Image Fusion for the Detection of Camouflaged People

Nádia Alexandra Ferreira Bento, nadiafbento@gmail.com

Academia Militar, Lisbon
Instituto Superior Técnico, Lisbon

Abstract— **The use of thermal imaging is a benefit for the Armed Forces. Due to their advantages, they have a large number of applications, including the detection of camouflaged people. For better results, the thermal information can be merged with the color information which allows a greater detail, resulting in a greater degree of security.**

The present study implemented as pixel level image fusion methods: Principal Components Analysis; Laplacian Pyramid; and Discrete Wavelet Transform.

A qualitative analysis concluded that the method which performs better is the one that uses Wavelets, followed by the Laplacian Pyramid and finally the PCA. A quantitative analysis was made using as performance metrics: 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. 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: **Laplacian Pyramid, Pixel level image fusion, Performance Metrics, Principal Component Analysis, Security, Wavelets.**

I. Introduction

Nowadays the majority of surveillance systems uses detection systems trough color, however these systems are highly limited by luminosity. Therefore, has been proposed the use of infrared cameras, which capture the thermal image of an object. So using thermal images has been a benefit for the military, due to its ability to daytime and nighttime use, as well under different weather conditions. These images are popularly used by the Army and the Navy to frontiers surveillance or coastal surveillance and establishment of order [1]. In this context, these images can be used to detect camouflaged people.

However, to enhance the results of people detection, the color information can be combined with thermal information. While the color images give a visual context to objects, thermal images give information about objects with high temperature. So the fusion of both images gives a better visual perception of the scene and it allows the detection of people more easily. Thus, the image fusion technologies allow to obtain surveillance, protection and detection that are necessary for military's safety.

This paper aims to implement pixel level image fusion methods to detect camouflaged people and it aims to discern which method gets the best results. For this will be done primarily a qualitative analysis followed by quantitative analysis.

II. Background

Over the years, various image fusion techniques have been proposed to cope with its growing demand, as there are several areas that benefit from this process, including: medical, military, surveillance and navigation.

Naidu and Raol [2] make a comparison of pixel level fusion methods using Wavelets and Principal Component Analysis (PCA). Regarding to Wavelets, the fusion rule used was the simple average for the approximation coefficients and detail coefficients with the largest absolute value, and were tested five levels of decomposition. In their study were implemented some metrics, with and without a reference image to evaluate the performance of image fusion algorithms. Metrics such as Standard Deviation, Entropy, Cross-Entropy and Spatial Frequency were considered appropriate when there is no reference image. Finally, concluded that the image fusion using Wavelets with a greater degree of decomposition has better performance.

Zheng [3] makes a comparison of multi-scale pixel level fusion algorithms, such as: different Pyramids, Discrete Wavelet Transform (DWT) and Iterative DWT. Zheng proposed the Advanced DWT algorithm for image fusion. In this algorithm, the approximation coefficients at the largest scale of the input image are fused by applying the PCA to the absolute values of these coefficients. It was subsequently optimized with an iterative procedure using the fusion metrics: Image Quality Index and Error Rate of the Spatial Frequency. The author used three pairs of images; the assessment of the fused images was qualitatively and quantitatively made. As quantitative performance metrics, he used the Entropy and Spatial Frequency. The quantitative results show that the iterative algorithms have better performances, followed by Laplacian Pyramid and finally the DWT.

Sadhasivam, Keerthivasan and Muttan [4] make the image fusion by implementing PCA using the maximum principle. Since the results obtained from the traditional PCA show a low performance when compared with other hybrid algorithms, in this study the authors implemented an algorithm that uses the DWT in conjunction with PCA. The low-frequency coefficient is chosen according to the maximum rule, and PCA is to be applied to the high frequency coefficients to determine their weights for the fusion. The final image is obtained by adding the low and high frequency images. They used three sets of images and the performance of this algorithm was measured by Entropy, Mutual Information and a measure based on Structural Similarity Index. While Entropy presents similar values for the compared methods, Structural Similarity and the Mutual Information in the fused image have better results for the proposed method.

