.

**Keywords:** Pixel level image fusion, wavelets, laplacian pyramid, principal component analysis, performance metrics, security.

# Índice

Agradecimentos .....	ii
Resumo .....	iii
Abstract .....	iv
Lista de Figuras .....	vii
Lista de Símbolos .....	x
Lista de Acrónimos e Siglas .....	xii
1 Introdução .....	1
1.1 Enquadramento e Motivação .....	1
1.2 Objetivos .....	2
1.3 Estrutura da Dissertação .....	3
2 Estado Da Arte .....	5
2.1 Resumo de alguns trabalhos em fusão de imagem .....	5
2.2 Fusão de Imagem .....	9
2.2.1 Fusão de Imagem ao Nível do Pixel .....	9
2.2.2 Fusão de Imagem ao Nível das Características e ao Nível da Decisão .....	12
3 Métodos .....	16
3.1 Enquadramento .....	16
3.2 Análise em Componente Principais .....	18
3.2.1 Cálculo das Componentes Principais .....	18
3.2.2 Fusão de Imagem .....	19
3.3 Pirâmide Laplaciana .....	19
3.3.1 Decomposição de Imagem .....	21
3.3.2 Reconstrução da Imagem .....	22
3.3.3 Fusão de Imagem .....	23
3.4 Wavelets .....	24
3.4.1 Famílias de Wavelets .....	24
3.4.2 Decomposição de Imagem .....	25
3.4.3 Reconstrução da Imagem .....	26
3.4.4 Fusão de Imagem .....	27
3.5 Métricas de Avaliação de Desempenho .....	27
3.5.1 Desvio Padrão .....	28
3.5.2 Entropia .....	28
3.5.3 Frequência Espacial .....	29
3.5.4 Informação Mútua .....	29
3.5.5 Índice de Qualidade de Fusão .....	30
3.5.6 Índice de Semelhança Estrutural .....	31
4 Resultados e Discussão .....	32
4.1 Conjunto de Dados .....	32

4.2	Análise em Componentes Principais .....	32
4.3	Pirâmide Laplaciana .....	35
4.4	Wavelets .....	36
4.5	Análise Qualitativa .....	39
4.6	Análise Quantitativa .....	43
4.6.1	Desvio Padrão .....	44
4.6.2	Entropia .....	46
4.6.3	Frequência Espacial .....	47
4.6.4	Informação Mútua .....	48
4.6.5	Índice de Qualidade de Fusão .....	48
4.6.6	Índice de Semelhança Estrutural .....	49
4.7	Comparação da Avaliação Qualitativa e Quantitativa .....	50
5	Conclusões .....	52
5.1	Trabalho Futuro .....	54
	Referências .....	56
	Anexo 1 .....	59
	Imagens utilizadas .....	59

## Lista de Figuras

Figura 1 – a) imagem térmica do porta-aviões b) imagem visível do porta-aviões (retirada de [7]) .....	6
Figura 2 – a) imagem fundida com a PCA b) imagem fundida com a DWT-PCA c) imagem fundida com PCA-máximo (retirada de [7]) .....	6
Figura 3 - a) imagem térmica da estrada b) imagem visível da estrada (retirada de [7]) .....	6
Figura 4 - a) imagem fundida com a PCA b) imagem fundida com a DWT-PCA c) imagem fundida com PCA-máximo (retirada de [7]) .....	7
Figura 5 - a) imagem térmica do campo b) imagem visível do campo (retirada de [7]) .....	7
Figura 6 - a) imagem fundida com a PCA b) imagem fundida com a DWT-PCA c) imagem fundida com PCA-máximo (retirada de [7]) .....	8
Figura 7 – a) imagem visível do campo b) imagem térmica do campo c) imagem fundida com pirâmide laplaciana d) imagem fundida com DWT (retirada de [8]) .....	8
Figura 8 – a) imagem visível do campo b) imagem térmica do campo c) imagem fundida com pirâmide laplaciana d) imagem fundida com DWT (retirada de [8]) .....	8
Figura 9 – Diagrama de radiação de Plank (adaptada de [22]) .....	16
Figura 10 – Processo de fusão pixel a pixel (adaptada de [6]) .....	17
Figura 11 – Componentes principais .....	18
Figura 12 – Fluxo de informação do algoritmo baseado em PCA (adaptada de [9]) .....	19
Figura 13 – Passos básicos da construção da pirâmide .....	20
Figura 14 – Pirâmide laplaciana com 5 níveis .....	21
Figura 15 – Decomposição da imagem utilizando a pirâmide laplaciana (adaptada de [14]) .....	22
Figura 16 – Reconstrução da imagem utilizando a pirâmide laplaciana (adaptada de [14]) .....	23
Figura 17 – Esquema de fusão de imagem com pirâmide laplaciana (adaptada de [24]) .....	24
Figura 18 – Decomposição da imagem usando wavelets (adaptada de [9]) .....	25
Figura 19 – a) coeficiente de detalhe b) coeficiente de detalhe horizontal c) coeficiente de detalhe vertical d) coeficiente de detalhe diagonal .....	26
Figura 20 – Reconstrução da imagem usando wavelets (adaptada de [9]) .....	26
Figura 21 – Esquema de fusão usando wavelets (adaptada de [9]) .....	27
Figura 22 – a) imagem visível b) imagem térmica c) imagem fundida com 1ª variante PCA .....	33
Figura 23 – a) imagem visível equalizada b) imagem térmica equalizada c) imagem fundida com 2ª variante PCA .....	33
Figura 24 – a) imagem térmica inicial b) imagem térmica equalizada .....	34
Figura 25 - a) imagem visível b) imagem térmica c) imagem fundida com 3ª variante PCA .....	34
Figura 26 – a) imagem fundida com 1ª variante PCA b) imagem fundida com 2ª variante PCA c) imagem fundida com 3ª variante PCA .....	35
Figura 27 a) imagem fundida com Pirâmide Laplaciana de 2 níveis b) imagem fundida com pirâmide laplaciana de 3 níveis c) imagem fundida com pirâmide laplaciana de 4 níveis .....	36
Figura 28 – a) coeficientes com 1 nível de decomposição b) coeficientes com 2 níveis de decomposição .....	37

Figura 29 – a) imagem fundida com ‘máximo’ b) imagem fundida com ‘média’ c) imagem fundida com ‘mínimo’ .....	37
Figura 30 – a) imagem fundida com 1ª regra de fusão com 1 nível de decomposição b) imagem fundida com 1ª regra de fusão com 2 níveis de decomposição .....	38
Figura 31 – a) imagem fundida com 1ª regra de fusão com 1 nível de decomposição b) imagem fundida com 1ª regra de fusão com 2 níveis de decomposição .....	38
Figura 32 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	39
Figura 33 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	40
Figura 34 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	40
Figura 35 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	40
Figura 36 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	41
Figura 37 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	41
Figura 38 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	42
Figura 39 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	42
Figura 40 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	42
Figura 41 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão .....	43
Figura 42 - Boxplot .....	44
Figura 43 – Boxplot do Desvio Padrão .....	45
Figura 44 – Boxplot do Desvio Padrão .....	45
Figura 45 – Boxplot da Entropia .....	46

Figura 46 – Boxplot da Frequência Espacial.....	47
Figura 47 – Boxplot da Informação Mútua .....	48
Figura 48 – Boxplot do Índice de Qualidade de Fusão .....	49
Figura 49 – Boxplot do Índice de Semelhança Estrutural .....	50

# Lista de Símbolos

## Símbolos Romanos

- $(i, j)$  Coordenadas do pixel.
- $A$  Matriz (Letra maiúscula a negrito).
- $a$  Vetor (Letra minúscula a negrito).
- $a$  Escalar (Letra minúscula).
- $H, \tilde{H}$  Filtro passa-alto.
- $I$  Imagem fonte.
- $I_{vis}$  Imagem visível.
- $I_{term}$  Imagem térmica.
- $I_{fund}$  Imagem fundida.
- $\bar{I}$  Média de uma imagem.
- $I_L, I_H$  Matrizes de coeficientes obtidas por subamostragem de colunas.
- $I_{LL}, I_{LH}, I_{HL}, I_{HH}$  Matrizes de coeficientes obtidas por subamostragem de linhas, coeficientes de aproximação e detalhe (horizontal, vertical e diagonal).
- $L, \tilde{L}$  Filtro passa-baixo.
- $LP_l$  n-ésimo nível da pirâmide laplaciana.
- $LP_N$  Nível máximo da pirâmide laplaciana.
- $LP_l^{vis}$  n-ésimo nível da pirâmide laplaciana da imagem visível.
- $LP_l^{term}$  n-ésimo nível da pirâmide laplaciana da imagem térmica.
- $LP_l^{fund}$  n-ésimo nível da pirâmide laplaciana da imagem fundida.
- $LP_N^{fund}$  Nível máximo da pirâmide laplaciana da imagem fundida.
- $n_l$  Número de linhas de uma imagem.
- $n_c$  Número de colunas de uma imagem.
- $n_{l_l}$  Número de linhas de uma imagem no n-ésimo nível da pirâmide.
- $n_{c_l}$  Número de colunas de uma imagem no n-ésimo nível da pirâmide.
- $p(fund, vis)$  Histograma conjunto da imagem fundida e da imagem visível.
- $p(fund, term)$  Histograma conjunto da imagem fundida e da imagem térmica.
- $p(i)$  Frequência normalizada.
- $Q_o$  Índice global de qualidade de duas imagens.
- $Q_w$  Índice de qualidade de uma imagem.
- $S(I_{vis})$  Saliência da imagem visível.
- $S(I_{term})$  Saliência da imagem térmica.
- $X_l$  n-ésimo nível da pirâmide gaussiana.
- $X_N$  Nível máximo da pirâmide gaussiana.
- $\hat{X}_l$  n-ésimo nível da pirâmide gaussiana expandido.
- $X_l^{vis}$  n-ésimo nível da pirâmide gaussiana da imagem visível.

- $X_l^{term}$  n-ésimo nível da pirâmide gaussiana da imagem térmica.  
 $X_l^{fund}$  n-ésimo nível da pirâmide gaussiana fundida.  
 $X_N^{fund}$  Nível máximo da pirâmide gaussiana fundida.

### **Símbolos Gregos**

- $\lambda$  Peso local.  
 $\mu$  Média de intensidade de uma imagem.  
 $\sigma$  Desvio padrão.  
 $\sigma^2_I$  Variância de uma imagem.  
 $\sigma_{I_{vis}I_{term}}$  Covariância.  
 $\phi$  Regra de fusão.  
 $\omega$  Função de custo.

## Lista de Acrónimos e Siglas

AM	Academia Militar
DCT	<i>Discrete Cosine Transform</i>
DWT	<i>Discrete Wavelet Transform</i>
EM	<i>Expectation Maximization</i>
IDWT	<i>Inverse Discrete Wavelet Transform</i>
ISE	Índice de Semelhança Estrutural
MDCT	<i>Multi-Resolution Discrete Cosine Transform</i>
MSD	<i>Multi-Scale Decomposition</i>
NMSD	<i>Non-Multi-Scale Decomposition</i>
PCA	<i>Principal Component Analysis</i>
PSNR	<i>Peak Signal Noise Ratio</i>
RAM	<i>Random Access Memory</i>
RBF	<i>Radial Base Function</i>
RMSE	<i>Root Mean Square Error</i>
SVM	<i>Support Vector Machine</i>

# 1 Introdução

Este capítulo apresenta uma visão geral da dissertação e da sua estrutura. Primeiramente é efetuado um enquadramento do tema e apresentado a motivação respetiva, ou seja, são explicados os motivos que levaram à sua realização. De seguida, são explanados os objetivos que se pretendem atingir com o presente estudo e, por último, é feita uma breve descrição do conteúdo de cada capítulo.

## 1.1 Enquadramento e Motivação

Segundo a Publicação Doutrinária do Exército 2.00 [1], a Segurança é definida como “a condição obtida quando a informação, o material, o pessoal, as atividades e as instalações estão protegidos contra a espionagem, a sabotagem, a subversão e o terrorismo, assim como contra perdas ou divulgações não autorizadas”. Assim, a Segurança “garante ao Comandante, liberdade para planear e conduzir operações, reduzindo as vulnerabilidades às ameaças e atividades hostis. De igual modo, garantindo a segurança, o comandante ficará em melhores condições de conduzir as operações usando a surpresa. A Segurança contribui, de uma forma geral, para a Proteção da Força, garantindo a manutenção do seu potencial de combate”.

Associado ao conceito de Segurança, aparece o conceito de Segurança Militar, que é definido como o “conjunto de medidas e disposições ativas e passivas, tomadas como proteção contra interferências, observação, espionagem e sabotagem, com a finalidade de impedir a surpresa, garantir a liberdade de ação, preservar o segredo e evitar a subversão, a captura ou aniquilamento do pessoal, bem como a destruição, dano ou roubo do material e instalações” [2]. Estas medidas são chamadas de Medidas de Segurança, e têm como principal objetivo neutralizar as ameaças do Inimigo. Ou seja, as “atividades do serviço de informações do inimigo podem ser contrariadas por medidas passivas, destinadas fundamentalmente a evitar as violações de segurança, e por medidas ativas destinadas à deteção e neutralização dos agentes inimigos” [2].

Dentro das medidas passivas inserem-se atividades tais como salvaguardar documentos, segurança das transmissões, controle de acessos, entre outros. Nas medidas ativas tem-se a detenção de agentes inimigos que procurem informações ou a procura de um elemento das Nossas Forças que fale inconscientemente. As medidas passivas são mais importantes que as ativas, uma vez que prevenir é melhor que remediar [3], [1]. Portanto, as Medidas de Segurança visam impedir tentativas de intrusão nas áreas controladas, por parte de indivíduos não autorizados; negar qualquer oportunidade aos Serviços de Informações hostis, indivíduos, grupos ou organizações terroristas ou subversivas, de obter informações, provocar perturbações ou subversões militares ou civis; estabelecer normas de segurança comuns, aplicáveis a todos os órgãos e unidades na máxima extensão possível; e apoiar as investigações de segurança sempre que haja comprometimentos [3].

Neste âmbito, de forma a contrariar o Serviço de Informações do Inimigo, surge um conceito importante, o de Contrainformação que se caracteriza como “o conjunto de atividades relacionadas com a identificação e neutralização da ameaça à segurança, constituída por serviços de informações

e organizações hostis ou por indivíduos relacionados com espionagem, sabotagem, subversão ou terrorismo” [2]. À medida que se procura alcançar uma maior segurança e vigilância, tornam-se necessários sistemas de vigilância mais robustos e sofisticados.

A maioria dos sistemas de vigilância atuais utiliza sistemas de detecção através da cor, no entanto, estes estão muito limitados pela iluminação. Por esse motivo, tem sido proposto o uso de câmaras do espectro infravermelho-térmico, que capturam a imagem térmica do objeto. Desta forma, a utilização de imagens térmicas tem vindo a constituir um benefício para as Forças Armadas, devido à sua capacidade de utilização diurna e noturna, em diversas condições atmosféricas e, devido às suas inúmeras vantagens, têm um grande número de aplicações. Estas imagens são popularmente utilizadas pelo Exército e pela Marinha para a vigilância de fronteiras, ou vigilância costeira, e estabelecimento da ordem [4]. Neste contexto, podem ser usadas também para detetar pessoas camufladas.

No entanto, para melhorar os resultados da detecção de pessoas, a informação capturada por câmaras do espectro visível pode ser combinada com a informação capturada pelas câmaras do espectro infravermelho-térmico. Enquanto as imagens do espectro visível fornecem um contexto visual aos objetos, neste caso as pessoas, as imagens do espectro infravermelho-térmico fornecem informação acerca de objetos com temperaturas elevadas. Assim, a fusão das duas imagens fornece uma boa perceção visual, e permite a detecção de pessoas mais facilmente. Deste modo, as tecnologias de fusão de imagem inserem-se nas medidas de Contrainformação, pois permitem aos militares obter uma vigilância, proteção e detecção necessárias à sua segurança contra a ameaça do Inimigo e, ao mesmo tempo, também permite aos militares obter informação acerca do Inimigo. Ao combinar a informação gerada por múltiplas fontes de imagem, a fusão permite um maior detalhe, resultando num maior grau de segurança.

Esta dissertação de mestrado surge da necessidade de implementar um método que permita a detecção de pessoas camufladas, visualizando-as enquanto estas estão enquadradas no seu ambiente, tornando ineficiente a sua camuflagem. Portanto, é apresentado um estudo onde se pretende fundir imagens da banda do visível com imagens da banda do infravermelho-térmico para obter uma imagem fundida com mais informação do que qualquer uma das imagens fonte vista separadamente, que sirva não só de Contrainformação contra potenciais ameaças inimigas como também de Informação contra o Inimigo.

Este tema tem sido abordado em diversos artigos e estudos ao longo dos últimos anos, ressaltando a sua importância e pertinência no que diz respeito à Segurança, quer de pessoas como de instalações.

## **1.2 Objetivos**

O objetivo da presente dissertação é a implementação de métodos de fusão de imagem ao nível do pixel para detetar pessoas camufladas. Depois de serem implementados métodos distintos, pretende-se discernir qual dos métodos obtém os melhores resultados. Para tal, primeiramente será

feita uma análise qualitativa, seguida de uma análise quantitativa, como forma de credibilizar a avaliação feita por inspeção visual.

Através da fusão de uma imagem da banda do visível com uma imagem da banda do infravermelho-térmico, obtém-se uma imagem que possui o detalhe da imagem da banda do visível e a informação térmica presente na imagem da banda do infravermelho-térmico, apresentando mais informação do que qualquer uma das imagens fonte vistas separadamente, e constituindo assim uma mais valia para a deteção de pessoas camufladas.

### **1.3 Estrutura da Dissertação**

Esta dissertação encontra-se dividida em cinco capítulos. O primeiro capítulo é composto pelo enquadramento do tema e a respetiva motivação, ou seja, as razões que justificam o seu desenvolvimento. É descrito o problema e são indicados os objetivos que se pretendem atingir com a realização do trabalho. Por último é feita uma descrição do conteúdo de cada capítulo.

No capítulo 2 é feita uma revisão da investigação e métodos desenvolvidos por diversos autores no que diz respeito a esta área de estudo, a fusão de imagem. Primeiramente é efetuado um resumo dos métodos desenvolvidos no que diz respeito à fusão de imagem ao nível do pixel e, por último, relativamente à fusão de imagem ao nível da característica e ao nível da decisão. Para cada um dos métodos, são apresentados os seus benefícios e resultados.

No capítulo 3 são expostos os métodos e algoritmos utilizados no presente estudo, que foram baseados em métodos desenvolvidos por outros autores. Primeiro, é feita uma breve explicação em que consiste a fusão ao nível do pixel, seguida da apresentação dos três métodos que serão utilizados, nomeadamente: a “análise em componentes principais” (*Principal Component Analysis – PCA*), a “pirâmide laplaciana” e a “análise em wavelets”. Após a apresentação dos métodos utilizados, são apresentadas as métricas de desempenho que vão ser utilizadas para realizar a análise quantitativa dos métodos.

No capítulo 4 são apresentados os resultados obtidos com cada um dos métodos utilizados. Em primeiro lugar, são expostos os principais resultados obtidos para cada um dos três métodos e as suas variantes, e são evidenciados os melhores resultados. De seguida, é feita uma análise qualitativa dos métodos, tendo por base dez conjuntos de três imagens, que foram previamente selecionadas como sendo as que obtêm melhores resultados subjetivos. Depois de efetuada a análise qualitativa, segue-se uma análise quantitativa, tendo por base 20 imagens fundidas obtidas para cada um dos métodos e as suas variantes, utilizando as métricas de desempenho descritas no capítulo anterior, particularmente: o desvio padrão, a entropia, a frequência espacial, a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural. Por último, é feita uma comparação dos resultados obtidos qualitativa e quantitativamente.

No quinto e último capítulo, é feita uma síntese do trabalho desenvolvido, ressaltando os resultados mais importantes, e relacionando-os com os objetivos propostos no primeiro capítulo. São apresentadas algumas recomendações resultantes do trabalho realizado, e algumas vantagens e

limitações das soluções implementadas. Por último, são sugeridos trabalhos futuros que possam ser feitos no âmbito deste estudo, e tendo este trabalho como base.

## 2 Estado Da Arte

No contexto da crescente procura por uma maior segurança e vigilância, tornam-se necessários sistemas de vigilância mais sofisticados e robustos. Atualmente, na maioria dos sistemas de vigilância são utilizados sistemas de deteção através da cor. Infelizmente, a aquisição de imagens de cor no espectro visível depende da iluminação e é limitada por fenómenos ambientais, tais como o nevoeiro ou o fumo. Principalmente em ambiente noturno, a informação visual captada pelas câmaras do espectro do visível é muito limitada, tornando difícil a vigilância utilizando apenas um sensor visual. Pelo contrário, as câmaras de infravermelhos capturam a imagem térmica do objeto. Por esse motivo, tem sido proposto o uso de câmaras do espectro infravermelho-térmico [5].

Para melhorar os resultados da deteção de pessoas, a informação de cor pode ser combinada com a informação térmica. Enquanto as imagens térmicas fornecem uma informação rica para objetos com temperaturas elevadas, as imagens no visível fornecem um contexto visual aos objetos. Assim, a fusão das duas imagens irá fornecer uma boa perceção à vista humana mesmo com reduzidas condições de luminosidade. Uma fusão inteligente da informação fornecida por ambos os sensores reduz erros de deteção, aumentando assim a eficiência e a robustez dos sistemas de vigilância. Para concretizar esta ideia, é necessário o desenvolvimento de uma técnica eficiente que possibilite a fusão de informação fornecida por dois sensores, um correspondente ao espectro do visível e outro do espectro do infravermelho-térmico [6].

Ao longo dos anos, têm sido propostas várias técnicas de fusão de imagem para fazer face à sua crescente procura, pois existem várias áreas que beneficiam deste processo, nomeadamente: área médica, militar, vigilância e navegação.

Este capítulo resume o trabalho desenvolvido por vários autores no que diz respeito à fusão de imagem. Na primeira secção, são apresentados os resultados obtidos noutros estudos e, na segunda secção, faz-se uma revisão dos métodos desenvolvidos por diversos autores para solucionar o problema da fusão de imagem, tendo em conta a classificação das técnicas de fusão.

### 2.1 Resumo de alguns trabalhos em fusão de imagem

Nesta secção, serão apresentados alguns resultados obtidos por outros autores, em estudos de fusão de imagem ao nível do pixel. Os primeiros três exemplos, ilustrados nas figuras 1 a 6, foram retirados dum estudo efetuado por Sadhasivam, Keerthivasan e Muttan [7], e os últimos dois exemplos, ilustrados nas figuras 7 e 8, foram retirados dum artigo elaborado por Zheng [8].

No primeiro exemplo, figura 1, pode-se observar uma imagem de um porta aviões que foi tirada durante o período noturno. A imagem da esquerda diz respeito à informação térmica, e a imagem da direita corresponde à informação visível.



Figura 1 – a) imagem térmica do porta-aviões b) imagem visível do porta-aviões (retirada de [7])

Na figura 2 observam-se três imagens fundidas, obtidas com diferentes métodos. Da esquerda para a direita, foram utilizados os seguintes métodos: PCA, a “transformada wavelet discreta” (*Discrete Wavelet Transform - DWT*) com a aplicação de PCA aos coeficientes de aproximação; PCA com a utilização do princípio do máximo. A primeira e segunda imagem são similares, e a terceira imagem é aquela que apresenta as zonas mais quentes (mais claras) mais destacadas. Os resultados obtidos para as métricas de desempenho utilizadas (entropia, índice de semelhança estrutural e informação mútua), apresentam os valores mais elevados para a imagem obtida com a PCA com a utilização do princípio do máximo, suportando a avaliação feita por inspeção visual.



Figura 2 – a) imagem fundida com a PCA b) imagem fundida com a DWT-PCA c) imagem fundida com PCA-máximo (retirada de [7])

No segundo exemplo, figura 3, pode-se observar uma imagem de uma estrada no período noturno. Tal como na figura 1, a imagem da esquerda diz respeito à informação térmica, e a imagem da direita corresponde à informação visível. É perceptível que na imagem térmica o cenário é muito mais detalhado, e na imagem visível as únicas áreas que aparecem destacadas são as correspondentes aos candeeiros e luzes dos automóveis.



Figura 3 - a) imagem térmica da estrada b) imagem visível da estrada (retirada de [7])

Tal como na figura 2, na figura 4 pode-se observar três imagens fundidas, com os respetivos métodos: PCA, DWT com a aplicação de PCA aos coeficientes de aproximação; PCA com a utilização do princípio do máximo. Neste exemplo, a primeira e segunda imagem mais uma vez são semelhantes, e a terceira imagem têm as zonas quentes mais destacadas. Tal como no exemplo anterior, as métricas de desempenho utilizadas apresentam os valores mais elevados para a imagem obtida com a PCA com a utilização do princípio do máximo, apoiando a avaliação feita por inspeção visual.



Figura 4 - a) imagem fundida com a PCA b) imagem fundida com a DWT-PCA c) imagem fundida com PCA-máximo (retirada de [7])

No terceiro exemplo, figura 5, pode-se observar uma imagem um campo, com uma casa e uma vedação, no período noturno. Tal como na figura 1 e na figura 3, a imagem da esquerda diz respeito à informação térmica, e a imagem da direita corresponde à informação visível. Na imagem térmica é notável que existe uma pessoa no lado de fora da vedação, enquanto na imagem visível essa pessoa não é perceptível. Por outro lado, na imagem visível os detalhes dos arbustos e da casa são mais visíveis.

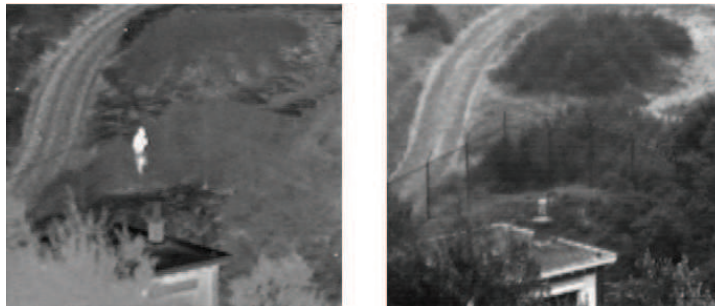


Figura 5 - a) imagem térmica do campo b) imagem visível do campo (retirada de [7])

Na figura 6 pode-se observar três imagens fundidas, com os respetivos métodos: PCA, DWT com a aplicação da PCA aos coeficientes de aproximação; PCA com a utilização do princípio do máximo. Tal como nos exemplos anteriores, a primeira e segunda imagem são similares, e a terceira imagem têm as zonas quentes mais destacadas. Tal como no primeiro e segundo exemplo, as métricas de desempenho utilizadas apresentam os valores mais elevados para a imagem obtida com a PCA com a utilização do princípio do máximo, suportando a avaliação feita por inspeção visual.



Figura 6 - a) imagem fundida com a PCA b) imagem fundida com a DWT-PCA c) imagem fundida com PCA-máximo (retirada de [7])

No quarto exemplo, figura 7, observam-se quatro imagens que correspondem a (da esquerda para a direita): imagem visível, imagem térmica, imagem fundida com o método da pirâmide laplaciana e imagem fundida com o método da DWT. Este exemplo é semelhante ao da figura 6. A imagem corresponde a um campo com uma casa e vedação. A pessoa que se encontra fora da vedação só é visível na imagem térmica. Entre as duas imagens fundidas, aquela que têm mais detalhe e contraste é a obtida com a pirâmide laplaciana. Os resultados numéricos obtidos, para as métricas de desempenho utilizadas, sustentam a avaliação feita por inspeção visual.

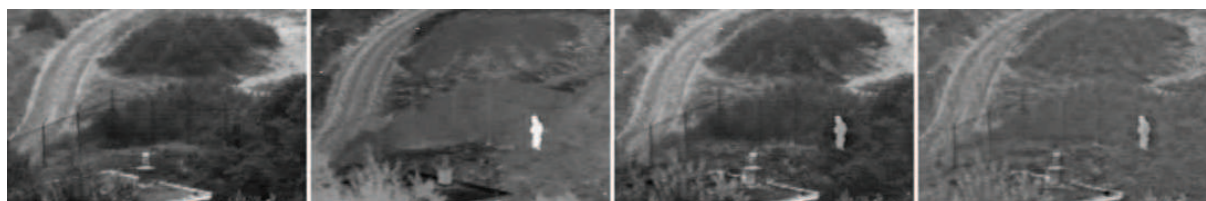


Figura 7 – a) imagem visível do campo b) imagem térmica do campo c) imagem fundida com pirâmide laplaciana d) imagem fundida com DWT (retirada de [8])

No quinto e último exemplo, figura 8, observam-se quatro imagens que correspondem a (da esquerda para a direita): imagem visível, imagem térmica, imagem fundida com o método da pirâmide laplaciana e imagem fundida com o método da DWT. Neste caso, as imagens correspondem a imagens médicas. Tal como na figura 7, a terceira imagem (obtida com a pirâmide laplaciana) é a que apresenta mais contraste e detalhe. Os resultados numéricos conseguidos, para as métricas de desempenho utilizadas, apoiam a avaliação feita por inspeção visual.

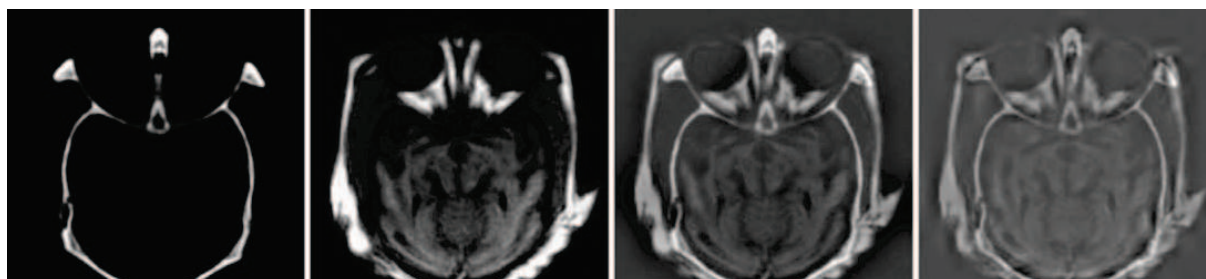


Figura 8 – a) imagem visível do campo b) imagem térmica do campo c) imagem fundida com pirâmide laplaciana d) imagem fundida com DWT (retirada de [8])

## 2.2 Fusão de Imagem

Existem duas formas de classificar as técnicas de fusão de imagem. A primeira forma de classificação consiste na distinção entre métodos de fusão baseados em “decomposição multi-escala” (*Multi-Scale Decomposition* - MSD) e métodos de fusão baseados em “decomposição não-multi-escala” (*Non-Multi-Scale Decomposition* - NMSD). A ideia básica dos métodos baseados em MSD consiste em fazer a transformação multi-escala nas imagens fonte, e depois a representação multi-escala destas imagens é construída baseando-se numa regra de seleção predeterminada. A imagem fundida é obtida através da transformada inversa da transformada multi-escala original. Os métodos mais conhecidos baseados em MSD incluem as transformadas em pirâmides e transformadas wavelet. Os métodos NMSD, como o próprio nome indica, não são baseados em transformadas multi-escala. Estes incluem PCA, técnica da média ponderada, métodos de estimação usando redes neurais artificiais.

A segunda forma de classificação das técnicas de fusão é baseada no nível de processamento onde a fusão toma lugar. A principal diferença entre os níveis é a quantidade de processamento que é feito na imagem antes da fusão. Assim, há três níveis onde a fusão de imagem pode tomar lugar: ao nível do pixel, ao nível da característica e ao nível de decisão [6]. Na fusão ao nível do pixel, a informação é adquirida por sensores, na forma de duas imagens, e estas são combinadas diretamente numa única imagem fundida que representa a informação presente nas duas imagens fonte. A fusão ao nível das características requer a extração de objetos das imagens fonte. Estes objetos são depois combinados com outros semelhantes presentes noutras imagens fonte, através de um processo de seleção, para formar a imagem final [6]. Na fusão ao nível de decisão, as saídas da deteção e classificação inicial dos objetos são usadas como entradas do algoritmo de fusão para desempenhar a integração dos dados.

Tendo em consideração que a fusão de imagem é um problema complexo e que pode ser abordado de diversas formas, tal como referido acima, neste trabalho é feita uma abordagem ao seu nível mais baixo, ou seja, ao nível do pixel.

Nas próximas subsecções faz-se uma revisão dos métodos desenvolvidos por diversos autores. A primeira subsecção aborda a fusão de imagem ao nível do pixel, a qual será adotada neste estudo; a segunda subsecção engloba a fusão de imagem ao nível das características e ao nível da decisão, que constituem alternativas à fusão ao nível do pixel.

### 2.2.1 Fusão de Imagem ao Nível do Pixel

Nesta secção apresentam-se várias abordagens e métodos implementados por diversos autores, no que concerne à fusão de imagem ao nível do pixel. Naidu e Raol [9] fazem uma comparação da fusão ao nível do pixel utilizando wavelets e PCA. Relativamente às wavelets, a regra de fusão utilizada foi a média simples para os coeficientes de aproximação e os coeficientes de detalhe com o maior valor absoluto, e foram testados cinco níveis de decomposição. Para efeitos comparativos, foi implementado o método de fusão usando a média simples. No seu estudo, foram implementadas

métricas, com e sem imagem de referência, para avaliar o desempenho dos algoritmos de fusão de imagem. Foram utilizados dois conjuntos de imagens com  $295 \times 400$  e  $512 \times 512$  pixels, respectivamente. O método de fusão utilizando a média simples mostrou um desempenho baixo e degradado. As métricas do desvio padrão, entropia cruzada e frequência espacial foram consideradas as adequadas quando não existe imagem de referência. Por último, concluiu-se que a fusão de imagem utilizando wavelets com um maior grau de decomposição possui melhor desempenho em algumas métricas, e noutras métricas PCA mostrou melhores resultados, nomeadamente no desvio padrão. Embora esses resultados sejam razoáveis para PCA, a imagem fundida tem pior qualidade, justificado pelos autores como sendo um efeito do contraste na imagem nos cálculos do desvio padrão. A frequência espacial é tida como o melhor indicador da qualidade da fusão independentemente da origem das imagens fonte.

Zheng [8] faz uma comparação de algoritmos de fusão multi-escala ao nível do pixel, nomeadamente: baseados em pirâmides, DWT e DWT Iterativa. Zheng propôs o algoritmo *Advanced Discrete Wavelet Transform* para a fusão de imagem. Neste algoritmo, os coeficientes de aproximação da maior escala da imagem fonte são fundidos ao aplicar PCA aos valores absolutos desses coeficientes. Posteriormente, foi otimizado com um procedimento iterativo utilizando as métricas de fusão: índice de qualidade de imagem e taxa de erro da frequência espacial. Neste estudo, foi implementada uma métrica de taxa de erro da frequência espacial modificada, considerando os gradientes de primeira e segunda ordem, que juntamente com a informação mútua constituem métricas de otimização para os dois algoritmos iterativos desenvolvidos. Para efeitos de comparação, foram implementados os algoritmos de fusão usando a pirâmide laplaciana e a DWT. Foram usados três pares de imagens. A avaliação das imagens fundidas foi feita qualitativamente e quantitativamente. Como métricas de avaliação quantitativa, o autor usou a entropia e frequência espacial. Os resultados quantitativos mostram que os algoritmos iterativos apresentam melhores desempenhos, seguidos da pirâmide laplaciana e por último a DWT.

Sadhasivam, Keerthivasan e Muttan [7] fazem a fusão de imagem via PCA utilizando o princípio do máximo, para a aplicação em sistemas de vigilância e navegação. Dado que os resultados obtidos a partir da PCA mostram um desempenho baixo quando comparados com outros algoritmos híbridos, neste estudo os autores implementam um algoritmo que utiliza a DWT em conjunto com PCA. O coeficiente de baixa frequência é escolhido de acordo com a regra do máximo, e vai ser aplicado PCA aos coeficientes de alta frequência para determinar os seus pesos para a fusão. A imagem final é obtida ao adicionar as imagens de baixa e alta frequência. Foram utilizados três conjuntos de imagens e o desempenho deste algoritmo foi medido através da entropia, da informação mútua e de uma medida baseada no índice de semelhança estrutural. Enquanto a entropia apresenta valores semelhantes para os métodos comparados, a semelhança estrutural e a informação mútua na imagem fundida apresentam melhores resultados para o método proposto.

Numa outra abordagem [10], é feito um estudo comparativo sobre a fusão dos coeficientes wavelet de imagens visíveis e térmicas para o reconhecimento de faces. São utilizadas transformadas wavelet diferentes, nomeadamente as Haar e Daubechies wavelets, e os coeficientes das imagens visíveis e térmicas são calculados separadamente e depois combinados. Depois é feita a transformada wavelet inversa para obter as imagens fundidas. O desempenho do algoritmo é analisado usando uma

base de dados visível e térmica de faces. Nesta base de dados, todas as imagens visíveis e térmicas das faces apresentam várias condições de luminosidade, várias expressões e poses. Ambas as imagens, visíveis e térmicas, têm dimensões de  $320 \times 240$  pixels, e foram utilizados 200 pares de imagens. Os resultados obtidos mostram que em muitos dos casos testados, para qualquer uma das duas transformadas, a taxa de desempenho atinge os 100%, ou seja, existe quase sempre uma correspondência entre a imagem fundida e uma imagem da base de dados.

Zheng, Essock e Hansen [11] desenvolvem um algoritmo que incorpora PCA na DWT, semelhante à abordagem seguida em [8]. A PCA é aplicada aos coeficientes de aproximação, ou seja, os coeficientes de baixa frequência, enquanto os coeficientes de detalhe são escolhidos de acordo com o maior valor absoluto. O algoritmo proposto é comparado com outras técnicas de fusão utilizando a entropia, a frequência espacial e o índice de qualidade de imagem, para os casos em que não existe imagem de referência. Nesse algoritmo foram selecionadas as *Symlets* wavelets e na DWT foram utilizadas as *Daubechies* Wavelets. São utilizados cinco grupos de imagens, com dimensões de  $512 \times 512$ ,  $256 \times 256$  e  $220 \times 220$  pixels. Em qualquer um dos grupos de imagens, o algoritmo proposto obteve os melhores resultados.

Tan, Huang, Huachun Tan e He [12] implementam um algoritmo de fusão de imagem ao nível do pixel baseado na máxima verosimilhança e na decomposição da pirâmide laplaciana. As imagens fonte são decompostas utilizando a pirâmide laplaciana e depois é utilizado o método da máxima verosimilhança para fundir as componentes da imagem nas bandas de baixa frequência. As bandas de alta frequência são fundidas usando o princípio do máximo. A imagem fundida final é obtida via transformada inversa da pirâmide laplaciana. O algoritmo foi testado em dois conjuntos de imagens, com dimensões de  $256 \times 256$  pixels, e a comparação quantitativa dos métodos foi feita usando a entropia, o gradiente médio da imagem fundida, a informação mútua entre as imagens e a frequência espacial da imagem fundida. Os resultados obtidos para as métricas utilizadas mostram que o método proposto obtém melhor desempenho do que os métodos de fusão convencionais tais como a DWT, o método de decomposição da pirâmide laplaciana, etc.

Naidu apresentou em [13] um método de fusão de imagem baseado na “transformada discreta do cosseno multi-resolução” (*Multi-Decomposition Cosine Transform* - MDCT). A “transformada discreta do cosseno” (*Discrete Cosine Transform* - DCT) importante no processamento de imagem. Os coeficientes DCT maiores estão concentrados na região de baixa frequência; por esse motivo, é conhecida pelas suas propriedades de compacidade de energia. A MDCT é muito semelhante à transformada wavelet, onde o sinal é filtrado separadamente por um filtro passa-baixo seguido de um filtro passa-alto e, à saída de cada filtro, o sinal é decimado por um fator de dois para alcançar o primeiro nível de decomposição. A regra de fusão utilizada nos coeficientes MDCT de detalhe é o maior valor absoluto, e nos coeficientes de aproximação é a média. Este método foi comparado com o método baseado na DWT, utilizando a percentagem de ajuste do erro, a “razão de pico sinal ruído” (*Peak Signal Noise Ratio* - PSNR), a medida de semelhança estrutural e, para os casos onde não existe imagem de referência, o desvio padrão, a entropia cruzada e frequência espacial. Foram utilizados dois conjuntos de imagens e os resultados mostram que este método obtém um desempenho muito semelhante ao

obtido com as wavelets; e é adequado para aplicações em tempo real. O autor concluiu também que a fusão com um nível mais elevado de decomposição obtém melhores resultados.

Naidu e Elias [14] propuseram uma nova técnica de fusão de imagem usando a pirâmide laplaciana baseada na DCT, e decompuseram as imagens fonte de um nível até oito níveis de decomposição. Dado que existe uma imagem de referência, as métricas utilizadas para a análise quantitativa foram a “raiz do erro quadrático médio” (*Root Mean Square Error* - RMSE), PSNR, frequência espacial e desvio padrão. Foi usado um conjunto de imagens, com dimensão de 512×512 pixels. Concluíram que a fusão com um nível mais alto de decomposição alcança melhores resultados, para qualquer uma das métricas usadas. O algoritmo proposto por Zheng e Elias também pode ser estendido a imagens a cores e pode ser usado em aplicações em tempo real.