Zheng, Essock and Hansen [5] develop an algorithm that incorporates the PCA in DWT. The PCA is applied to the approximation coefficients (low frequency coefficients), whereas the detail coefficients are chosen in accordance with the largest absolute value. The proposed algorithm is compared to other fusion techniques using Entropy, the Spatial Frequency and Image Quality Index, in cases where there is no reference image. The developed algorithm obtained the best results.

Naidu and Elias [6] proposed a new image fusion technique using a Discrete Cosine Transform (DCT)-based Laplacian Pyramid, and decomposed the input images up to eight levels of decomposition. Given that there is a reference image, the metrics used for the quantitative analysis were the Root Mean Square Error (RMSE), the Peak Signal-to-Noise Ratio (PSNR), the Spatial Frequency and Standard Deviation. They concluded that fusion with a higher level of decomposition achieves better results, to any one of the used metrics.

III. Methods

In this section, it is done a brief explanation of the methods used in this study: PCA, Laplacian Pyramid and Wavelets. Then, will be presented the performance metrics used for quantitative analysis.

A. Principal Component Analysis

The PCA involves a mathematical procedure that transforms a number of correlated variables into a number of uncorrelated variables called principal components.

The image fusion, using this method, is achieved through a weighted average of the images to be fused. The weights for each input image are obtained from the eigenvectors corresponding to the highest eigenvalue of covariance matrices for each input image.

1) Principal components computation

The input images (images to be fused) are arranged in two columns arrays. The steps to design the data in a two dimensional subspace are the following:

1. Arrange the data in a matrix $\boldsymbol{Z} = [\,\boldsymbol{z}_1,\ \boldsymbol{z}_2, \dots, \boldsymbol{z}_n]$, where $\boldsymbol{z}_i \in \mathbb{R}^2$ e $\boldsymbol{z}_i = \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix}$.
2. Compute $\bar{\boldsymbol{z}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i$.
3. Obtain $\boldsymbol{X} = [\,\boldsymbol{x}_1,\ \boldsymbol{x}_2, \dots, \boldsymbol{x}_n]$, where $\boldsymbol{x}_i = \boldsymbol{z}_i - \bar{\boldsymbol{z}}$.
4. Find the covariance matrix $\boldsymbol{C} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$.
5. Compute $\boldsymbol{C} = \boldsymbol{V} \sum \boldsymbol{V}^T$, where $\boldsymbol{V} = [v_1, v_2]$ and $\sum diag(\lambda_1, \lambda_2)$. The pair $(\boldsymbol{v}_i, \lambda_i)$ with $i = 1{,}2$ corresponds to the pair eigenvector/eigenvalue of de C, where $\boldsymbol{v}_i = \begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix}$.
6. Consider the highest eigenvalue to compute $p_{1_i} = \frac{v_{1i}}{v_{1i}+v_{2i}}$ and $p_{2_i} = \frac{v_{2i}}{v_{1i}+v_{2i}}$.

2) Image Fusion

The information flow diagram of the fusion algorithm based on PCA is shown in Fig. 1.

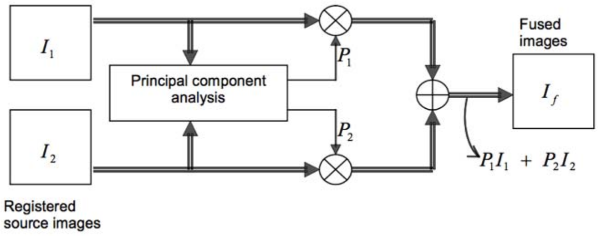

Fig. 1. Image fusion with PCA [2]

The fused image is given by:

$$I_{fus} = p_{1_i} I_{vis} + p_{2_i}\, I_{therm} \tag{1}$$

B. Laplacian Pyramid

The principle of this method is to decompose the input image in sub-images with different spatial resolutions. A fusion rule is used to construct a representation of a fused pyramid and the fused image is obtained by doing the inverse transform pyramid (see Fig. 2).