Mandhare, Upadhyay e Gupta [15] implementaram métodos de fusão de imagem ao nível do pixel baseados na transformada de Brovey e na DWT, com o objetivo de testar a sua capacidade de melhorar a informação espacial e espectral. Com esse propósito implementaram os métodos da média amostral, o método multiplicativo, o método Brovey e o método baseado na DWT. Para avaliar o desempenho de cada um destes métodos foram utilizados parâmetros de avaliação tais como: entropia, desvio padrão, RMSE e PSNR. Foi utilizado um par de imagens com dimensões de 1000×1000 pixels. Os resultados obtidos para as métricas utilizadas mostram que as imagens fundidas através dos métodos multiplicativo e baseado em wavelets obtém uma resolução espacial mais elevada e têm características espectrais melhores do que as imagens originais.

Rockinger [16] introduziu uma alteração ao método de fusão baseado na DWT. Propôs uma nova abordagem à fusão de imagens sequenciais, que incorpora uma extensão *shift invariant* da DWT, que fornece uma representação sobre completa do sinal. A vantagem do método proposto é o melhoramento da estabilidade temporal e consistência da sequência fundida ao ser comparada com outros métodos de fusão. É utilizada uma medida quantitativa baseada na informação mútua para verificar a estabilidade e consistência do método proposto. Foram utilizados três conjuntos de imagens, e os resultados numéricos adquiridos indicam que o método proposto tem um desempenho superior ao método baseado em DWT e é ideal para sequências de imagens.

## **2.2.2 Fusão de Imagem ao Nível das Características e ao Nível da Decisão**

Nesta secção apresentam-se diversos métodos de fusão ao nível da característica e da decisão. Khan, Fayyaz e Gilani [17] apresentaram um novo algoritmo baseado na fusão de imagens do espectro visível e de infravermelhos. Neste caso, as características lineares e não-lineares são tratadas separadamente. Primeiro, a imagem é transformada para o domínio da transformada wavelet, separando as características lineares das não-lineares, e fundindo-as utilizando diferentes conjuntos de regras de fusão. O principal problema inerente à fusão de imagens do espectro do visível e térmicas é o facto de que esta fusão depende do conteúdo das baixas frequências, presente nos objetos que aparecem nas imagens térmicas, e do conteúdo das altas frequências, tais como os detalhes nas imagens do espectro visível. Assim, para explorar o conteúdo das baixas frequências é utilizada PCA não linear via *Radial Base Function* (RBF) kernels, enquanto para explorar o conteúdo das altas

frequências foi desenvolvido um modelo baseado numa “máquina de suporte vetorial” (*Support Vector Machine* - SVM). A técnica incorpora, com sucesso, um classificador linear para fundir as características lineares e kernels RBF para fundir as características não-lineares dos dados. Conseqüentemente, a imagem resultante é melhor tanto subjetivamente como objetivamente. Esta metodologia foi aplicada em 32 pares de imagens, mas dado que o desempenho do sistema está fortemente dependente do treino do classificador SVM, um treino mais exaustivo implica resultados mais estáveis. Segundo os resultados obtidos, este método é bastante eficaz em aplicações em tempo-real e aplicações médicas, tais como ultrassons e ressonância magnética.

Carson, Belongie, Greenspan e Malik [18] propuseram uma técnica de segmentação de imagem utilizando um algoritmo de *Expectation Maximization* (EM). Apresentam uma nova representação de imagem, na qual a imagem está dividida num conjunto pequeno de regiões que são coerentes em cor e textura. Esta representação é conseguida ao agrupar os pixels num espaço de características comuns de cor-textura-localização. Para a segmentação automática de cada imagem, foi modelada uma distribuição conjunta de características de cor, textura e posição com uma mistura de gaussianas usando EM. O algoritmo de EM é utilizado para estimar os parâmetros do modelo. Os agrupamentos de pixels resultantes proporcionam a segmentação da imagem e, depois de estar segmentada em regiões, é produzida uma descrição das características de cor e textura de cada região. Com esta metodologia, ao encontrar as regiões da imagem correspondentes aos objetos, é permitida a pesquisa ao nível dos objetos ao invés das características globais da imagem. O algoritmo de segmentação foi testado numa coleção de 10000 imagens naturais, os resultados indicam que a pesquisa de objetos distintos, usando esta técnica, fornece uma maior precisão do que a pesquisa usando os histogramas de cor e textura da imagem completa.

Davis e Sharma [19] sugeriram uma nova técnica de subtração do plano de fundo, para extrair objetos no plano principal, baseada no contorno em imagens térmicas. Primeiro, é utilizada uma subtração de plano de fundo para identificar as regiões de interesse locais. Dentro de cada região, a informação de entrada (imagem fonte) e do gradiente do plano de fundo é combinada para formar o mapa de saliência do contorno. Este contorno é depois afinado e é aplicado um limiar para formar segmentos. De seguida, é feita uma pesquisa ao longo do contorno para completar e fechar quaisquer segmentos que não estivessem ligados. Finalmente, o contorno da imagem é preenchido para produzir silhuetas. Esta abordagem é mais estável e robusta para diferentes condições ambientais e distintos meios circundantes e está preparada para lidar com a presença indesejada de halos nas imagens térmicas, que são o efeito em forma de auréola (mais clara ou mais escura) que por vezes aparece à volta de objetos quentes ou frios. A generalidade deste método foi demonstrada ao usar um único conjunto de parâmetros ao longo do conjunto de dados. Foram testadas seis seqüências de vídeo em diferentes condições, e foram segmentadas manualmente 30 imagens para efeitos de comparação. Os pixels correspondentes a pessoas detetadas coincidiram bastante bem com as silhuetas segmentadas manualmente. A abordagem não se limita a extrair apenas silhuetas de pessoas. Também foi demonstrada a extração de silhuetas de cães e veículos. Além disto, foi demonstrada a sua aplicação em imagens a cores.

Davis e Keck [20] mostram uma abordagem de duas fases para detetar pessoas em imagens térmicas, que combina um método especial de subtração do plano de fundo com uma técnica de classificação que utiliza um classificador AdaBoost. Primeiro, o método faz um procedimento de seleção baseado na correlação para produzir uma localização aproximada das pessoas na imagem. Para melhorar a taxa de deteção, é empregue uma técnica de subtração do plano de fundo baseada na temperatura. As regiões candidatas são examinadas mais profundamente por um classificador AdaBoost. Este classificador (AdaBoost) utiliza um conjunto de classificadores simples que, quando utilizados em conjunto, fornecem uma melhor classificação. Cada classificador tem um peso consoante a sua taxa de erro no conjunto de dados de treino. Os filtros utilizados são modelados a partir dos pesos calculados para os classificadores ao invés de serem selecionados a partir de um conjunto de filtros pré-definidos, ou seja, para cada exemplo de treino positivo ou negativo são criadas quatro imagens de características. Depois, é feita uma média ponderada das imagens de características para cada classe. O filtro ótimo do classificador AdaBoost é escolhido ao encontrar a sub-região de uma das quatro imagens de características que forneça a menor taxa de erro quando aplicada como filtro às imagens de características de treino. O conjunto de classificação resultante é então reduzido numa tentativa de fornecer uma única deteção por pessoa. Esta abordagem foi demonstrada com um conjunto de imagens térmicas de  $360 \times 240$  pixéis, em diversas condições atmosféricas. Os resultados mostram que, com esta abordagem, são necessários nove filtros/classificadores para atingir os 100% de classificação correta de pessoas.

Num artigo elaborado por Deng, Manjunath e Shin [21], é exposta uma nova forma de segmentação de imagens de cor, designada por JSEG. Primeiro, as cores da imagem são quantificadas para representar diferentes classes que são usadas para diferenciar as regiões na imagem. De seguida, os pixéis de cor da imagem são substituídos pelas etiquetas correspondentes às suas classes, formando assim um mapa de classes da imagem. Normalmente, cada região da imagem contém pixéis de um subconjunto mais pequeno das classes de cores, e cada classe é distribuída em algumas regiões da imagem. O mapa de classes pode ser visto como um conjunto de pontos espaciais localizados num plano de duas dimensões. O valor de cada ponto é a sua posição dentro da imagem. Estes pontos são classificados e a cada ponto é atribuída uma etiqueta, que é o valor do mapa de classes na posição da imagem. O critério para uma boa segmentação é calculado através do quociente entre a distância entre as diferentes classes e a distância entre membros da mesma classe. Um valor elevado significa que as classes estão mais separadas umas das outras, e os membros dentro da mesma classe estão mais próximos e vice-versa. Uma boa segmentação acontece quando dentro de cada região apenas existe uma classe de pontos. Para um número fixo de regiões, uma segmentação melhor tende a ter um valor menor para o critério. Se a segmentação é boa, cada região segmentada contém um conjunto de classes de cor uniformemente distribuídas, e o valor resultante é pequeno. Ao aplicar este critério às janelas locais do mapa de classes, têm-se um bom indicador se a área está no centro da região ou perto das fronteiras da região. Quanto maior for o valor do critério, mais provável é que o ponto esteja perto da fronteira da região. O tamanho das janelas locais determina o tamanho das regiões que podem ser detetadas; janelas pequenas são úteis para localizar fronteiras de intensidade e cor, enquanto janelas grandes são indicadas para detetar fronteiras de textura. Finalmente, é utilizado o método de

crescimento de regiões para segmentar a imagem. Os resultados fornecem bons resultados, ou seja, imagens bem segmentadas, mas devido à falta de imagens para comparação não é possível fazer uma avaliação objetiva ou uma comparação com outros métodos de segmentação. Este método também fornece bons resultados quando aplicado a imagens na escala dos cinzentos quando os valores da intensidade são quantificados da mesma forma das cores.

### 3 Métodos

Neste capítulo, apresenta-se uma breve explicação dos métodos utilizados neste estudo. Primeiramente será feita uma rápida explicação em que consiste a fusão ao nível do pixel, seguida da apresentação dos três métodos utilizados: PCA, pirâmide laplaciana e wavelets. Por último, serão apresentadas as métricas de avaliação de desempenho que são utilizadas para a análise quantitativa.

#### 3.1 Enquadramento

Ao longo dos anos têm sido desenvolvidos inúmeros algoritmos de fusão de imagem para corresponder à sua crescente necessidade. São várias as áreas que beneficiam desta técnica, nomeadamente: a vigilância, aplicações de navegação, aplicações militares, robótica e imagens médicas.

O processo de fusão de imagem pode ser definido como a integração de informação de um número de imagens registadas sem a introdução de distorção [7]. Um dos objetivos da fusão de imagem é combinar informação relevante de duas ou mais imagens originais numa única imagem, de tal forma que essa única imagem contenha a maior informação possível das imagens originais. Desta forma, a imagem fundida corresponderá de uma forma mais precisa ao cenário real e será mais útil, nomeadamente, à percepção visual humana [12].

A imagem de infravermelhos é um mapa das radiações infravermelhas que são emitidas por objetos que tenham uma temperatura acima de zero graus Kelvin. A quantidade de informação transmitida está intimamente ligada com a temperatura do objeto. Assim, objetos quentes e seres vivos, tais como os humanos e animais, podem ser detetados facilmente numa imagem de infravermelhos. Na figura 9 pode-se observar um diagrama de radiação de Plank, onde a curva dos 300 graus Kelvin corresponde aproximadamente à temperatura dos corpos. Contudo, para representar objetos tais como árvores, folhas ou relva, as câmaras da banda do visível são as mais indicadas. Por esse motivo há a necessidade da integração da informação visual com a informação térmica, sendo esta fusão muito útil em diversas situações [7].

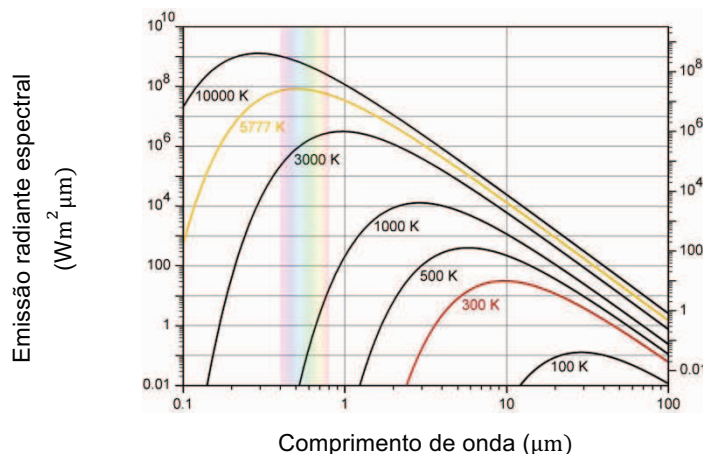


Figura 9 – Diagrama de radiação de Plank (adaptada de [22])

### 3.1.1 Fusão ao Nível do Pixel

Quando se fala em fusão ao nível do pixel significa que a fusão é efetuada ao nível mais baixo de processamento, onde o elemento fundamental é o pixel. Dentro dos três níveis de fusão, a fusão ao nível do pixel é a mais madura e engloba a maioria dos algoritmos de fusão.

Primeiro, todas as imagens são alinhadas e depois o processamento é realizado ao longo de todos os pixels das imagens fonte. Assim, para realizar a fusão ao nível do pixel todas as imagens fonte necessitam de estar registadas, ou seja, cada pixel das imagens a fundir representa a mesma localização. A fusão de imagens deve cumprir os seguintes requisitos:

- O processo de fusão deve preservar na imagem fundida toda a informação relevante contida nas imagens fonte;
- O processo de fusão não deve introduzir artefactos ou inconsistências;
- O processo de fusão deve ser invariante a translações e rotações, ou seja, o resultado de fusão não deve depender na localização ou orientação do objeto nas imagens fonte.

A figura 9 ilustra o fluxo de informação ao nível do pixel.

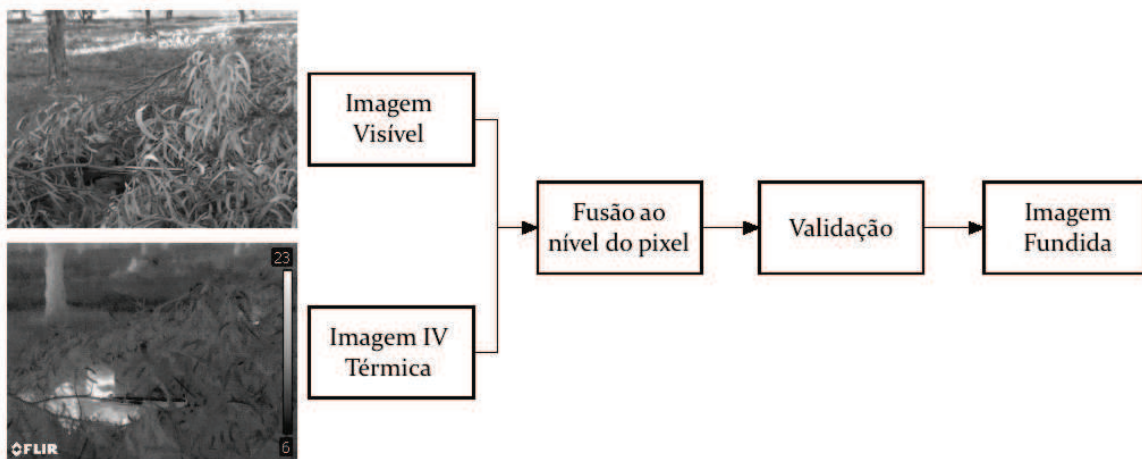


Figura 10 – Processo de fusão pixel a pixel (adaptada de [6])

Os algoritmos mais comuns de fusão ao nível do pixel são [6]: média amostral, PCA, esquemas de fusão em pirâmide e DWT. Embora a média amostral seja a forma mais básica de fusão, que consiste em calcular a média dos pixels das imagens originais na escala de cinzentos pixel a pixel, produz vários efeitos indesejados e um contraste de características baixo [9]. Portanto, torna-se necessário adotar outras técnicas de fusão mais elaboradas para um melhor resultado final.

Assim, este estudo vai-se basear na implementação de técnicas de fusão baseadas em técnicas estatísticas e em análise de multi-resolução, ou seja, em PCA, pirâmide laplaciana e em wavelets.

## 3.2 Análise em Componente Principais

A PCA, também chamada de transformada de Karhunen-Loève, envolve um procedimento matemático que transforma um número de variáveis correlacionadas num número de variáveis não-correlacionadas que são chamadas de componentes principais. É utilizada em compressão e classificação de imagem. A fusão de imagem, utilizando este método, é alcançada através de uma média ponderada das imagens a serem fundidas. Os pesos para cada imagem fonte são obtidos a partir dos vetores próprios correspondentes ao maior valor próprio das matrizes de covariância de cada imagem de fonte.

A primeira componente principal encontra-se na direção da máxima variância dos dados. A segunda componente principal tem a restrição de se encontrar no subespaço perpendicular à primeira. Dentro deste subespaço, esta componente aponta na direção da máxima variância, tal como se pode ver na figura 11.

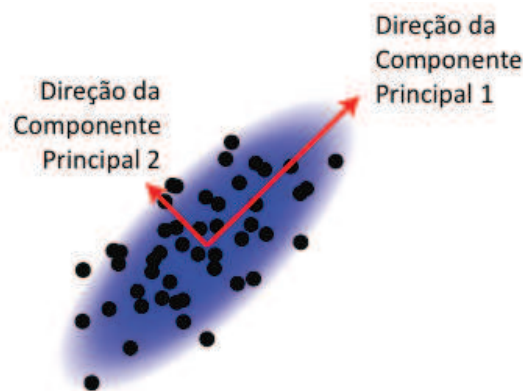


Figura 11 – Componentes principais

A terceira componente principal encontra-se na direção de máxima variância no subespaço perpendicular às duas primeiras, e assim sucessivamente.

### 3.2.1 Cálculo das Componentes Principais

As imagens fonte (imagens a serem fundidas) são organizadas em vetores de duas colunas. Os passos seguidos para projetar os dados num subespaço de duas dimensões são:

1. Organizar os dados numa matriz  $Z = [z_1, z_2, \dots, z_n]$ , em que  $z_i \in \mathbb{R}^2$  e  $z_i = \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix}$ .
2. Calcular  $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ .
3. Obter  $X = [x_1, x_2, \dots, x_n]$ , em que  $x_i = z_i - \bar{z}$ .
4. Calcular a matriz de covariância  $C = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$ , em que  $()^T$  demonstra a operação de transposição.

5. Calcular  $C = V \Sigma V^T$ , em que  $V = [v_1, v_2]$  e  $\Sigma = \text{diag}(\lambda_1, \lambda_2)$ . O par  $(v_i, \lambda_i)$  com  $i = 1, 2$  corresponde ao par vetor próprio / valor próprio de  $C$ , em que  $v_i = \begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix}$ .
6. De acordo com o maior valor próprio calcula-se:

$$p_{1i} = \frac{v_{1i}}{v_{1i} + v_{2i}} \text{ e } p_{2i} = \frac{v_{2i}}{v_{1i} + v_{2i}} \quad (1)$$

### 3.2.2 Fusão de Imagem

As imagens fonte  $I_{vis}$  e  $I_{term}$  são organizadas em dois vetores e as suas médias amostrais são subtraídas. Os vetores próprios e os valores próprios são calculados e os vetores próprios correspondentes aos maiores valores próprios são obtidos. A partir do vetor próprio são calculadas as componentes principais, de onde se obtêm dois valores,  $p_{1i}$  e  $p_{2i}$  [9].

O diagrama de fluxo de informação do algoritmo de fusão baseado em PCA é apresentado na figura 12.