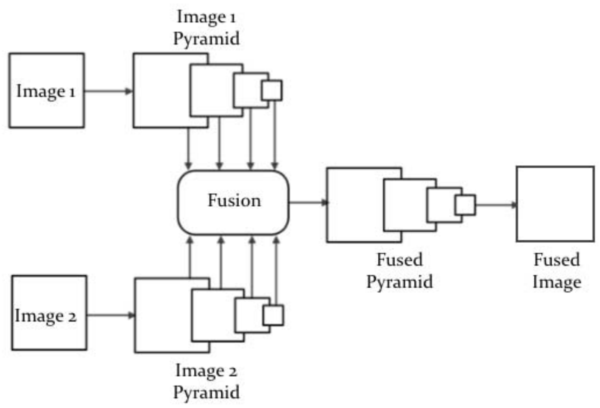

Fig. 2. Image fusion with Laplacian Pyramid [7]

In a study by Zheng, Essock and Hansen [5], the Laplacian pyramid showed the best results and, therefore, is the one that has been implemented in this study. In the Laplacian Pyramid, the lowest level of the pyramid is built from the original input image and each of the other levels is built recursively from its lowest level by applying four basic steps [8]:

1. Blurring;
2. Subsampling;
3. Expansion by interpolation;
4. Differentiation.

1) Image Decomposition

There are two standard operations, the operation "Reduce", and the operation "Expand", which is the inverse of "Reduce", which aims to expand an array $(M+1)\times(N+1)$ in an array $(2M+1)\times(2N+1)$, when filling zeroes in the horizontal direction and zeroes in the vertical direction.

So "Expand" applied to $\hat{X}_l$ array would generate an array with the same size that X_{l-1}. The two operators are defined by:

$$X_l = Reduce(X_{l-1}) \tag{2}$$

$$\hat{X}_l = Expand\ (X_l) \tag{3}$$

The Laplacian Pyramid is applied to decompose the input images at N levels, which corresponds to the n-th level of the pyramid. Thus, the Laplacian Pyramid is obtained from the following definition:

$$\begin{cases} LP_l = X_l - \hat{X}_{l+1}, 0 \le l < N \\ LP_N = X_N, l = N \end{cases} \tag{4}$$

2) Image Reconstruction

The reconstruction of the image from the Laplacian Pyramid is the inverse process of decomposition and in the reverse direction, from the top level to the bottom level, with the following definition:

$$\begin{cases} X_N = LP_N, l = N \\ X_l = LP_l + \hat{X}_{l+1}, 0 \le l < N \end{cases} \tag{5}$$

3) Image Fusion

Having two images to fuse, I_{vis} and I_{therm}, the construction of the pyramid is made for each image individually. At the n-th level of the pyramid the fusion rule is the following: $X_{l+1}^{fus} = \frac{X_{l+1}^{vis} + X_{l+1}^{therm}}{2}$. At level N, $X_N^{fus} = LP_N^{fus}$, where $LP_N^{fus} = \frac{LP_N^{vis} + LP_N^{therm}}{2}$. From level $N-1$ to level 0 of the pyramid, $X_l^{fus} = LP_l^{fus} + \widehat{X_{l+1}}^{fus}$, where $LP_l^{fus} = \begin{cases} LP_l^{vis}, & |LP_l^{vis}| \ge |LP_l^{therm}| \\ LP_l^{therm}, & |LP_l^{vis}| < |LP_l^{therm}| \end{cases}$, and amplitude comparison is made in the corresponding pixels. The pyramid $I_{fus} = X_0^{fus}$ is the fused image.

C. Wavelets

Wavelets theory has been widely used in image processing and provides a multi-resolution decomposition of an image. Wavelets give a good resolution both in time and frequency domain.

There are several families of Wavelets: Haar, Daubechies, Coiflets, Symlets, Discrete Meyer, Biorthogonal and Reverse Biorthogonal. In the implemented method is used the Haar Wavelet family, being the simplest and the one that achieved better results in this work.

1) Image Decomposition

The wavelet filters and subsample the image in vertical and horizontal directions. The input image I is filtered by a low pass filter and a high pass filter in the horizontal direction and then is subsampled by a factor of 2 to create the matrix of coefficients I_L and I_H. Both matrices of coefficients I_L and I_H are filtered through a low-pass filter and high-pass in the vertical direction and subsampled by a factor of 2 to create the matrix of coefficients, or sub-images, I_{LL}, I_{LH}, I_{HL} and I_{HH} (see Fig. 3).

The coefficient matrix I_{LL} contains the average information of the image corresponding to the low frequency band of the multi-scale decomposition. Can be considered a smoothed and subsampled version of the input image. The coefficient matrices I_{LH}, I_{HL} and I_{HH} represent detailed sub-images containing directional information (horizontal, vertical and diagonal) of the input image.

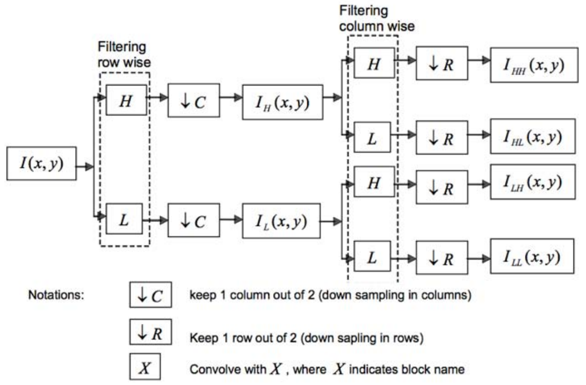

Fig. 3. Image Decomposition [2]

The multi-resolution can be achieved by applying recursively the same algorithm in the low pass coefficients of the previous decomposition.

2) Image Reconstruction

The inverse wavelet transform is used to reconstruct the image I from the sub-images, I_{LL}, I_{LH}, I_{HL} and I_{HH}.

This process involves column up sampling (inserting zeroes between samples) and filtering using a low pass filter and a high pass filter for each sub image. The oversampling and filtering of the lines with the low pass filter and the high pass filter of the resultant image and the sum of all arrays reconstruct the image (see Fig. 4).

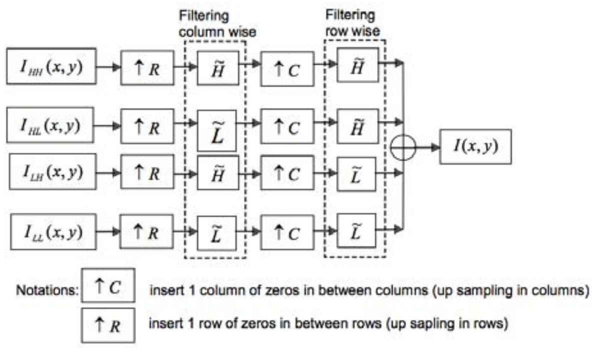

Fig. 4. Image Decomposition [2]

3) Image Fusion

In the image fusion scheme using wavelets, the input images are decomposed into approximation and detail coefficients at a certain level by using the DWT. Then approximation and detail coefficients are combined using the fusion rule ϕ. The fused image can be obtained by using the inverse DWT as (see Fig. 5):

$$I_{fus} = IDWT\left[\phi\{DWT(I_{vis}(x,y)), DWT(I_{therm}(x,y))\}\right] \quad (6)$$

The fusion rules used in this study were:

i) the maximum approximation coefficient at the largest scale and the largest absolute value of the detail coefficients in each transformed scale;
ii) the mean of approximation coefficients at the largest scale and the largest absolute value of the detail coefficients in each transformed scale.

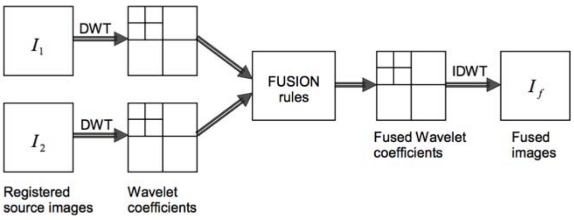

Fig. 5. Image fusion with wavelets [2]

D. Performance Metrics

To measure the quality of the fused image, it is important to make a quantitative assessment such that fusion algorithms can be analyzed and compared objectively.