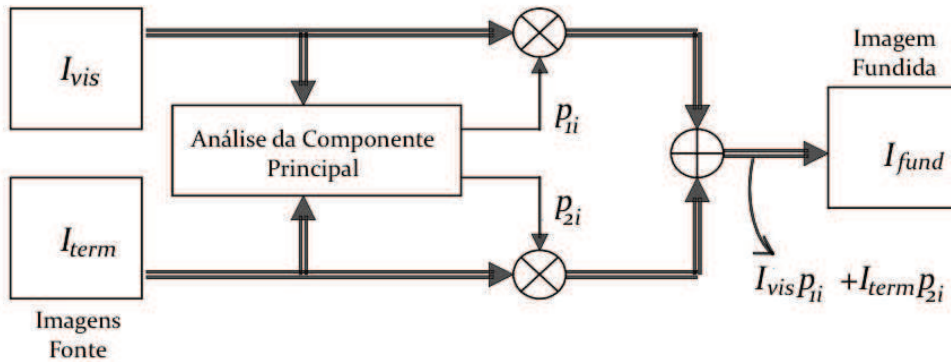


Figura 12 – Fluxo de informação do algoritmo baseado em PCA (adaptada de [9])

A imagem fundida é então dada por:

$$I_{fund} = p_{1i}I_{vis} + p_{2i}I_{term} \quad (2)$$

### 3.3 Pirâmide Laplaciana

A estrutura da pirâmide foi proposta para descrever a análise de uma imagem de multi-resolução, que é uma imagem que pode ser apresentada com diferentes resoluções, desde a sua resolução mais alta até à resolução mais baixa [11]. O princípio deste método consiste em decompor a imagem fonte em subimagens com diferentes resoluções espaciais, através de operações matemáticas. Por sua vez, a estratégia de fusão de imagem baseada em pirâmides consiste em usar uma regra de fusão para construir uma representação de uma pirâmide fundida a partir das representações em pirâmide das imagens fonte. A imagem fundida é obtida ao fazer a transformada inversa da pirâmide.

Ao longo dos anos têm sido propostos vários esquemas de fusão baseados em pirâmides, entre as quais: pirâmides laplacianas, pirâmides de gradiente e pirâmides morfológicas. Num estudo efectuado por Zheng, Essock e Hansen [11], entre as várias pirâmides testadas, a pirâmide laplaciana foi a que apresentou melhores resultados e, por esse motivo, é esta que foi implementada no presente estudo. Na pirâmide laplaciana, o nível mais baixo da pirâmide é construído a partir da imagem original e cada um dos restantes níveis é construído recursivamente a partir do seu nível mais baixo, ao aplicar quatro passos básicos [23], tal como se observa na figura 13:

1. *Blurring* (filtragem passa-baixo);
2. Sub Amostragem (reduzir o tamanho);
3. Expansão por Interpolação;
4. Diferenciação (subtrair duas imagens pixel a pixel).

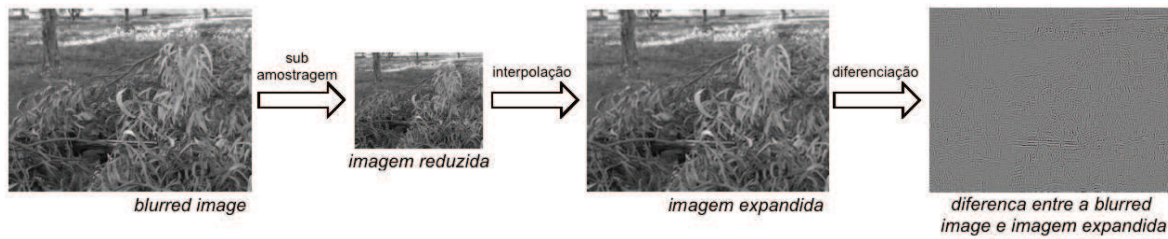


Figura 13 – Passos básicos da construção da pirâmide

A pirâmide laplaciana é derivada da pirâmide gaussiana. Constitui uma representação multi-escala e é obtida através duma filtragem recursiva passa-baixo e decimação, que consiste em reduzir a taxa de amostragem de um sinal (complementar à interpolação).

Admitindo que, a imagem fonte  $I_0$  com  $n_{l_0}$  linhas e  $n_{c_0}$  colunas corresponde ao nível zero da pirâmide, na base da pirâmide, e que  $X_l$  denota o  $l$ -ésimo nível da pirâmide gaussiana, que pode ser obtido pela equação (3), a qual é aplicada de forma recursiva para gerar todos os níveis da pirâmide.

$$X_l(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 \omega(m, n) X_{l-1}(m + 2i, n + 2j) \quad (3)$$

$$0 < l \leq N, 0 \leq i \leq n_{l_k}, 0 \leq j \leq n_{c_k}$$

Na equação (3),  $n_{l_l}$  e  $n_{c_l}$  são respetivamente a linha e coluna no  $l$ -ésimo nível da pirâmide, e  $N$  é o nível máximo da pirâmide.  $\omega(m, n)$  é chamada a função de custo, que satisfaz as seguintes três condições [12]: (i)  $\omega(m, n)$  satisfaz a condição  $\omega(m, n) = \hat{\omega}(m)\hat{\omega}^T(n)$ ,  $T$  é a operação de transposição, e  $\hat{\omega}$  representa um vetor coluna; (ii) a função  $\hat{\omega}$  de uma dimensão de comprimento 5 deve satisfazer a condição de normalização  $\sum_{m=-2}^2 \hat{\omega}(m) = 1$ ; (iii)  $\hat{\omega}$  também precisa de satisfazer a condição de simetria  $\hat{\omega}(i) = \hat{\omega}(-i)$  para  $i = 0, 1, 2$ .

Através do uso recursivo da equação (3), obtém-se  $N$  níveis da pirâmide,  $X_1, \dots, X_N$ . Dado que  $X_0$  corresponde ao nível mais baixo da pirâmide, o número total de níveis da pirâmide é  $N + 1$ , e o nível do topo é  $X_N$ , tal como se pode observar na figura 14.

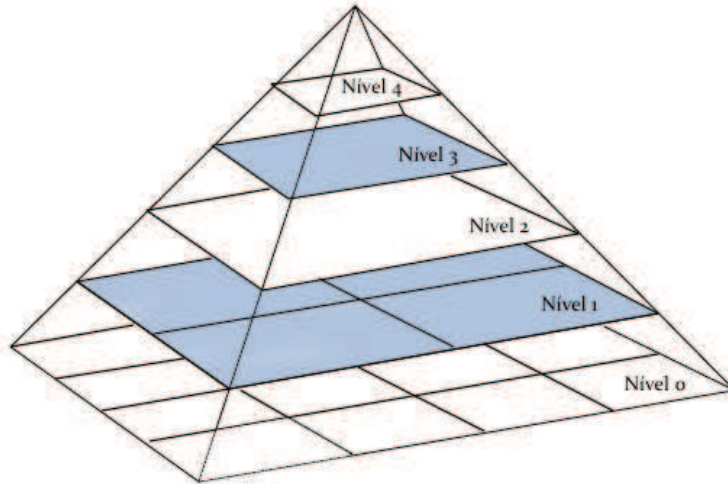


Figura 14 – Pirâmide laplaciana com 5 níveis

### 3.3.1 Decomposição de Imagem

De forma a derivar a pirâmide laplaciana, há duas operações padrão: a operação “Reduzir”, que é a versão simplificada da equação (3), e a operação “Expandir”, que é o inverso da “Reduzir”, cujo objetivo é expandir um array  $(M + 1) \times (N + 1)$  num array  $(2M + 1) \times (2N + 1)$ , ao preencher  $M$  zeros na direção horizontal e  $N$  zeros na direção vertical.

Assim, “Expandir” aplicado ao array  $X_l$  da pirâmide gaussiana gera um array  $\hat{X}_l$  com o mesmo tamanho que  $X_{l-1}$ . Os dois operadores são definidos por (ver Apêndice 1):

$$X_l = \text{Reduzir}(X_{l-1}) \quad (4)$$

$$\hat{X}_l = \text{Expandir}(X_l) \quad (5)$$

A pirâmide laplaciana é aplicada para decompor as imagens fonte em  $N$  níveis, onde  $LP_l$  corresponde ao  $l$ -ésimo nível da pirâmide. Assim, a pirâmide laplaciana  $LP_0, LP_1, \dots, LP_N$  é obtida a partir da seguinte definição:

$$\begin{cases} LP_l = X_l - \hat{X}_{l+1}, 0 \leq l < N \\ LP_N = X_N, l = N \end{cases} \quad (6)$$

A imagem fonte é decomposta num conjunto de resoluções em que a resolução espacial reduz gradualmente do nível mais baixo para o nível mais alto da pirâmide. A figura 15 ilustra o esquema de decomposição da imagem utilizando a pirâmide laplaciana, ou seja, a construção da pirâmide.

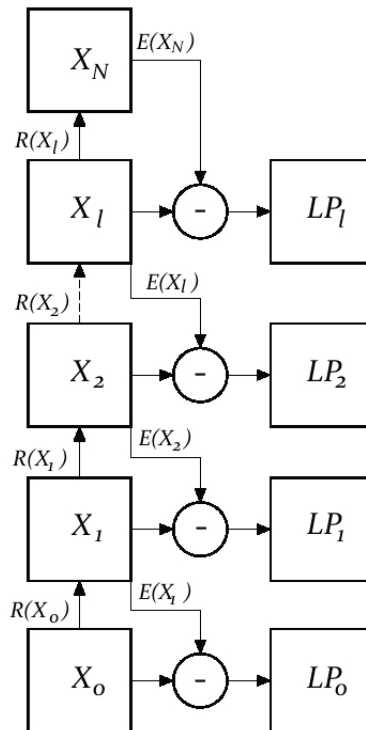


Figura 15 – Decomposição da imagem utilizando a pirâmide laplaciana (adaptada de [14])

### 3.3.2 Reconstrução da Imagem

A reconstrução da imagem a partir da pirâmide laplaciana é o processo inverso da decomposição e na direção inversa, ou seja, do nível do topo para o nível de base, com a seguinte definição:

$$\begin{cases} X_N = LP_N, l = N \\ X_l = LP_l + \hat{X}_{l+1}, 0 \leq l < N \end{cases} \quad (7)$$

A figura 16 representa a reconstrução da imagem.

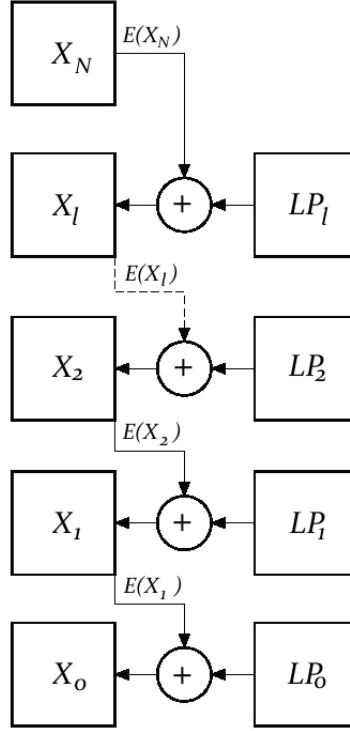


Figura 16 – Reconstrução da imagem utilizando a pirâmide laplaciana (adaptada de [14])

### 3.3.3 Fusão de Imagem

Existindo duas imagens para fundir,  $I_{vis}$  e  $I_{term}$ , a construção da pirâmide é feita para cada imagem individualmente. Considerando os  $N$  níveis construídos da pirâmide laplaciana para cada imagem, para a primeira tem-se  $P_l^{vis} = \{LP_0^{vis}, LP_1^{vis}, \dots, LP_l^{vis}, LP_N^{vis}\}$  e, da mesma forma, para a segunda imagem tem-se  $P_l^{term} = \{LP_0^{term}, LP_1^{term}, \dots, LP_l^{term}, LP_N^{term}\}$ . Ao nível mais alto, ou seja no nível  $N$ ,  $LP_N = X_N$ , dado que não há mais decomposição a partir deste nível.

Sabendo que a pirâmide laplaciana representa o limite de detalhe de cada imagem em cada nível, ao comparar os níveis correspondentes de imagens fonte, é possível obter a imagem fundida que junta os respectivos detalhes, e efetua a integração da imagem mantendo a quantidade de informação o mais rica possível [14].

Assim, ao  $n$ -ésimo nível da pirâmide temos a seguinte regra de fusão:  $X_{l+1}^{fund} = \frac{X_{l+1}^{vis} + X_{l+1}^{term}}{2}$ .

No nível  $N$  tem-se que  $X_N^{fund} = LP_N^{fund}$ , onde  $LP_N^{fund} = \frac{LP_N^{vis} + LP_N^{term}}{2}$ . Do nível  $N - 1$  até ao nível 0 da pirâmide, tem-se que  $X_l^{fund} = LP_l^{fund} + \widehat{X_{l+1}^{fund}}$ , onde  $LP_l^{fund} = \begin{cases} LP_l^{vis}, & |LP_l^{vis}| \geq |LP_l^{term}| \\ LP_l^{term}, & |LP_l^{vis}| < |LP_l^{term}| \end{cases}$ , e a comparação de amplitude é feita nos pixels correspondentes.

A pirâmide  $I_{fund} = X_0^{fund}$  é a imagem fundida. Na figura 17 encontra-se representado o esquema de fusão de duas imagens usando a pirâmide laplaciana.

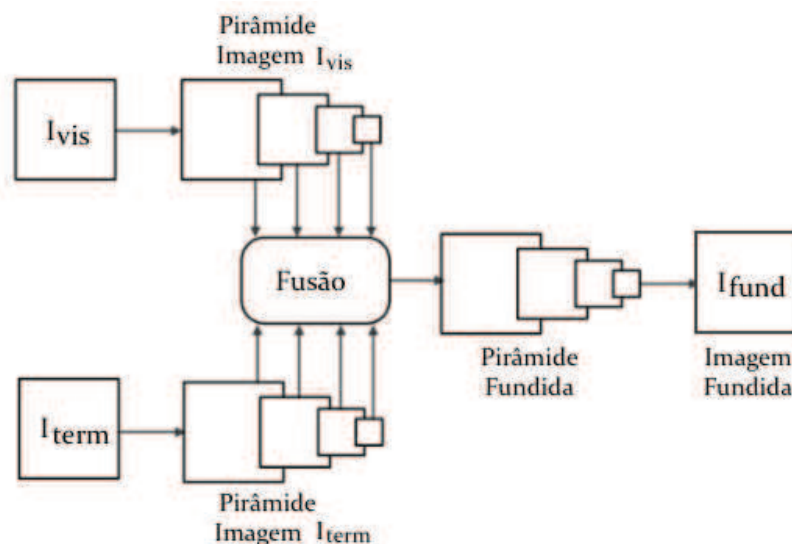


Figura 17 – Esquema de fusão de imagem com pirâmide laplaciana (adaptada de [24])

### 3.4 Wavelets

A teoria das wavelets é muito utilizada em processamento de imagem e fornece uma decomposição multi-resolução de uma imagem. É uma extensão da teoria de Fourier em vários aspectos. Na teoria de Fourier, o sinal é decomposto em senos e cossenos, mas nas wavelets o sinal é projetado num conjunto de funções wavelet. Estas funções wavelet são geradas por translação e dilatação da wavelet mãe.

Enquanto a transformada de Fourier fornece uma boa resolução no domínio da frequência, as wavelet dão uma boa resolução tanto do domínio do tempo como da frequência. Uma wavelet, como o seu próprio nome indica, é uma onda pequena que essencialmente cresce e decresce num período de tempo limitado. Para uma wavelet ser uma onda pequena tem de satisfazer duas propriedades básicas: o integral do tempo deve ser zero, e o quadrado do integral da wavelet ao longo do tempo deve ser unitário.

#### 3.4.1 Famílias de Wavelets

Existem várias famílias de wavelets [25]: *Haar*, *Daubechies*, *Coiflets*, *Symlets*, *Discrete Meyer*, *Biorthogonal* e *Reverse Biorthogonal*. No método implementado é usada a família Haar wavelet, por ser a mais simples e por ser aquela que obteve melhores resultados neste estudo específico.

A Haar wavelet foi a primeira wavelet conhecida, proposta por Alfred Haar, e faz parte das wavelets ortogonais, ou seja, é uma wavelet cuja transformada é ortogonal, e é equivalente à primeira wavelet da família das Daubechies. A transformada Haar é a transformada wavelet mais simples, e pode ser considerada um processo de amostragem no qual as linhas da matriz da transformada atuam como amostras de resolução cada vez mais simples [10].

As Daubechies Wavelets são a família de wavelets ortogonais que definem a DWT e são caracterizadas por um número máximo de *vanishing moments*, sendo que o número de *vanishing moments* limita a capacidade da wavelet representar o comportamento polinomial num sinal [26].

### 3.4.2 Decomposição de Imagem

Numa imagem de duas dimensões, para o nível um de decomposição, a informação flui tal como mostra a figura 18, onde  $I$  corresponde à imagem fonte.

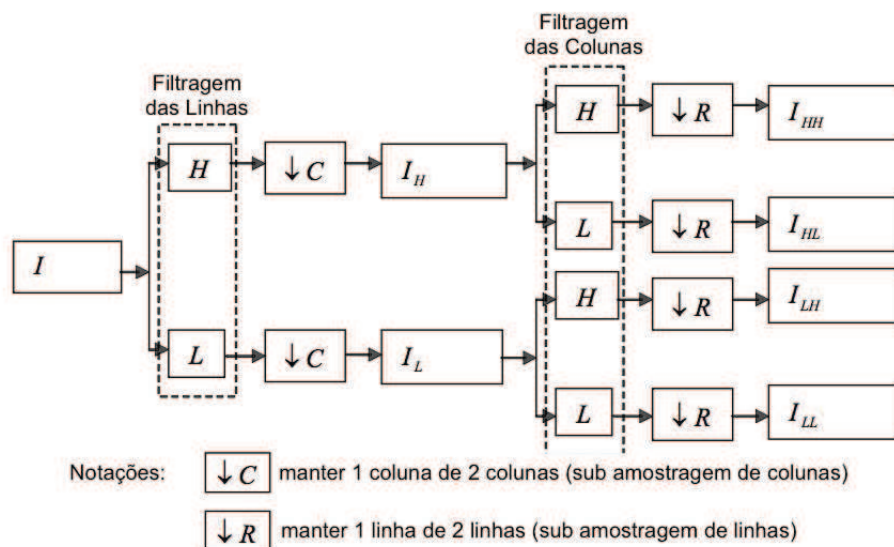


Figura 18 – Decomposição da imagem usando wavelets (adaptada de [9])

A wavelet filtra separadamente e subamostra a imagem nas direções vertical e horizontal. A imagem fonte  $I$  é filtrada por um filtro passa-baixo  $L$  e por um filtro passa-alto  $H$  na direção horizontal e depois é subamostrada por um factor de 2 para criar as matrizes de coeficientes  $I_L$  e  $I_H$ .

Ambas as matrizes de coeficientes  $I_L$  e  $I_H$  são filtradas por um filtro passa-baixo e passa-alto na direção vertical e subamostradas por um fator de 2 para criar as matrizes de coeficientes, ou subimagens,  $I_{LL}$ ,  $I_{LH}$ ,  $I_{HL}$  e  $I_{HH}$ . A matriz de coeficientes  $I_{LL}$  contém a média de informação da imagem correspondente à banda de baixa frequência da decomposição de multi-escala. Pode ser considerada uma versão suavizada e subamostrada da imagem fonte  $I$ . Representa a aproximação da imagem fonte  $I$ . As matrizes de coeficientes  $I_{LH}$ ,  $I_{HL}$  e  $I_{HH}$  representam subimagens detalhadas que contém a informação direcional (horizontal, vertical e diagonal) da imagem fonte  $I$ , devido à orientação espacial, como se pode observar na figura 19.

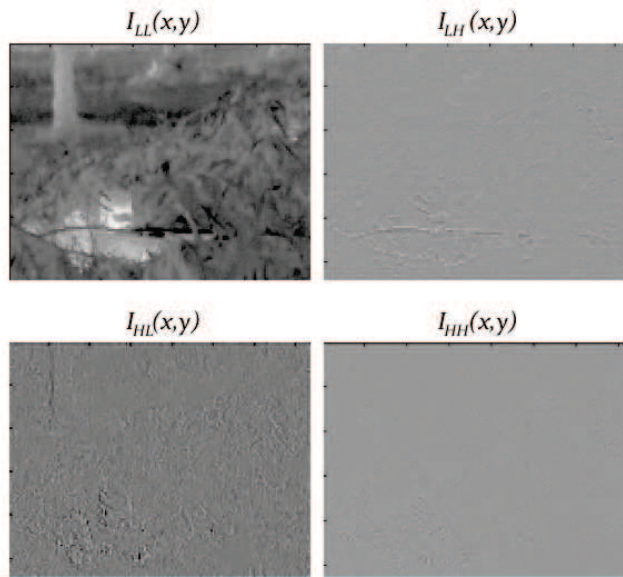


Figura 19 – a) coeficiente de detalhe b) coeficiente de detalhe horizontal c) coeficiente de detalhe vertical d) coeficiente de detalhe diagonal

A multi-resolução pode ser alcançada ao aplicar recursivamente o mesmo algoritmo nos coeficientes passa-baixo da decomposição anterior.

### 3.4.3 Reconstrução da Imagem

A transformada wavelet inversa é utilizada para reconstruir a imagem  $I$ , a partir das subimagens  $I_{LL}$ ,  $I_{LH}$ ,  $I_{HL}$  e  $I_{HH}$ .