The metrics used to measure the efficiency of fusion methods are: Standard Deviation, Entropy, the Spatial Frequency, the Mutual Information, Quality Fusion Index and Similarity Structural Index [2], [4].

1) Standard Deviation

The Standard Deviation measures the contrast of the fused image and is defined as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i - \bar{I_i})^2} \quad (7)$$

where I_i is a column vector of N observations and $\bar{I_i}$ is the mean of that same vector. An image with high contrast have a high standard deviation value [2].

2) Entropy

The Entropy can measure the information content of an image, but it can not distinguish the noise information. The entropy is given by [9]:

$$E_n = -\sum_{i=1}^{G} p(i)\log_2 p(i) \quad (8)$$

where G is the number of gray levels in the histogram of the image, typically between 0 and 255, and $p(i)$ is the normalized frequency of occurrence of each gray level.

An image with high information content will have a high entropy.

3) Spatial Frequency

The Spatial Frequency indicates the overall activity level of an image, and is defined as [10]:

$$SF = \sqrt{RF^2 + CF^2} \quad (9)$$

where RF corresponds to row frequency and CF to columns frequency. The frequencies of lines and columns are given by:

$$FL = \sqrt{\frac{1}{n_l n_c}\sum_{i=1}^{n_l}\sum_{j=2}^{n_c}\left[I_{fus}(i,j) - I_{fus}(i,j-1)\right]^2} \quad (10)$$

$$FC = \sqrt{\frac{1}{n_l n_c}\sum_{j=1}^{n_c}\sum_{i=2}^{n_l}\left[I_{fus}(i,j) - I_{fus}(i-1,j)\right]^2} \quad (11)$$

where n_l is the number of lines and n_c the number of columns of an image. A high value for the spatial frequency indicates a high overall activity. The higher its value the more information has the fused image.

4) Mutual Information

This metric measures the degree of dependence between two images. It is calculated by setting the joint histogram of input images I_{vis}, I_{ir}, and the fused image I_{fus} [4]. The mutual information between the input images and the fused image is given by:

$$IM_1(fus, vis) = -\sum p(fus, vis)\log_2\left(\frac{p(fus, vis)}{p(fus).p(vis)}\right) \quad (12)$$

$$IM_2(fus, therm) = -\sum p(fus, therm)\log_2\left(\frac{p(fus, therm)}{p(fus).p(therm)}\right) \quad (13)$$

where $p(fus, vis)$ and $p(fus, therm)$ are the joint histograms of input images and fused image. The efficiency of the fusion algorithm is determined by the IM metric which is defined by:

$$IM = IM_1(fus, vis) + IM_2(fus, therm) \quad (14)$$

A greater dependence, that is, a larger value means better quality.

5) Fusion Quality Index

The Fusion Quality Index of an image measures the similarity between the fused image and both input images and it is defined by:

$$Q_W = \lambda Q_0(I_{vis}, I_{fus}) + (1-\lambda)Q_0(I_{therm}, I_{fus}) \quad (15)$$

where Q_0 is the overall quality index, and λ is a local weight that indicates the relative importance of the input image compared to the fused image. The overall quality index Q_0 of two images is defined by:

$$Q_0(I_{vis}, I_{therm}) = \frac{\sigma_{I_{vis}I_{therm}}}{\sigma_{I_{vis}}\ \sigma_{I_{therm}}} \cdot \frac{2\overline{I_{vis}}\ \overline{I_{therm}}}{(\overline{I_{vis}}^2 + \overline{I_{therm}}^2)} \cdot \frac{2\sigma_{I_{vis}}\ \sigma_{I_{therm}}}{(\sigma_{I_{vis}}{}^2 + \sigma_{I_{therm}}{}^2)} \qquad (16)$$

This metric can assume values between 0 and 1, wherein value 1 corresponds to a better quality of the fused image.

6) Structural Similarity Index

The Structural Similarity Index has been used to indicate the similarity of the structure of information between two images and is defined by [11]:

$$ISE(I_{vis}, I_{therm}) = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{I_{vis}\,I_{therm}} + C_2)}{(\mu_{I_{vis}}^2 + \mu_{I_{therm}}^2 + C_1)(\sigma_{I_{vis}}^2 + \sigma_{I_{therm}}^2 + C_2)} \qquad (17)$$

where μ is the mean of the intensity of the image, σ is the standard deviation of the image and C_1 and C_2 are constants.

The information from each input image in the fused image is calculated separately for each input image, and the total Structural Similarity Index is given by:

$$ISE_T = ISE(fuse, vis) + ISE(fuse, therm) \qquad (18)$$

A larger value indicates that the information present in each input image is present in the fused image.

IV. RESULTS

In this section, the results are presented from each of the methods used, followed by a qualitative analysis and quantitative analysis.

A. Data Set

The images used for this study were acquired with a FLIR T440bx camera and have 320×240 pixels. The images were obtained in two Military Academy exercises. For the performed tests, were selected 20 images, which represents camouflaged people in different scenes and light conditions. The algorithms were implemented using Matlab® software.

B. Principal Component Analysis

Three variants of the PCA method were implemented. In the first variant, the principal components are computed from the original input images, and then multiplied by the input images. In the second variant are used equalized images (obtained from the input images) instead of the original input images. The third and last variant is a hybrid version, where the principal components are obtained through the equalized images, however these values will be multiplied by the original input images.

Fig. 6. a) First variant of PCA b) Second variant of PCA c) Third variant of PCA

The third variant was the one which obtained the best results, therefore was chosen to do the qualitative analysis.

C. Laplacian Pyramid

To compare the results, the decomposition of the input images was performed up to four levels, which means that were built pyramids with two, three and four levels (see Fig. 7).

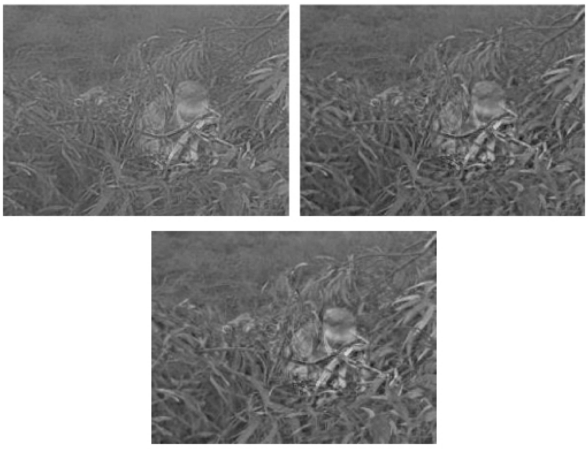

Fig. 7. a) Fused image with 2 levels b) Fused image with 3 levels c) Fused image with 4 levels

While the image obtained with two levels of decomposition presents little detail and a very low contrast, the image obtained with four levels of decomposition provides more detail and contrast.

As the main objective is to obtain a fused image which presents as much information as possible, it becomes clear that the Laplacian Pyramid with four levels of decomposition is the most suitable for the intended purpose.

D. Wavelets

After the decomposition of the input images in approximation and detail coefficients, were used two distinct fusion rules (see section III.C.3) and two levels of decomposition.

The use of the 1st fusion rule gives the clearest results, the fused image has the most marked thermal component, making it easier to identify the camouflaged people. With the 2nd fusion rule, although the camouflaged people is identifiable, the image contrast is low.

Regarding the decomposition levels were tested just one and two levels of decomposition. The images obtained with two levels of decomposition are more degraded, that is, have a worse quality (see Fig. 8). Then, was not meaningful to fuse the image with more levels of decomposition because the quality of the image would be worse.