Este processo envolve a sobreamostragem de coluna (inserindo zeros entre as amostras) e filtragem, utilizando um filtro passa baixo  $\tilde{L}$  e um filtro passa alto  $\tilde{H}$  para cada subimagem. A sobreamostragem das linhas e a filtragem com o filtro passa baixo  $\tilde{L}$  e o filtro passa alto  $\tilde{H}$  da imagem resultante e a soma de todas as matrizes reconstruirá a imagem  $I$ , tal como se vê na figura 20.

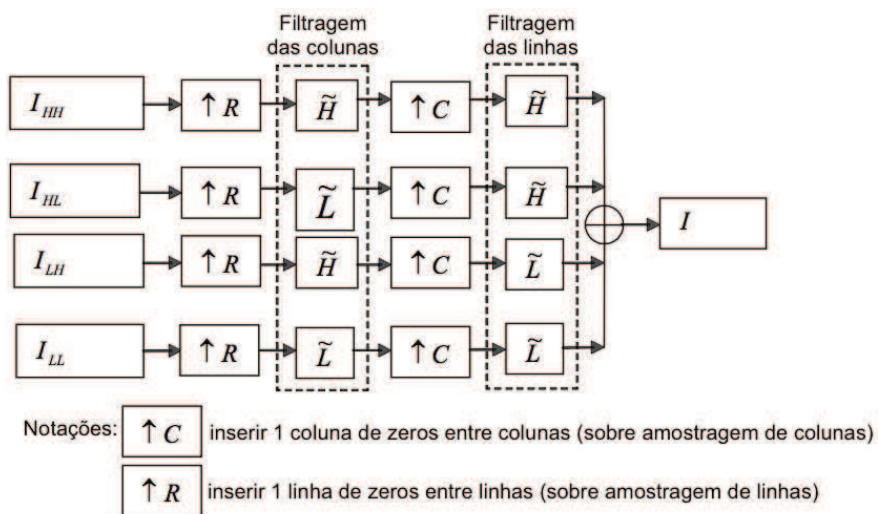


Figura 20 – Reconstrução da imagem usando wavelets (adaptada de [9])

### 3.4.4 Fusão de Imagem

No esquema de fusão de imagem usando wavelets, as imagens fonte  $I_{vis}$  e  $I_{term}$  são decompostas em coeficientes de aproximação e detalhe num determinado nível usando a DWT. Depois, os coeficientes de aproximação e de detalhe são combinados usando a regra de fusão  $\phi$ . A imagem fundida  $I_{fund}$  pode ser obtida ao utilizar a DWT Inversa (*Inverse Discrete Wavelet Transform – IDWT*) tal como:

$$I_{fund} = IDWT [\phi\{DWT(I_{vis}), DWT(I_{term})\}] \quad (8)$$

As regras de fusão utilizadas neste estudo foram: (i) o máximo dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada; (ii) e a média dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada.

Sendo o coeficiente de aproximação aquele que mais influencia a imagem final, nas regras de fusão considerou-se a utilização do máximo e da média por serem aquelas que forneceriam melhores resultados (resultados mais claros). Quanto aos coeficientes de detalhe, optou-se por utilizar como regra de fusão o coeficiente de detalhe com maior valor absoluto das duas imagens a fundir, por ser o que contribuirá com maior informação para a imagem fundida.

A figura 21 ilustra o esquema de fusão usando wavelets.

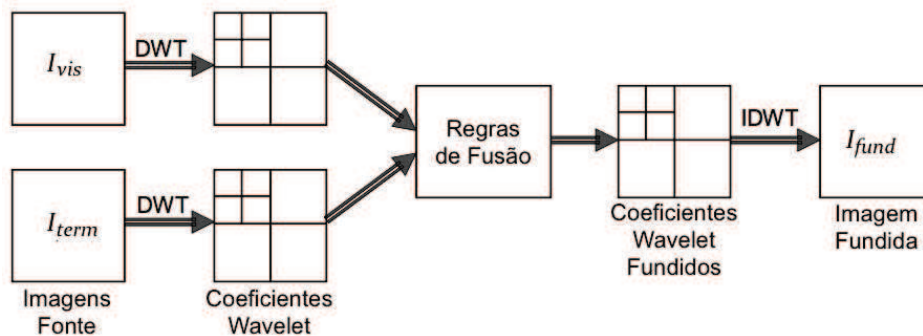


Figura 21 – Esquema de fusão usando wavelets (adaptada de [9])

### 3.5 Métricas de Avaliação de Desempenho

Para medir a qualidade da imagem fundida, é importante fazer uma avaliação quantitativa de tal forma que as comparações de desempenho dos respectivos algoritmos de fusão possam ser analisadas de forma objetiva. As métricas de qualidade usadas para avaliar o desempenho da fusão de imagem são normalmente classificadas em avaliação com referência completa e em avaliação sem referência [7]. As abordagens de referência completa comparam a imagem fundida com uma imagem de referência da mesma cena sem distorção, como sendo produzida por um esquema de fusão perfeito.

Um processo de fusão de imagem ideal deveria preservar todos os padrões úteis das imagens fonte e minimizar artefactos que possam interferir com análises subsequentes ou distrair os observadores [8].

Os algoritmos de fusão usados neste estudo estão destinados a aplicações na área militar e, devido à natureza de tempo-real da cena a ser visualizada, não há imagens de referência disponíveis. Assim, para avaliar o desempenho destes esquemas de fusão é necessário o uso de métricas de qualidade sem referência.

Os métricas utilizados para medir a eficiência dos métodos de fusão são: o desvio padrão, a entropia, a frequência espacial, a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural [7], [9].

### 3.5.1 Desvio Padrão

O desvio padrão mede o contraste da imagem fundida. É constituído por sinal e ruído, sendo que seria mais eficiente na ausência de ruído. É definido por:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_i - \bar{I}_i)^2} \quad (9)$$

onde  $I_i$  é um vetor coluna de  $N$  observações e  $\bar{I}_i$  é a média desse mesmo vetor. Uma imagem com um elevado contraste terá um elevado valor de desvio padrão [9].

### 3.5.2 Entropia

A entropia é uma medida de informação e o seu conceito tem sido empregue em vários campos científicos, inclusive em processamento de imagem. Pode medir o conteúdo de informação de uma imagem, mas tem uma limitação: não consegue distinguir a informação do ruído. A entropia de uma imagem é dada por [27]:

$$E = - \sum_{i=1}^G p(i) \log_2 p(i) \quad (10)$$

onde  $G$  é o número de níveis de cinzento do histograma da imagem, tipicamente entre 0 e 255, e  $p(i)$  é a frequência normalizada de ocorrência de cada nível de cinzento.

A entropia pode refletir diretamente a média do conteúdo de informação de uma imagem [27]. Uma imagem com elevado conteúdo de informação terá uma entropia elevada.

### 3.5.3 Frequência Espacial

Esta frequência no domínio espacial indica o nível de atividade global<sup>1</sup> de uma imagem, e é definida como [28]:

$$FE = \sqrt{FL^2 + FC^2} \quad (11)$$

onde  $FL$  corresponde à frequência das linhas e  $FC$  à frequência das colunas. As frequências das linhas e das colunas são dadas por:

$$FL = \sqrt{\frac{1}{n_l n_c} \sum_{i=1}^{n_l} \sum_{j=2}^{n_c} [I_{fund}(i, j) - I_{fund}(i, j - 1)]^2} \quad (12)$$

$$FC = \sqrt{\frac{1}{n_l n_c} \sum_{j=1}^{n_c} \sum_{i=2}^{n_l} [I_{fund}(i, j) - I_{fund}(i - 1, j)]^2} \quad (13)$$

onde  $n_l$  é o número de linhas e  $n_c$  é o número de colunas de uma imagem.

Um valor elevado para a frequência espacial indica uma atividade global elevada, ou seja, quanto maior for o seu valor mais informação possui a imagem fundida.

### 3.5.4 Informação Mútua

Esta métrica mede o grau de dependência entre duas imagens. É calculada ao definir o histograma conjunto das imagens fonte  $I_{vis}$ ,  $I_{term}$  e a imagem fundida  $I_{fund}$  como  $p(fund, vis)$  e  $p(fund, term)$  [7]. A informação mútua entre a imagem fonte e a imagem fundida é dada por:

$$IM_1(fund, vis) = - \sum p(fund, vis) \log_2 \left( \frac{p(fund, vis)}{p(fund) \cdot p(vis)} \right) \quad (14)$$

$$IM_2(fund, term) = - \sum p(fund, term) \log_2 \left( \frac{p(fund, term)}{p(fund) \cdot p(term)} \right) \quad (15)$$

onde  $p(fund, vis)$  e  $p(fund, term)$  são os histogramas conjuntos das imagens de fonte  $I_{vis}$ ,  $I_{term}$  e a imagem fundida  $I_{fund}$ . A eficiência do algoritmo de fusão é determinada pela métrica  $IM$  que é definida por:

$$IM = IM_1(fund, vis) + IM_2(fund, term) \quad (16)$$

---

<sup>1</sup> Traduzido de "overall activity"

Uma dependência maior, ou seja, um valor maior, implica melhor qualidade.

### 3.5.5 Índice de Qualidade de Fusão

O índice de qualidade de imagem  $Q_W$  inicialmente foi introduzido por Wang e Bovik. Dadas duas imagens  $I_{vis}$  e  $I_{term}$ , seja  $\overline{I_{vis}}$  a média de  $I_{vis}$ , e  $\sigma_{I_{vis}}^2$  e  $\sigma_{I_{vis}I_{term}}$  a variância de  $I_{vis}$  e covariância de  $I_{vis}$  e  $I_{term}$ , respetivamente [29], dadas por:

$$\sigma_{I_{vis}}^2 = \frac{1}{n-1} \sum_{i=1}^n (I_{vis_i} - \overline{I_{vis}})^2 \quad (17)$$

$$\sigma_{I_{vis}I_{term}} = \frac{1}{n-1} \sum_{i=1}^n (I_{vis_i} - \overline{I_{vis}})(I_{term_i} - \overline{I_{term}}) \quad (18)$$

O índice global de qualidade de duas imagens é definido por:

$$Q_0(I_{vis}, I_{term}) = \frac{4\sigma_{I_{vis}I_{term}}\overline{I_{vis}}\overline{I_{term}}}{(\overline{I_{vis}}^2 + \overline{I_{term}}^2)(\sigma_{I_{vis}}^2 + \sigma_{I_{term}}^2)} \quad (19)$$

que pode ser decomposto em:

$$Q_0(I_{vis}, I_{term}) = \frac{\sigma_{I_{vis}I_{term}}}{\sigma_{I_{vis}}\sigma_{I_{term}}} \cdot \frac{2\overline{I_{vis}}\overline{I_{term}}}{(\overline{I_{vis}}^2 + \overline{I_{term}}^2)} \cdot \frac{2\sigma_{I_{vis}}\sigma_{I_{term}}}{(\sigma_{I_{vis}}^2 + \sigma_{I_{term}}^2)} \quad (20)$$

A primeira componente corresponde ao coeficiente de correlação entre  $I_{vis}$  e  $I_{term}$ . Este valor é a medida de semelhança das imagens  $I_{vis}$  e  $I_{term}$ , e toma valores entre 0 e 1. A segunda componente corresponde à distorção da luminância que tem uma gama dinâmica de [0,1]. O terceiro fator mede a distorção do contraste e seu alcance também é de [0,1]. Concluindo, o valor de  $Q_0(I_{vis}, I_{term})$  está entre [0,1], e o valor máximo  $Q_0 = 1$  é alcançado quando  $I_{vis}$  e  $I_{term}$  são idênticas.

Piella e Heijmans introduziram um procedimento de pesos no cálculo de  $Q_0$  [30]. O peso reflete a relevância local de uma imagem fonte que pode depender da sua variância local, contraste, nitidez ou entropia. Dada a saliência local de duas imagens fonte  $I_{vis}$  e  $I_{term}$ , calcula um peso local  $\lambda$  que indica a importância relativa da imagem  $I_{vis}$  comparada com a imagem  $I_{term}$ . Quão maior for  $\lambda$ , maior é o peso dado à imagem  $I_{vis}$ . Uma escolha típica para  $\lambda$  é:

$$\lambda = \frac{S(I_{vis})}{[S(I_{vis}) + S(I_{term})]} \quad (21)$$

onde  $S(I_{vis})$  e  $S(I_{term})$  denotam a saliência da imagem  $I_{vis}$  e  $I_{term}$ , respectivamente. Depois, o índice de qualidade da imagem com os pesos pode ser definido por:

$$Q_W = \lambda Q_0(I_{vis}, I_{fund}) + (1 - \lambda) Q_0(I_{term}, I_{fund}) \quad (22)$$

Dado que os sinais de imagens são geralmente não-estacionários, é mais apropriado medir o índice de qualidade de imagem  $Q_W$  em regiões locais e depois combinar estes resultados locais numa única medida como a medida global da imagem inteira.

Piella também sugeriu usar a variância local como a saliência de uma imagem. De facto, a métrica do índice de qualidade de uma imagem mede a semelhança entre a imagem fundida  $I_{fund}$  e ambas as imagens fonte  $I_{vis}$  e  $I_{term}$ , ao assumir que a imagem fundida ideal deve apresentar ambas as imagens de entrada originais [8]. Esta métrica pode assumir valores entre 0 e 1, onde o valor 1 corresponde a uma melhor qualidade da imagem fundida.

### 3.5.6 Índice de Semelhança Estrutural

O índice de semelhança estrutural (ISE) tem sido utilizado para indicar a semelhança da estrutura da informação presente nas duas imagens a serem comparadas. O ISE de duas imagens  $I_{vis}$  e  $I_{term}$ , é definido por [31]:

$$ISE(I_{vis}, I_{term}) = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{I_{vis}I_{term}} + C_2)}{(\mu_{I_{vis}}^2 + \mu_{I_{term}}^2 + C_1)(\sigma_{I_{vis}}^2 + \sigma_{I_{term}}^2 + C_2)} \quad (23)$$

onde  $\mu$  é a média da intensidade da imagem,  $\sigma$  é o desvio padrão da imagem e  $C_1$  e  $C_2$  são constantes.

No entanto, o ISE é uma abordagem de referência-completa e requer uma imagem de referência para o seu cálculo. Este problema é ultrapassado ao calcular separadamente a informação presente de cada imagem fonte na imagem fundida,  $ISE(fund, vis)$  e  $ISE(fund, term)$ .

Assim, o índice de semelhança estrutural total, para a imagem fundida, é calculado por:

$$ISE_T = ISE(fund, vis) + ISE(fund, term) \quad (24)$$

Um valor maior, indica que a informação presente em cada uma das imagens fonte se encontra presente na imagem fundida.

## 4 Resultados e Discussão

Neste capítulo, serão apresentados os resultados relativos aos métodos utilizados, seguidos de uma análise qualitativa e de uma análise quantitativa dos mesmos. Por último, serão extraídas conclusões face aos resultados obtidos.

### 4.1 Conjunto de Dados

As imagens utilizadas para o presente estudo foram adquiridas com uma câmara fotográfica FLIR e têm dimensão de 320×240 pixéis. Cada imagem adquirida é constituída por uma imagem da banda do visível e uma imagem térmica. No total foram obtidas 182 imagens.

As imagens foram obtidas presencialmente em dois exercícios de campo da Academia Militar (AM). O primeiro exercício foi realizado no início de fevereiro, em Vendas Novas (Alentejo) e o segundo exercício realizou-se no fim de fevereiro, em Santa Margarida (Ribatejo). As pessoas presentes nas imagens, camufladas no meio envolvente, são alunos da AM que se disponibilizaram para a aquisição de imagens.

Para os testes efetuados, foram seleccionadas 20 imagens (constituídas por uma imagem visível e uma imagem térmica), que correspondem a pessoas camufladas em diferentes meios circundantes e em diferentes condições de luminosidade (Anexo 1).

Os resultados apresentados nas seguintes secções foram obtidos recorrendo a um computador com processador Intel Core i5 @ 2,5GHz, 8 GB de memória *Random Access Memory* (RAM) e sistema operativo Mac OS X 10.10.4. Para a implementação de todos os algoritmos desenvolvidos utilizou-se o software Matlab® R 8.4.0.150421 (R2014b).

### 4.2 Análise em Componentes Principais

Em relação ao método da PCA, foram implementadas três variantes do mesmo. A primeira variante consiste em efetuar o cálculo das componentes principais a partir das imagens fonte e depois multiplicar os valores obtidos pelas imagens fonte. A segunda variante difere da primeira no sentido em que são utilizadas imagens equalizadas, obtidas a partir das imagens fonte. A terceira e última variante é uma versão híbrida, ou seja, as componentes principais são obtidas através das imagens equalizadas, no entanto esses valores vão ser multiplicados pelas imagens fonte.

Ao utilizar a primeira variante do método da PCA, os resultados obtidos foram, na sua maioria, satisfatórios. Na figura abaixo (figura 22) pode-se observar um exemplo de duas imagens fonte e a imagem fundida obtida com este método. É perceptível que a imagem fundida para além de ter o detalhe que se encontra na imagem visível, possui também o pormenor da imagem térmica. Enquanto na imagem visível a presença de uma pessoa no meio das folhas passa despercebida, na imagem fundida

consegue-se observar a presença dessa pessoa camuflada entre os ramos, que é o pretendido. Por esse motivo pode-se considerar que os resultados obtidos são razoáveis.



Figura 22 – a) imagem visível b) imagem térmica c) imagem fundida com 1ª variante PCA

Em vários exemplos, as imagens fonte têm naturezas muito distintas e o seu histograma é bastante diferente. Para melhorar estes os casos onde a fusão não ocorre de forma satisfatória, dado que uma das imagens fonte tem muito mais influência para a imagem fundida do que a outra, testou-se a utilização de uma equalização de histograma para que a distribuição do histograma nas duas imagens fonte fosse similar, constituindo esta a segunda variante da PCA testada.

Desta forma, foi feita uma equalização do histograma das imagens fonte em 256 níveis. Enquanto na figura 22 se observa as imagens fonte sem equalização, na figura 23 observa-se as imagens fonte após a equalização e a respetiva imagem fundida.



Figura 23 – a) imagem visível equalizada b) imagem térmica equalizada c) imagem fundida com 2ª variante PCA

Apesar dos histogramas estarem equalizados, o que teoricamente corresponderia a uma melhoria da qualidade da imagem fundida, neste caso isso não se verifica (figura 23). Tal como se pode observar, a qualidade subjetiva da imagem fundida não é boa. Para além do detalhe da imagem visível estar degradado, a componente térmica é praticamente inexistente, o que não contribui para a deteção da pessoa camuflada entre as folhas.

Depois de ser testada a fusão de outras imagens fonte, os resultados obtidos, na maioria dos casos, não foram os esperados. Dado que as imagens térmicas são imagens onde as componentes quentes ou térmicas são bastante distintas da restante informação, ao ser feita a equalização de histograma a parte branca da imagem (a parte mais quente na imagem), vai ser redistribuída numa escala de cinzentos, o que fará com que haja uma perda de detalhe da imagem térmica, como se pode observar na figura 24.

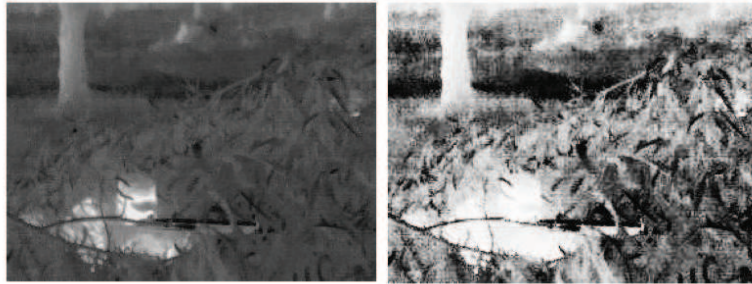


Figura 24 – a) imagem térmica inicial b) imagem térmica equalizada

Como os resultados obtidos com a segunda variante da PCA não foram os esperados, testou-se a terceira variante da PCA. Nesta variante, os valores das componentes principais foram calculados a partir das imagens equalizadas, no entanto, a fusão de imagem fez-se ao multiplicar os valores das componentes principais pelas imagens fonte não equalizadas.

Desta forma, obteve-se o resultado apresentado na figura 25. Aqui, pode-se observar que a imagem fundida, para além de possuir os detalhes presentes a imagem visível, também tem presente a componente térmica, que permite a identificação da pessoa escondida entre os ramos.



Figura 25 - a) imagem visível b) imagem térmica c) imagem fundida com 3ª variante PCA

Por inspeção visual, torna-se evidente que os piores resultados são obtidos quando se utiliza a segunda variante da PCA. Por outro lado, os resultados adquiridos ao utilizar a primeira variante e a terceira são bastante semelhantes.

Teoricamente, os melhores resultados são encontrados quando se calcula as componentes principais a partir das imagens equalizadas, na medida em que ambas as imagens terão a mesma quantidade de informação em cada nível, e se multiplicam esses valores pelas imagens fonte originais.

Na figura 26, mostra-se um exemplo dos resultados obtidos para as três variantes de PCA utilizadas, em que se adquire resultados razoáveis para qualquer uma delas.



Figura 26 – a) imagem fundida com 1ª variante PCA b) imagem fundida com 2ª variante PCA c) imagem fundida com 3ª variante PCA

Os resultados adquiridos para este exemplo são satisfatórios no seu todo porque em qualquer uma das três imagens é possível identificar a pessoa escondida entre os arbustos. Neste caso, a segunda variante da PCA obteve bons resultados porque a imagem térmica ao ser equalizada não perdeu o contraste existente entre as partes quentes e frias da imagem, permitindo desta forma que a componente quente continue identificável.

### 4.3 Pirâmide Laplaciana

A pirâmide escolhida para este estudo foi a pirâmide laplaciana, por ser aquela que apresenta melhores resultados em relação a outros tipos de pirâmides [11].

A fusão de imagem que utiliza este método processa-se da seguinte forma: as imagens de fonte são decompostas em  $N$  níveis e, depois de construídas as pirâmides com os respectivos  $N$  níveis, a imagem fundida vai ser reconstruída fazendo uso das regras de fusão já descritas. A fusão é feita a cada nível, a partir do topo até à base da pirâmide, sendo desta forma reconstruída a imagem fundida.

Para comparar os resultados, a decomposição das imagens de fonte foi efetuada a dois a quatro níveis, ou seja, foram construídas pirâmides com dois, três e quatro níveis paralelamente. Dado que as imagens fonte têm uma dimensão reduzida, de  $320 \times 240$  pixéis, não é possível decompor as imagens além dos quatro níveis. Na decomposição com um nível obtém-se uma imagem de  $160 \times 120$  pixéis, com dois níveis uma imagem de  $80 \times 60$  pixéis, com três níveis uma imagem de  $40 \times 30$  pixéis e com quatro níveis uma imagem de  $20 \times 15$  pixéis. A partir deste nível a decomposição da imagem deixa de ser conveniente, porque o número de pixéis deixa de ser inteiro.

Na figura 27 podem-se observar três imagens fundidas, onde a imagem da esquerda corresponde à imagem obtida com dois níveis, a do meio com três níveis e a da direita com quatro níveis de decomposição.



Figura 27 a) imagem fundida com Pirâmide Laplaciana de 2 níveis b) imagem fundida com pirâmide laplaciana de 3 níveis c) imagem fundida com pirâmide laplaciana de 4 níveis

Ao observar as três imagens, constata-se que a principal diferença se encontra entre a primeira e a terceira. Enquanto a primeira imagem (obtida com dois níveis de decomposição) apresenta pouco detalhe e um contraste muito baixo, a imagem obtida a partir da decomposição em quatro níveis apresenta mais detalhe e contraste. Embora nas três imagens se consiga identificar a pessoa camuflada, é mais evidente nas imagens obtidas com três e quatro níveis de decomposição. Por outro lado, também é perceptível que a terceira imagem tem os detalhes mais afinados, ou seja, as folhas e ramos são mais visíveis na última imagem.

Dado que o objetivo principal é obter uma imagem fundida que apresente a maior quantidade de informação possível, torna-se evidente que a pirâmide laplaciana com quatro níveis de decomposição é a mais adequada para o objetivo pretendido.

#### 4.4 Wavelets

Quando se utiliza as wavelets para decompor uma imagem, esta é decomposta em quatro coeficientes que podem ser chamados de subimagens. Estes coeficientes são chamados de coeficientes de aproximação e de detalhe. O coeficiente de aproximação é aquele que contém mais informação relativa à imagem fonte e por esse motivo é considerado uma aproximação da imagem fonte. Os coeficientes de detalhe contêm informação de detalhe horizontal, vertical e diagonal da imagem fonte.

Quando se efetua mais do que um nível de decomposição das imagens fonte, será o coeficiente de aproximação obtido na primeira decomposição que irá dar origem aos novos quatro coeficientes, e assim sucessivamente, como se pode observar na figura 28.

$I_{LL}(x,y)$	$I_{LH}(x,y)$	$I_{LL}(x,y)$	$I_{LH}(x,y)$	$I_{LH}(x,y)$
		$I_{HL}(x,y)$	$I_{HH}(x,y)$	
$I_{HL}(x,y)$	$I_{HH}(x,y)$	$I_{HL}(x,y)$		$I_{HH}(x,y)$

Figura 28 – a) coeficientes com 1 nível de decomposição b) coeficientes com 2 níveis de decomposição

Após a decomposição das imagens fonte em coeficientes de aproximação e de detalhe, vão ser utilizadas as regras de fusão que foram definidas no capítulo métodos, nomeadamente: (i) o máximo dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada; (ii) e a média dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada.

Estas regras de fusão foram definidas após terem sido efetuados vários testes. O coeficiente de aproximação é aquele que mais influencia a imagem fundida, sendo por isso o coeficiente a que se deve dar maior relevância. Ou seja, a escolha da regra de fusão aplicada aos coeficientes de aproximação é a que mais influencia o resultado final. Assim, para discernir qual a regra de fusão a aplicar, foram testadas as três formas de fusão que se podem aplicar aos coeficientes: o máximo, a média e o mínimo dos coeficientes de aproximação à maior escala.

A figura 29 mostra a influencia da escolha do máximo, da média e do mínimo na fusão dos coeficientes de aproximação, na imagem fundida.



Figura 29 – a) imagem fundida com 'máximo' b) imagem fundida com 'média' c) imagem fundida com 'mínimo'

Torna-se perceptível que a utilização do mínimo na fusão do coeficiente de aproximação vai dar origem a uma imagem fundida sem qualquer detalhe térmico, o que não é o desejado. A utilização do máximo é a que obtém resultados mais claros, seguida da utilização da média, sendo, portanto, a regra de fusão aplicada ao coeficiente de aproximação.

Os coeficientes de detalhe têm pouca influência na imagem fundida e por esse motivo optou-se por escolher aquele que apresenta o maior valor absoluto, pois desta forma contribuirá com mais informação para a imagem fundida.

Quanto à wavelet utilizada para a decomposição e reconstrução da imagem, foi utilizada a Haar Wavelet, pois não se verificaram diferenças substanciais com a utilização de outras wavelets.

Relativamente aos níveis de decomposição foram testados apenas um nível e dois níveis de decomposição. A figura 30 mostra a imagem fundida com um nível e dois níveis de decomposição, utilizando a primeira regra de fusão, e a figura 31 mostra a imagem fundida com um nível e dois níveis de decomposição, utilizando a segunda regra de fusão.

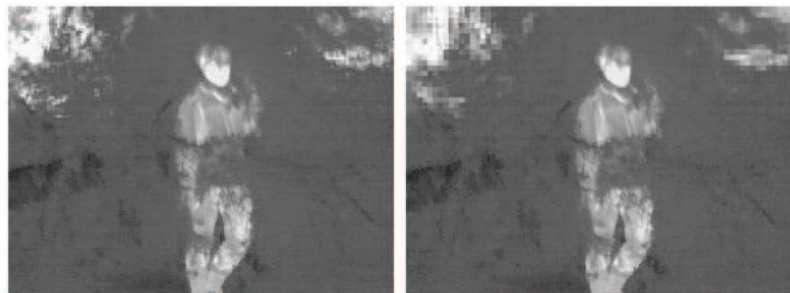


Figura 30 – a) imagem fundida com 1ª regra de fusão com 1 nível de decomposição b) imagem fundida com 1ª regra de fusão com 2 níveis de decomposição

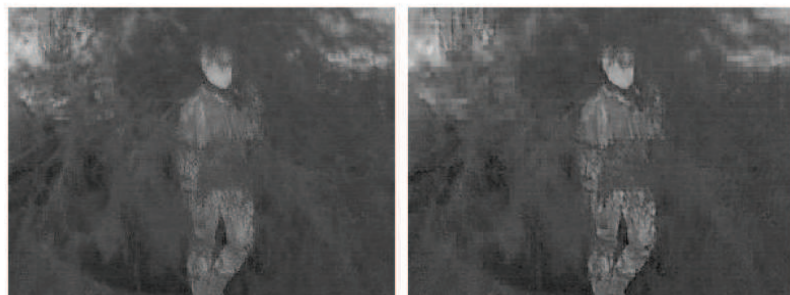


Figura 31 – a) imagem fundida com 1ª regra de fusão com 1 nível de decomposição b) imagem fundida com 1ª regra de fusão com 2 níveis de decomposição

Por inspeção visual pode-se constatar que as imagens obtidas com dois níveis de decomposição estão mais degradadas, ou seja, têm pior qualidade. Devido a esse facto, não se considerou pertinente fazer a fusão da imagem com mais níveis de decomposição, pois a qualidade iria ser pior certamente.

Quanto às regras de fusão utilizadas, consegue-se perceber que ao utilizar a primeira regra de fusão (o máximo dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada) a imagem fundida tem a componente térmica mais marcada, sendo mais fácil a identificação da pessoa camuflada. Com a segunda regra de fusão (e a média dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada), apesar da pessoa continuar identificável, o contraste da imagem é menor.

Como o objetivo é identificar a pessoa camuflada, para a análise qualitativa vai-se optar por considerar as imagens obtidas com a primeira regra de fusão.

#### 4.5 Análise Qualitativa

De forma a poder comparar os três métodos implementados, apresentam-se dez conjuntos de três imagens fundidas, da figura 32 à figura 41, onde a primeira imagem corresponde ao método que utiliza a terceira variante da PCA (por ter sido considerado o mais adequado), a segunda imagem diz respeito ao método da pirâmide laplaciana com quatro níveis de decomposição e a terceira imagem corresponde ao método que utiliza as wavelets com um nível de decomposição usando como regra de fusão o máximo dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada. Estes três métodos foram selecionados como sendo os que produziram melhores resultados.

Na figura 32, pode-se observar que a qualidade das três imagens é diferente. A imagem obtida com a pirâmide laplaciana aparenta ter o maior contraste das três, seguida da imagem obtida com a PCA e por último a imagem obtida com as wavelets. A última imagem é a que apresenta menos contraste, e onde os detalhes são menos perceptíveis.



Figura 32 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Neste segundo exemplo, figura 33, a qualidade subjetiva das três imagens é muito semelhante, sendo que a segunda imagem apresenta um contraste menor, visível na silhueta das pessoas na sombra da árvore, relativamente às outras duas.



Figura 33 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis  
c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Na figura 34, a qualidade das duas primeiras imagens é semelhante, sendo que na terceira imagem a pessoa encontra-se mais destacada em relação às outras duas, sendo por esse motivo preferível. O detalhe presente nas três imagens é idêntico.



Figura 34 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis  
c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

No quarto conjunto de imagens, figura 35, a imagem que apresenta um maior detalhe e contraste é a terceira, onde é mais evidente a componente térmica, ou seja, a zona a que corresponde a uma pessoa ocorre mais destacada (mais clara) quando comparada com as restantes regiões na imagem. Neste exemplo, tanto na primeira como na segunda imagem a componente térmica é menos visível, apesar de na segunda imagem ser mais perceptível.



Figura 35 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis  
c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Tal como acontece na figura 35, na figura 36 a imagem que apresenta melhor qualidade é a terceira, onde se utiliza o método que utiliza as wavelets. Neste caso, apesar de na primeira imagem a componente térmica estar mais destacada do que na segunda imagem, os detalhes e o contraste presente é menor relativamente à segunda imagem, aquela que utiliza a pirâmide laplaciana. Já na terceira imagem, a zona que corresponde à pessoa camuflada encontra-se mais evidente sendo por esse motivo a preferível. Também na terceira imagem, ao contrário do que acontece nas primeiras duas, a área superior da imagem encontra-se mais clara, o que corresponde à incidência de raios solares.



Figura 36 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Na figura 37, em qualquer uma das três imagens não existe muito detalhe, devido à natureza das imagens fonte (ver imagem O, Anexo 1). Nesta situação, encontra-se uma pessoa atrás de arbustos, o que por si só contribui para a falta de detalhe, pois os arbustos são esparsos e nas imagens fonte aparentam ser apenas uma mancha. Na terceira imagem é onde se pode observar maior intensidade da componente térmica, no entanto a segunda imagem é aquela que aparenta ter uma melhor qualidade, porque tem um equilíbrio entre o detalhe e a componente térmica; apesar de se conseguir identificar a pessoa também se consegue perceber que as partes mais escuras da imagem têm tonalidades diferentes (ao contrário do que acontece na terceira imagem).



Figura 37 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Tal como ocorre na figura 37, a figura 38 é pouco detalhada pelos mesmos motivos (a pessoa encontra-se atrás de arbustos, que são esparsos e aparentam ser apenas uma mancha nas imagens fonte, o que contribui para a falta de detalhe), e tal como acontece com a figura 36, a segunda imagem que é obtida com a pirâmide laplaciana é a que parece obter um compromisso entre o detalhe dos

arbustos e a componente térmica da pessoa camuflada. Na segunda imagem, não se consegue só identificar a pessoa, como também ter uma perspectiva do seu ambiente envolvente.

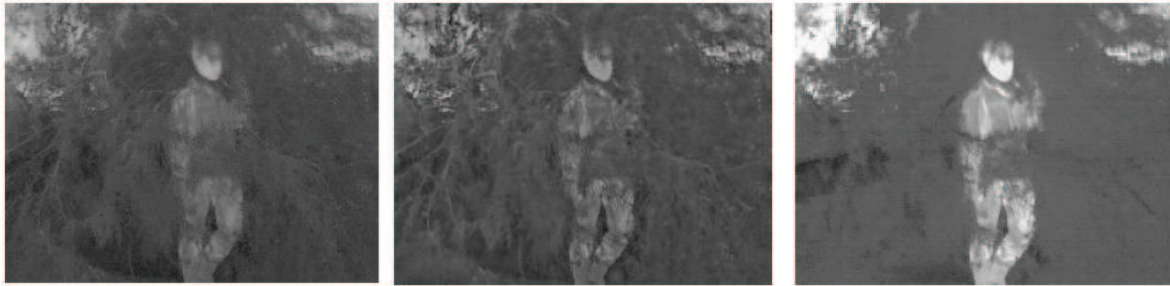


Figura 38 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Dos 10 exemplos apresentados, a figura 39 é a que apresenta os piores resultados. Com qualquer um dos métodos utilizados não se obtém a imagem ideal, pois não se consegue identificar a pessoa camuflada. Nesta imagem encontra-se uma pessoa sentada atrás de arbustos, o que não é perceptível ao observar a imagem fundida. Contudo, a imagem que aparenta ter a componente térmica mais destacada é a terceira (obtida com as wavelets), sendo por isso a mais indicada para o objetivo pretendido (identificar a pessoa camuflada).



Figura 39 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Na figura 40, a qualidade subjetiva das três imagens é idêntica, sendo que a terceira imagem mostra a componente térmica de uma forma mais clara, sendo por isso a melhor.



Figura 40 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

No décimo e último conjunto de imagens, figura 41, mais uma vez a qualidade subjetiva é semelhante nas três imagens, sendo que a primeira apresenta menos contraste relativamente às outras duas, e a terceira é a que possui a componente térmica mais destacada, sendo visível a face e os braços, pois estas regiões (da pessoa indicada na figura) são as que se encontram mais quentes.



Figura 41 – a) imagem fundida com 3ª variante PCA b) imagem fundida com pirâmide laplaciana com 4 níveis c) imagem fundida com wavelets com 1 nível de decomposição, com a 1ª regra de fusão

Resumindo a análise qualitativa a partir destes 10 exemplos, o método que aparenta ter melhores resultados é aquele que utiliza wavelets, seguido do método que utiliza a pirâmide laplaciana e por último o que utiliza a PCA.

Um dos fatores que contribuí para os fracos resultados apresentados pelo método que utiliza a PCA é o facto de quando ambos os valores das componentes principais são próximos de 0.5, a fusão assemelha-se à fusão através da média simples o que produz um contraste de características baixo.

Por outro lado, o método que utiliza as wavelets é o que apresenta melhores resultados, pois é aquele que faz a fusão das várias componentes da imagem (aproximação, horizontal, vertical e diagonal) de acordo com as regras de fusão definidas como sendo as melhores para o objetivo pretendido, que é detetar as pessoas camufladas.

Ao observar as imagens fundidas que foram geradas, apesar de se conseguir ter uma percepção visual de qual o melhor e pior método utilizado, para medir a qualidade das imagens é necessário fazer uma avaliação quantitativa que permita a comparação objetiva e automática do desempenho dos métodos utilizados.

## 4.6 Análise Quantitativa

Nesta secção é efetuada uma avaliação quantitativa dos métodos utilizados na fusão de imagem para uma comparação objetiva e fundamentada do desempenho de cada método.

Foram utilizadas 20 imagens (ver Anexo 1) fundidas para testar cada uma das métricas de avaliação de desempenho e os resultados serão apresentados em figuras *boxplot*.

Cada caixa (retângulo) constitui um intervalo de dispersão de valores e tem uma linha vermelha que representa a sua mediana, acima dessa linha estão os valores acima da mediana e abaixo os valores menores do que a mediana. A linha superior (fora da caixa) representa o valor máximo, e a linha inferior (fora da caixa) representa o valor mínimo. Por vezes podem aparecer cruces vermelhas,

chamadas de *outliers*, que representam valores que são bastante diferentes do resto do conjunto de dados e por isso estão fora do intervalo de dispersão (ver figura 42).

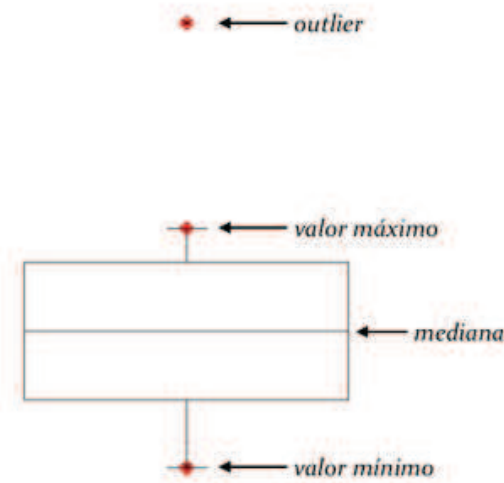


Figura 42 - Boxplot

Nestas figuras, é feita uma comparação objetiva entre os métodos testados, em que cada caixa corresponde a um método. Os métodos são designados da seguinte forma: PCA1, PCA2 e PCA3 correspondem às três variantes da PCA testadas; LP02 e LP04 correspondem ao método que utiliza a pirâmide laplaciana com dois e quatro níveis, respetivamente; WV11 e WV12 correspondem à wavelet com um nível de decomposição com a 1ª e 2ª regra de fusão, respetivamente; por último, WV21 e WV22 equivalem à wavelet com dois níveis de decomposição com a 1ª e 2ª regra de fusão.

#### 4.6.1 Desvio Padrão

O desvio padrão mede o contraste da imagem fundida. As imagens obtidas a partir do método que utiliza a PCA têm valores próximos de zero, o que corresponde à avaliação que tinha sido feita por inspeção visual (ver figura 43), ou seja, têm muito pouco contraste. Esta falta de contraste corresponde à falta de detalhe na imagem fundida.

Os métodos que utilizam a pirâmide laplaciana e wavelets apresentam valores muito mais elevados, sendo que aquele que mostra melhores resultados (mediana mais elevada), quanto a esta métrica, é o que utiliza wavelets com a 1ª regra de fusão. Este método é o que possui os valores mais elevados por apresentar maior contraste nas imagens fundidas, devido à presença bem marcada da componente térmica (parte mais clara na imagem) e da componente visível (parte mais escura na imagem).