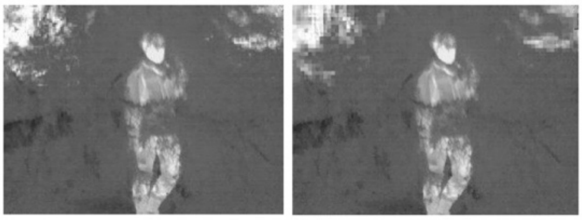

Fig. 8. a) Fused image with the 1st fusion rule and one level of decomposition b) Fused image with the 1st fusion rule and two levels of decomposition

Since the goal is to identify the camouflaged people, for qualitative analysis will be considered the images taken with the 1st rule of fusion and one level of decomposition.

E. Qualitative Analysis

In order to compare the three implemented methods there are presented four sets of three fused images from Fig. 9 to Fig. 13, wherein the first image corresponds to the method that uses the third variant of the PCA (as it was considered the most suitable) the second image refers to the method that uses the Laplacian Pyramid with four levels of decomposition and the third image corresponds to the method of Wavelets with one level of decomposition using the 1st fusion rule.

In this first example, Fig. 9, the subjective quality of the three images is very similar, second image has a lower contrast visible in the silhouette of people in the tree shade, relatively to the other two.

Fig. 9. a) Third variant of PCA b) Laplacian Pyramid with four levels of decomposition c) Wavelet with one level of decomposition and 1st fusion rule

In this second example, Fig. 10, the quality of the first two images is similar, and in the third image the person is most prominent in relation to the other two, being preferred for that reason. The detail in the three images is identical.

Fig. 10. a) Third variant of PCA b) Laplacian Pyramid with four levels of decomposition c) Wavelet with one level of decomposition and 1st fusion rule

In the third set of images, Fig. 11, the image that shows best quality is the third, which uses the method of Wavelets. Comparing the first two images, although the thermal component is more significant in the first image than in the second image, the details and contrast are lower.

Fig. 11. a) Third variant of PCA b) Laplacian Pyramid with four levels of decomposition c) Wavelet with one level of decomposition and 1st fusion rule

In the fourth image (Fig. 11), the area that corresponds to the camouflaged person is more evident and therefore preferable. Also in the third image, unlike what happens in the first two, the upper area of the image is lighter, which corresponds to the incidence of sunlight.

In Fig. 12, in any of the three images there is not much detail due to the nature of the input images. In this situation, a person is lying behind bushes, which by itself contributes to the lack detail because the bushes are sparse, and the input images appear to be a stain. The third image is where you can observe increased intensity of the thermal component, however the second image is the one that appears to have a better quality because it has a balance between detail and thermal component; despite being able to identify the person you can also see that the darker parts of the image have different shades (unlike what happens in the third image).

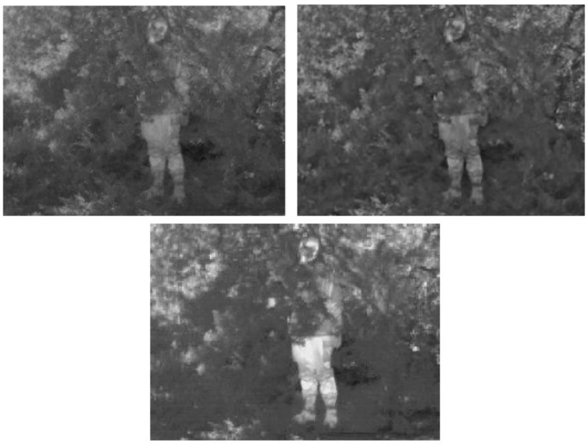

Fig. 12. a) Third variant of PCA b) Laplacian Pyramid with four levels of decomposition c) Wavelet with one level of decomposition and 1st fusion rule

Of the five presented examples, Fig. 13 is the one with the worst results. Any of the methods used is not able to identify the camouflaged person. This image consists of a person sitting behind bushes, which is not noticeable when observing the fused image. However, the image that appears to be the most outstanding thermal component is the third (obtained with Wavelets) and is therefore the most suitable for the intended purpose (to identify the camouflaged person).