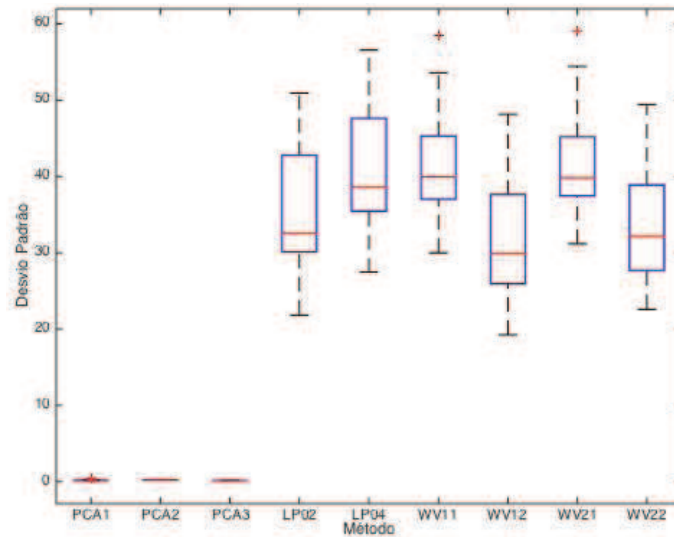


Figura 43 – Boxplot do Desvio Padrão

Embora por inspeção visual, as imagens obtidas com dois níveis de decomposição nas wavelets apresentem pior qualidade, aqui verifica-se que a wavelet com dois níveis de decomposição que utiliza a 1ª regra de fusão (WV21) tem resultados muito próximos daquela com apenas um nível de decomposição, tal como se pode ver na figura 44. Isto acontece devido à consideração de todas as bandas de frequência no processo de fusão [9].

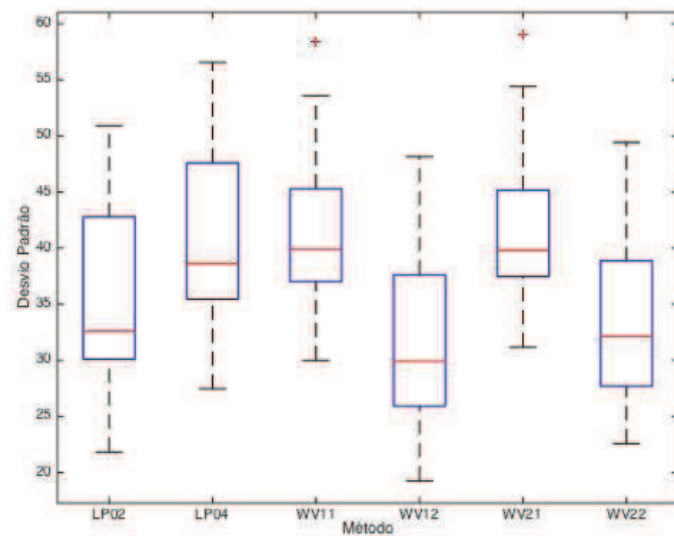


Figura 44 – Boxplot do Desvio Padrão

A pirâmide laplaciana com quatro níveis é aquela que, a seguir às wavelets, apresenta melhores resultados por ser capaz de preservar os contornos e reduzir os artefactos à sua volta. Este resultado está de acordo com a avaliação visual.

Os valores obtidos estão dentro da gama de valores adquiridos noutros estudos [9], em que os resultados do desvio padrão têm média de 47. No presente estudo, a média dos valores obtidos é aproximadamente 35 (se excluirmos os valores obtidos para a PCA), o que não difere muito dos resultados obtidos em [9].

#### 4.6.2 Entropia

A entropia mede o conteúdo de informação da imagem fundida, logo uma entropia elevada indica o melhoramento do conteúdo de informação [27]. Tal como acontece no desvio padrão, a entropia das imagens obtidas com o método que utiliza a PCA apresenta novamente valores próximos de zero, o que indica a inexistência de informação relevante.

Ao excluir os valores obtidos com o método que utiliza a PCA, observa-se que os métodos que obtêm melhores resultados são a pirâmide laplaciana com quatro níveis, seguida das wavelets com um e dois níveis de decomposição que utilizam a 1ª regra de fusão (ver figura 45). A pirâmide laplaciana com quatro níveis obtém os melhores resultados (o conjunto de valores obtidos possui a mediana mais elevada) porque preserva os contornos das imagens e contribui para um melhor detalhe da imagem fundida.

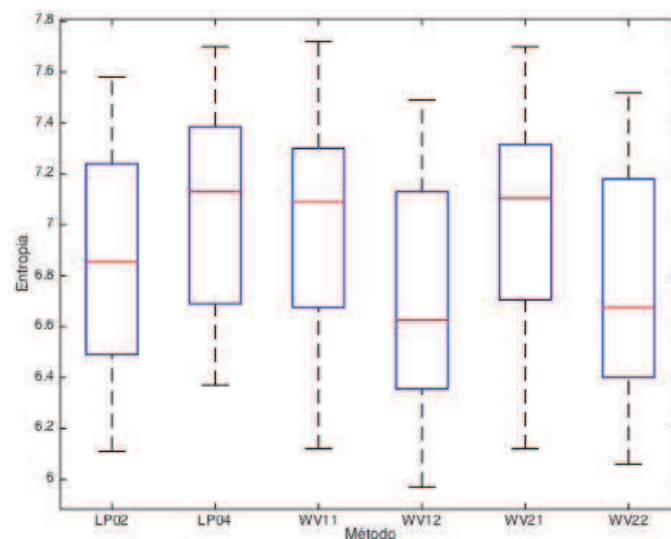


Figura 45 – Boxplot da Entropia

Tal como acontece na métrica anterior, a wavelet com dois níveis de decomposição apresenta resultados muito próximos (*mediana* ~ 7.1) à que utiliza apenas um nível de decomposição (*mediana* ~ 7.1). No entanto, como se observa na análise qualitativa, com dois níveis a imagem é degradada (ver secção 4.4, figuras 29 e 30). Ou seja, embora a entropia seja elevada, a fusão não está a aumentar o conteúdo de informação, mas está sim a deteriorar a imagem através da introdução de ruído [27].

É importante salientar que, a entropia é uma métrica apropriada para a comparação da qualidade das imagens fundidas, fornecendo por isso uma avaliação relativa entre os métodos implementados. No entanto, para a classificação da fusão de imagem propriamente dita, por si só não é uma métrica adequada, pois não reflete a relação entre as imagens fonte e a imagem fundida.

Os valores que foram obtidos para esta métrica (*média* ~ 6.9) estão próximos dos valores alcançados por outros estudos [7] [11] [27], em que a média dos valores obtidos é ~ 7.1.

### 4.6.3 Frequência Espacial

A frequência espacial indica o nível de atividade global<sup>2</sup> de uma imagem, ou seja, quanto maior for o seu valor mais informação possui a imagem, exceto nos casos onde se verifica a introdução de ruído no processo de fusão. Assim como acontece na entropia e no desvio padrão, a frequência espacial das imagens obtidas com o método que utiliza a PCA apresenta valores próximos de zero.

Pode-se observar que, o método que obtém melhores resultados é a pirâmide laplaciana com quatro níveis por preservar os contornos das imagens e contribuir dessa forma para um melhor detalhe da imagem fundida, seguido pelas wavelets com dois níveis de decomposição, que têm um desempenho melhor do que aquelas com um nível de decomposição, ao contrário do que se seria de esperar (ver figura 46).

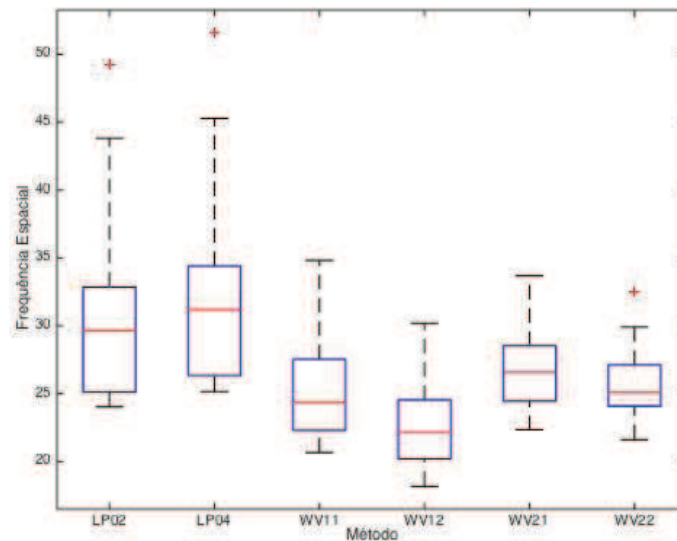


Figura 46 – Boxplot da Frequência Espacial

Ao comparar os métodos que utilizam as wavelets, pode-se deduzir que a distorção introduzida pelo processo de fusão é o motivo pelo qual as wavelets com dois níveis de decomposição alcançam melhores resultados, pois a fraca qualidade da imagem final vai contribuir para o aumento da frequência espacial da imagem fundida [30].

Tal como acontece com a entropia, a frequência espacial é uma métrica adequada para comparar a qualidade de imagens fundidas entre os métodos implementados. Isoladamente, a Frequência Espacial não pode medir a qualidade da fusão da imagem fundida porque não sabe distinguir a informação útil do ruído ou artefactos. A frequência espacial é o melhor indicador de desempenho quando não se considera as imagens fonte [9].

Os valores adquiridos no presente estudo (*média* ~ 27) encontram-se na mesma gama de valores obtidos por outros estudos [9], em que a *média* dos resultados é ~ 17.

<sup>2</sup> Traduzido de “overall activity”

#### 4.6.4 Informação Mútua

A informação mútua mede o grau de dependência entre duas imagens. O valor da informação mútua é a soma entre a informação mútua de cada imagem fonte com a imagem fundida, portanto quanto maior for o seu valor maior será a dependência entre as imagens fonte e a imagem fundida.

Os melhores resultados são obtidos quando se utiliza as wavelets com a 1ª regra fusão, tanto com um como com dois níveis de decomposição; os resultados com a PCA são os piores (ver figura 47). Os valores mais altos devem-se ao facto de, ao serem selecionados os pixéis com maior intensidade, a transferência de informação é maior. Por outro lado, na pirâmide laplaciana e nas wavelets com a 2ª regra de fusão é utilizada a média dos pixéis, e por essa razão os valores da informação mútua são mais baixos.

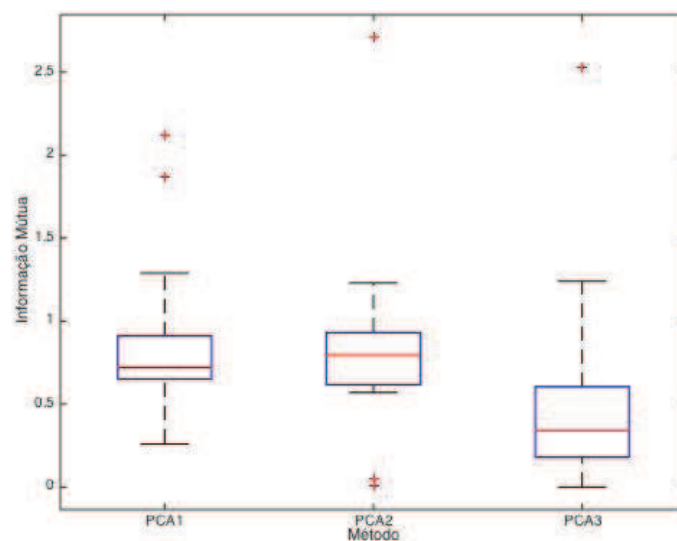


Figura 47 – Boxplot da Informação Mútua

Os valores obtidos (*média* ~2.5) estão dentro da gama de valores adquiridos noutros estudos (*média* ~ 4) [7].

#### 4.6.5 Índice de Qualidade de Fusão

O índice de qualidade fusão mede a semelhança entre a imagem fundida e as imagens fonte. Toma valores entre zero e um, sendo que quanto maior for a semelhança mais próximo estará do valor um. Idealmente, a imagem fundida deve apresentar ambas as imagens fonte.

Tal como se pode ver na figura 48, os melhores resultados são obtidos quando se utiliza o método que utiliza a PCA, seguidos do método que utiliza a pirâmide laplaciana. Contrariamente ao que se tem verificado nas métricas anteriores, os métodos baseados na PCA obtém os valores mais elevados. Isto pode ser explicado pelo facto de que, neste método, é feita uma média ponderada das imagens fonte para obter a imagem fundida e, assim, ambas as imagens fonte contribuem para a

imagem final da mesma forma. Ou seja, enquanto nos métodos que utilizam as wavelets existem coeficientes que são selecionados apenas de uma das imagens fonte (havendo coeficientes de detalhe de uma das imagens fonte que não contribuem para a imagem fundida), nos métodos baseados na PCA isso não acontece.

Por outro lado, a pirâmide laplaciana produz resultados melhores do que as wavelets porque preserva os detalhes e os contornos das imagens fonte, mantendo desta forma a semelhança entre as imagens e a imagem fundida.

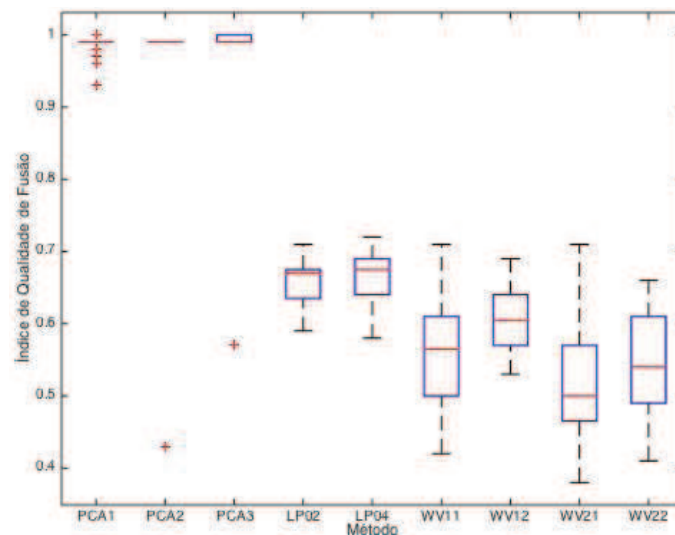


Figura 48 – Boxplot do Índice de Qualidade de Fusão

Esta métrica é considerada como sendo a mais fidedigna para as imagens fundidas que não têm imagem de referência [11], no entanto, como se pode verificar, os resultados obtidos para os métodos que utilizam a PCA são enganadores.

Os valores obtidos (*média* ~ 0.6, excluindo os valores da PCA) encontram-se um pouco abaixo da média dos obtidos por outros autores (*média* ~0.8) [9].

#### 4.6.6 Índice de Semelhança Estrutural

O índice de semelhança estrutural mede a semelhança da estrutura da informação presente nas imagens a ser comparadas; é a soma entre o índice de semelhança estrutural de cada imagem fonte com a imagem fundida. Quanto maior for o valor obtido, maior será a semelhança entre imagens.

Os valores obtidos para as três variantes da PCA foram próximos de zero, e por esse motivo optou-se por apresentar apenas os restantes valores, que se encontram na figura 49. Como se pode ver na figura 49, os resultados obtidos são bastante diferentes dos que foram adquiridos para as métricas anteriores. Ao contrário do que aconteceu até agora, a pirâmide laplaciana com dois níveis obteve melhores resultados, bem como as wavelets que utilizaram a 2ª regra de fusão.

Estes resultados são explicados por quanto menor for o nível de decomposição da pirâmide laplaciana mais semelhante a imagem fundida será com as imagens fonte, e como a 2ª regra de fusão das wavelets utiliza a média dos coeficientes de aproximação a imagem fundida também será mais próxima das imagens fonte.

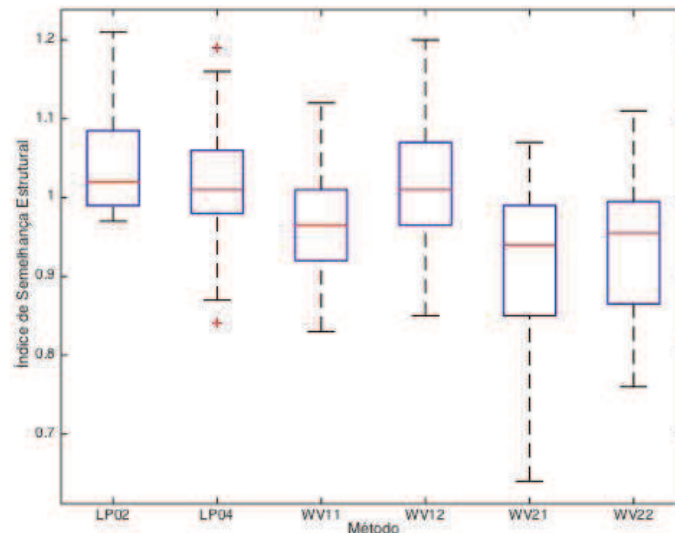


Figura 49 – Boxplot do Índice de Semelhança Estrutural

#### 4.7 Comparação da Avaliação Qualitativa e Quantitativa

Nesta secção é feita uma comparação entre os resultados obtidos pela análise qualitativa e a análise quantitativa, de forma a comentar se a percepção visual está de acordo com os resultados numéricos obtidos. Na análise qualitativa, os métodos que apresentaram melhores resultados foram os que utilizam as wavelets com a 1ª regra de fusão e a pirâmide laplaciana com quatro níveis. Os melhores resultados foram tidos em imagens cuja componente térmica se encontra bem marcada, bem como a presença de detalhe proveniente da imagem da banda do visível. No entanto, por vezes, torna-se difícil escolher entre dois resultados que sejam melhores do que os outros, pois apenas com a inspeção visual é difícil tirar conclusões objetivas. Por esse motivo, recorre-se a uma avaliação quantitativa, para verificar se os melhores resultados obtidos por inspeção visual correspondem aos melhores resultados numéricos.

Das seis métricas de desempenho implementadas, conclui-se que o desvio padrão, a entropia e a frequência espacial são métricas apropriadas para fazer uma comparação relativa entre os métodos implementados (pois medem a qualidade da imagem fundida), contudo, por si só estas métricas não têm em consideração a relação existente entre a imagem fundida e as imagens fonte. Assim, a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural já vão ter em conta essa relação, sendo por isso consideradas as métricas mais adequadas para medir a qualidade da fusão de imagem.

Os resultados obtidos com estas métricas de desempenho coincidiram, na sua maioria, com os resultados da inspeção visual e, a única métrica que não forneceu resultados coincidentes, foi o índice de semelhança estrutural. No entanto, os resultados que dela advieram são compreensíveis, porque ao medir a semelhança entre os pixels da imagem fundida e das imagens fonte, os valores mais elevados vão ser atribuídos aos métodos baseados na PCA que fazem uma média ponderada entre as imagens fonte, e dessa forma ambas as imagens contribuem em grande parte para a imagem fundida.

Face à discussão apresentada, conclui-se que na maioria dos casos a inspeção visual constitui uma boa avaliação dos métodos atendendo a que os resultados qualitativos são frequentemente similares aos resultados quantitativos. Mas quando é necessário documentar quantitativamente a similaridade entre imagens, então deve-se usar as métricas acima indicadas.

## 5 Conclusões

A maioria dos sistemas de vigilância atuais utiliza sistemas de detecção através da cor, no entanto, estes estão muito limitados pela iluminação. Assim, tem sido proposto o uso de câmaras do espectro infravermelho-térmico, que capturam a imagem térmica do objeto. Devido à sua capacidade de utilização diurna e noturna, em diversas condições atmosféricas, têm um grande número de aplicações, entre as quais a detecção de pessoas camufladas. Para melhorar os resultados da detecção de pessoas, a informação visível pode ser combinada com a informação térmica. Enquanto as imagens da banda do visível fornecem um contexto visual às pessoas, as imagens térmicas fornecem informação acerca de objetos com temperaturas elevadas. Desta forma, a fusão das duas imagens proporciona uma boa percepção visual, e permite a detecção de pessoas mais facilmente. As tecnologias de fusão de imagem permitem aos militares obter uma vigilância, proteção e detecção necessárias à sua segurança contra a ameaça do Inimigo e, ao mesmo tempo, também permite aos militares obter informação acerca do Inimigo. Ao combinar a informação gerada por múltiplas fontes de imagem, a fusão permite um maior detalhe, resultando assim num maior grau de segurança.

O presente estudo surge da necessidade da implementação de um método que permita a detecção de pessoas camufladas, enquanto estas estão enquadradas no seu ambiente, tornando a sua camuflagem ineficiente. Assim, o objetivo da presente dissertação é a implementação de métodos de fusão de imagem ao nível do pixel, que pretendem fundir imagens da banda do visível com imagens da banda do infravermelho-térmico, obtendo uma imagem fundida mais rica do que as imagens fonte. Depois de serem implementados métodos distintos, pretende-se discernir qual dos métodos obtém os melhores resultados, sendo por isso o mais adequado à fusão de imagem. Para tal, primeiramente será feita uma análise qualitativa, seguida de uma análise quantitativa, como forma de fundamentar a avaliação feita por inspeção visual. Os três métodos de fusão de imagem que foram implementados são: o método da análise em componentes principais (PCA), o método que utiliza a pirâmide laplaciana e o método que usa a análise em wavelets.

A PCA envolve um procedimento matemático que transforma um número de variáveis correlacionadas num número de variáveis não-correlacionadas que são chamadas de componentes principais. A fusão de imagem, utilizando este método, é alcançada através de uma média ponderada das imagens a serem fundidas. Os pesos para cada imagem fonte são obtidos a partir dos vetores próprios correspondentes ao maior valor próprio das matrizes de covariância de cada imagem de fonte. Neste método, foram implementadas três variantes: a primeira variante, onde são calculadas as componentes principais a partir das imagens fonte e depois multiplicadas pelas imagens fonte; a segunda variante da PCA, onde as componentes principais são calculadas a partir das imagens fonte, após ter sido feita uma equalização de histograma, e são multiplicadas pelas imagens fonte equalizadas; a terceira e última variante, onde os valores das componentes principais são calculados a partir das imagens fonte equalizadas, e depois são multiplicados pelas imagens fonte originais.

O princípio do método que utiliza a pirâmide laplaciana consiste em decompor a imagem fonte em subimagens com diferentes resoluções espaciais, e a estratégia de fusão de imagem consiste em usar uma regra de fusão para construir uma representação de uma pirâmide fundida a partir das

representações em pirâmide das imagens fonte. A imagem fundida é obtida ao fazer a transformada inversa da pirâmide. Neste método, foi implementada a construção de pirâmides com dois, três e quatro níveis, não podendo ser feita a decomposição além dos quatro níveis devido ao tamanho das imagens fonte não o permitir. Foram implementados três níveis de forma a poder comparar os resultados obtidos.

O método que utiliza as wavelets decompõe as imagens fonte em matrizes de coeficientes ou subimagens. Esses coeficientes são denominados de coeficiente de aproximação, que representa a aproximação da imagem fonte, e coeficientes de detalhe (horizontal, vertical e diagonal) que representam subimagens detalhadas que contém informação direcional. Estes coeficientes são fundidos de acordo com regras de fusão, e a imagem fundida é obtida fazendo a transformada wavelet inversa. Para este método, foram testados dois níveis de decomposição e duas regras de fusão. A primeira regra de fusão consiste na utilização do máximo dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada. A segunda regra de fusão consiste na utilização da média dos coeficientes de aproximação à maior escala e o maior valor absoluto dos coeficientes de detalhe em cada escala da transformada. Apenas foram experimentados dois níveis de decomposição porque, com dois níveis de decomposição, a qualidade da imagem fundida já é degradada.

As imagens utilizadas para o presente estudo têm dimensão de  $320 \times 240$  pixels e cada uma é constituída por uma imagem da banda do visível e uma imagem térmica. Foram selecionadas 20 imagens que correspondem a pessoas camufladas em diferentes meios circundantes e em diferentes condições de luminosidade. Após ter sido feita uma comparação inicial das imagens obtidas a partir de cada método, concluiu-se que as melhores imagens foram obtidas utilizando a terceira variante da PCA, a pirâmide laplaciana com quatro níveis e as wavelets com um nível de decomposição que utilizam a primeira regra de fusão. De seguida, foi efetuada uma análise qualitativa das imagens fundidas tendo como base dez conjuntos de três imagens, imagens essas selecionadas a partir da comparação inicial. Desta análise, conclui-se que o método que obtém melhores resultados é aquele que utiliza as wavelets, seguido do método que utiliza a pirâmide laplaciana e por último o que utiliza a PCA. O método que utiliza a PCA muitas vezes assemelha-se à fusão através da média simples o que produz um contraste de características baixo, sendo por isso o que obtém piores resultados. O método que utiliza as wavelets é o que apresenta melhores resultados, pois é aquele que faz a fusão das várias componentes da imagem (aproximação, horizontal, vertical e diagonal) de acordo com as regras de fusão definidas como sendo as melhores para o objetivo pretendido, que é detetar as pessoas camufladas. Contudo, uma análise qualitativa é insuficiente para comparar os diferentes métodos. É necessária uma análise quantitativa que permita a comparação objetiva e automática do desempenho dos métodos utilizados.

A análise quantitativa foi realizada usando como métricas de desempenho o desvio padrão, a entropia, a frequência espacial, a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural. Esta análise foi efetuada tendo como base 20 imagens fundidas para cada um dos métodos e das suas variantes, perfazendo um total de nove metodologias. Os resultados obtidos para cada uma das métricas foram apresentados em gráficos *boxplot*, onde se pode ver a mediana dos resultados obtidos para cada um dos métodos, e a dispersão dos resultados acima e abaixo da

mediana. Com estes gráficos tem-se uma percepção melhor do conjunto de resultados obtidos, e por esse motivo foram utilizados para apresentar os resultados. Ao observar os gráficos obtidos, os valores adquiridos para as três variantes do método que utiliza a PCA são os que mais se destacam pela negativa, pois estão bastante aquém dos obtidos para os restantes métodos, o que apenas com a inspeção visual não era perceptível. Os valores adquiridos para os métodos que utilizam a pirâmide laplaciana e a transformada wavelet discreta encontram-se dentro dos intervalos obtidos por outros autores, e suportam as conclusões retiradas da análise qualitativa. Das seis métricas de desempenho implementadas, conclui-se que o desvio padrão, a entropia e a frequência espacial são métricas apropriadas para fazer uma comparação relativa entre os métodos implementados (pois medem a qualidade da imagem fundida), contudo, por si só estas métricas não têm em consideração a relação existente entre a imagem fundida e as imagens fonte. Assim, a informação mútua, o índice de qualidade de fusão e o índice de semelhança estrutural já vão ter em conta essa relação, sendo por isso consideradas as métricas mais adequadas para medir a qualidade da fusão de imagem.

Concluindo, ao comparar os resultados obtidos qualitativa e quantitativamente, observa-se que os métodos tidos como melhores na avaliação qualitativa são de facto os melhores na avaliação quantitativa, ou seja, o método que utiliza as wavelets com um nível de decomposição com a primeira regra de fusão e o método que usa a pirâmide laplaciana com quatro níveis. Apesar da análise qualitativa constituir uma boa avaliação dos métodos, é necessária uma análise quantitativa que suporte as conclusões retiradas por inspeção visual, e que nos permita identificar qual o método mais adequado para o objetivo pretendido.

## **5.1 Trabalho Futuro**

O trabalho desenvolvido mostrou-se eficaz na deteção de pessoas camufladas, na maioria dos casos. No entanto, houve imagens onde tal não aconteceu de forma clara e as imagens fundidas são pobres em informação. Desta forma, uma das possibilidades para trabalho futuro é fazer a fusão de imagem ao nível das características, e posteriormente ao nível da decisão.

Os métodos de fusão ao nível das características são o próximo nível de processamento onde a fusão de imagem pode tomar lugar. A fusão a este nível requer a extração de objetos/características das imagens fonte, e estas características são depois combinadas com as características presentes nas outras imagens de entrada através de um processo de seleção predeterminado, para formar a imagem fundida final. Dado que um dos objetivos da fusão é preservar as características da imagem, os métodos ao nível das características têm a capacidade de produzir imagens fundidas com melhor qualidade subjetiva. Os algoritmos mais comuns para a fusão de imagens ao nível das características incluem os métodos de deteção de arestas e classificadores.

Os métodos de fusão ao nível da decisão são o nível de processamento mais alto onde a fusão de imagem pode ser efetuada. A fusão ao nível da decisão leva a fusão ao nível das características um passo mais à frente ao declarar identidades aos objetos reconhecidos, nas imagens de entrada individuais, e depois atribuir uma medida de qualidade às características extraídas. A informação obtida

é depois combinada ao aplicar regras de decisão para reforçar a interpretação comum e resolver diferenças dos objetos observados. Alguns algoritmos comuns usados ao nível da decisão incluem a lógica de Fuzzy, fusão baseada em regras e redes Bayesianas.

## Referências

- [1] Ministério da Defesa Nacional, *Informações, Contra-Inteligência e Segurança*.: Exército Português, 2009.
- [2] Ministério da Defesa Nacional, *Instruções de Segurança Militar do Exército Português*.: Exército Português, 2013.
- [3] Núcleo de Operações/Informações e Tática, *Manual de Informações*. Portugal: Escola da Guarda, 2010.
- [4] Aparna Akula, Ripul Ghosh, and H K Sardana, "Thermal Imaging And Its Application In Defence Systems," *Optics: Phenomena, Materials, Devices, and Characterization* , pp. 333-335, 2011.
- [5] Juan Serrano-Cuerda, Maria Teresa Lopez, and Antonio Fernandez-Caballero, "Robust Human Detection and Tracking in Intelligent Environments by Information Fusion of Color and Infrared Video," in *Seventh International Conference on Intelligent Environments*, 2011, pp. 354-357.
- [6] Thi Thi Zin, Hideya Takahashi, Takashi Toriu, and Hiromitsu Hama, *Fusion of Infrared and Visible Images for Robust Person Detection*, Osamu Ukimura, Ed. Osaka, Japan: InTech, 2011.
- [7] Senthil Kumar Sadhasivam, Mahesh Bharath Keerthivasan, and Muttan S, "Implementation of Max Principle with PCA in Image Fusion for Surveillance and Navigation Application," *Electronic Letters on Computer Vision and Image Analysis*, Universitat Autònoma de Barcelona, Barcelona, 1577-5097, 2011.
- [8] Yufeng Zheng, "Multi-scale Fusion Algorithm Comparisons: Pyramid, DWT and Iterative DWT," in *12th International Conference on Information Fusion*, Seattle, 2009, pp. 1060-1067.
- [9] V. P. S. Naidu and J. R. Raol, "Pixel-Level Image Fusion using Wavelets and Principal Component Analysis," *Defence Science Journal*, vol. 58, no. 3, pp. 338-352, May 2008.
- [10] Mrinal Kanti Bhowmik, Debotosh Bhattacharjee, Mita Nasipuri, Dipak Kumar Basu, and Mahantapas Kundu, "Fusion of Wavelet Coefficients from Visual and Thermal Face Images for Human Face Recognition - A Comparative Study," *International Journal of Image Processing (IJIP)*, vol. 4, no. 1, pp. 1-12.
- [11] Yufeng Zheng, Edward A. Essock, and Bruce C. Hansen, "An Advanced Image Fusion Algorithm Based on Wavelet Transform - Incorporation with PCA and Morphological Processing," Department of Psychological & Brain Sciences, University of Louisville, Louisville, 2004.
- [12] Hang Tan, Xianhe Huang, Huachum Tan, and Changtao He, "Pixel-level Image Fusion Algorithm Based on Maximum Likelihood and Laplacian Pyramid Transformation," *Journal of Computational Information Systems*, pp. 327-334, January 2013.
- [13] V. P. S. Naidu, "Discrete Cosine Transform-based Image Fusion," *Defence Science Journal*, vol. 60, no. 1, pp. 48-54, January 2010.
- [14] V. P. S. Naidu and Bindu Elias, "A Novel Image Fusion Technique using DCT based Laplacian Pyramid," *International Journal of Inventive Engineering and Sciences*, vol. 1, no. 2.


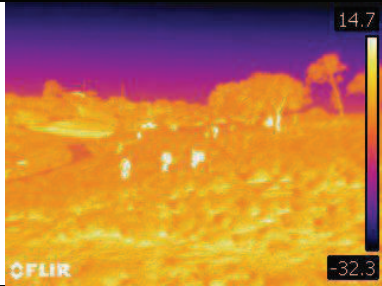

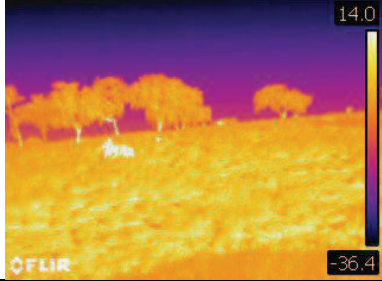

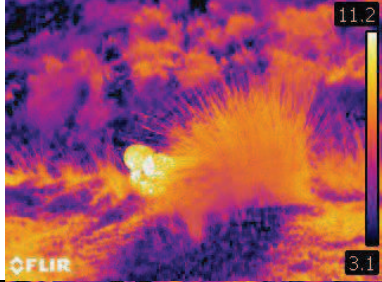

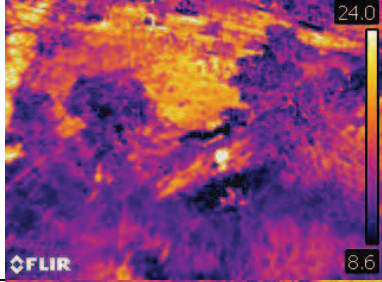

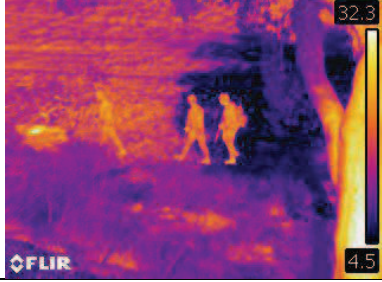
- [15] Rohan Ashok Mandhare, Pragati Upadhyay, and Sudha Gupta, "Pixel-Level Image Fusion Using Brovey Transform and Wavelet Transform," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 6, pp. 2690-2695, June 2013.
- [16] Oliver Rockinger, "Image Sequence Fusion Using a Shift-Invariant Wavelet Transform," Systems Technology Research, Intelligent Systems Group, Berlin.
- [17] A. Mujahid Khan, M. Fayyaz, and S. A. M. Gilani, "Thermal and Visible Image Fusion: a Machine Learning Approach," *International Review on Computers and Software*, vol. 2, no. 6, pp. 643-652, November 2007.
- [18] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik, "Blobworld: Image segmentation using Expectation-Maximization and its application to image querying".
- [19] James W. Davis and Vinay Sharma, "Background-Subtraction in Thermal Imagery Using Contour Saliency," *International Journal of Computer Vision*, vol. 71, no. 2, pp. 161-181, 2007.
- [20] James W. Davis and Mark A. Keck, "A Two-Stage Template Approach to Person Detection in Thermal Imagery," Department of Computer Science, Ohio State University, Columbus,.
- [21] Yining Deng, B. S. Manjunath, and Hyundoo Shin, "Color Image Segmentation," Department of Electrical and Computer Engineering, University of California, Santa Barbara.
- [22] AR15.com. [Online]. <https://www.ar15.com/archive/topic.html?b=6&f=18&t=370436>
- [23] Peter J. Burt and Edward H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions On Communications*, vol. 31, no. 4, pp. 532-540, April 1983.
- [24] G. Geetha, S. Raja Mohammad, and Y. S. S. R. Murthy, "Multifocus Image Fusion Using Multiresolution Approach With Bilateral Gradient Based Sharpness Criterion," Computer Science & Engineering Department, Sri Vishnu Engineering College for Women, Bhimavaram, 2012.
- [25] (2015, Setembro) Wikipedia, The Free Encyclopedia. [Online]. <https://en.wikipedia.org/wiki/Wavelet>
- [26] (2015, Setembro) Wikipedia, The Free Encyclopedia. [Online]. [https://en.wikipedia.org/wiki/Daubechies\\_wavelet](https://en.wikipedia.org/wiki/Daubechies_wavelet)
- [27] Lau Wai Leung, Bruce King, and Vijay Vohora, "Comparison Of Image Data Fusion Techniques Using Entropy and INI," Centre for Remote Imaging, Sensing and Processing, National University of Singapore, Singapore, November 2001.
- [28] Ahmet M. Eskicioglu and Paul S. Fisher, "Image Quality Measures and Their Performance," *IEEE Transactions On Communications*, vol. 43, no. 12, pp. 2959-2965, December 1995.
- [29] Gemma Piella and Henk Heijmans, "A New Quality Metric For Image Fusion," Amsterdam.
- [30] Yufeng Zheng, Edward A. Essock, Bruce C. Hansen, and Andrew M. Haun, "A new metric based on extended spatial frequency and its application to DWT based fusion algorithms," *Information Fusion*, no. 8, pp. 177-192, April 2007.


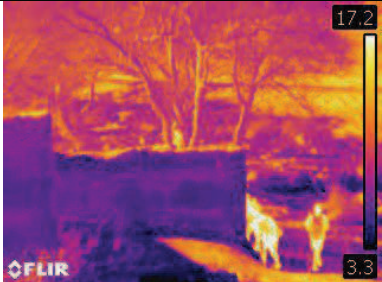



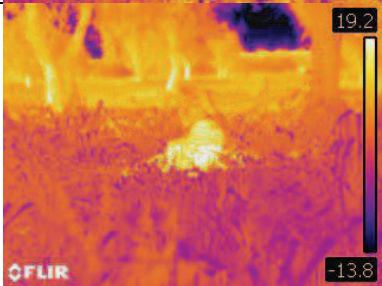



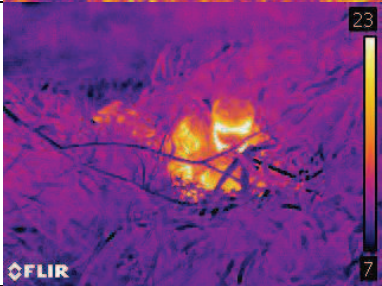


- [31] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions On Image Processing*, vol. 13, no. 4, pp. 1-14, April 2004.




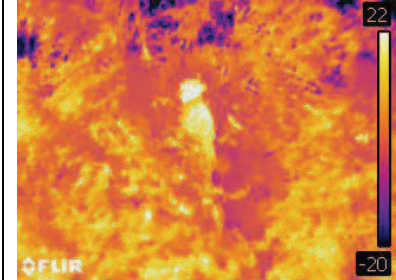



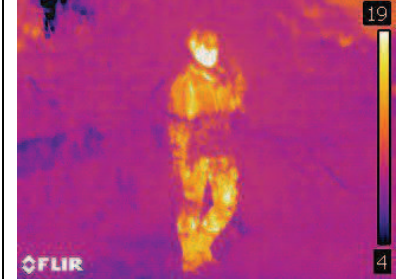

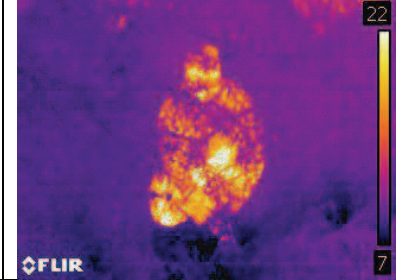

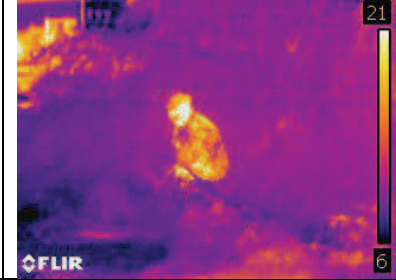
# Anexo 1


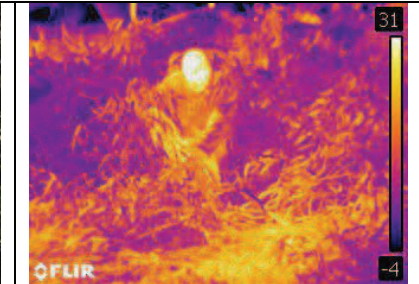



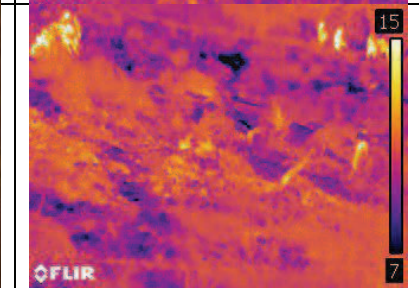
## Imagens utilizadas

Tabela 1 - Imagens utilizadas para o estudo, onde a imagem da esquerda corresponde à imagem da banda do visível e a imagem da direita corresponde à imagem térmica

Imagem	Imagem Visível	Imagem Térmica	Descrição
A			Pessoas a progredir numa encosta
B			Pessoas paradas na encosta
C			Pessoa escondida atrás de um arbusto
D			Pessoas paradas no vale
E			Três pessoas a progredir no terreno plano

F			<p>Pessoas a circundar uma ruína</p>
G			<p>Pessoa deitada e camuflada na relva e arbustos</p>
H			<p>Pessoa deitada e camuflada na relva e arbustos</p>
I			<p>Pessoa camuflada na relva e arbustos</p>
J			<p>Pessoa deitada e camuflada entre troncos e folhas secas</p>
K			<p>Pessoa deitada e camuflada entre troncos e folhas secas</p>

L			Pessoa escondida atrás de um tronco de árvore
M			Pessoa camuflada entre um arbusto esparso
O			Pessoa camuflada entre um arbusto esparso
P			Pessoa camuflada entre um arbusto esparso
Q			Pessoa camuflada entre um arbusto esparso
R			Pessoa camuflada entre um arbusto esparso

S			<p>Pessoa sentada entre folhas secas e troncos</p>
T			<p>Pessoa dentro de um abrigo, a empunhar uma espingarda automática G3</p>
U			<p>Pessoas a andar no terreno inclinado</p>