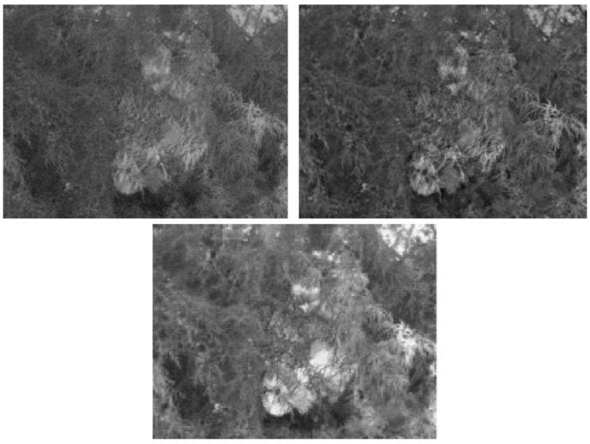

Fig. 13. a) Third variant of PCA b) Laplacian Pyramid with four levels of decomposition c) Wavelet with one level of decomposition and 1st fusion rule

Summing up the qualitative analysis from these four examples, the method that seems to have better results is one that uses Wavelets, followed by the method using the Laplacian Pyramid and finally the one that uses the PCA.

One of the factors that contributed to the poor results presented by the method using the PCA is the fact that when both of the principal component values are close to 0.5, the fusion resembles the fusion through the simple average which produces a low contrast of features.

On the other hand, the method using the Wavelets is the one with better results because it is the one that does the fusion of the various image components (approximation, horizontal, vertical and diagonal coefficients) in accordance with the established fusion rules to be the best for the intended purpose, which is to detect the camouflaged people.

F. *Quantitative Analysis*

In this section is performed a quantitative assessment for comparison of the performance of each method used. Twenty fused images were used to test each one of the performance evaluation metrics and the results are presented in boxplot figures.

In these figures, an objective comparison is made, wherein each bin corresponds to a method. The methods are designated as follows: PCA1, PCA2 and PCA3 corresponding to the three variants of the PCA; LP02 and LP04 match method using Laplacian Pyramid with two and four levels, respectively; WV11 and WV12 match the wavelet with one level of decomposition with the 1st and 2nd fusion rule, respectively; finally, WV21 and WV22 are equivalent to Wavelet with two levels of decomposition with the 1st and 2nd fusion rule.

1) *Standard Deviation*

The Standard Deviation measures the contrast of the fused image. The images obtained with the method using PCA have values close to zero, which agrees with evaluation made by visual inspection (see Fig. 14), which means they have very little contrast. This lack of contrast corresponds to a lack of detail in the fused image.

The methods using the Laplacian Pyramid and Wavelets have much higher values, and the ones that show better results (higher values) are the Wavelets with the 1st fusion rule. This method is the one with the highest values as a result of its higher contrast on the fused image, due to the marked presence of the thermal component (lighter part in the picture) and visible component (darkest part of the image).

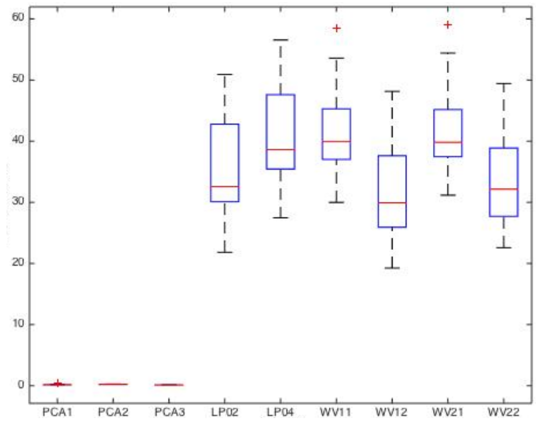

Fig. 14. Standard Deviation

After the Wavelets, the Laplacian Pyramid with four levels is one with the best results by being able to preserve the contours and reduce artifacts around. This result is in agreement with the visual assessment.

2) *Entropy*

The Entropy measures the information content of the fused image; a high entropy indicates the improvement of information content [9]. As the Standard Deviation, the Entropy of images obtained with the method based on PCA has values close to zero, which indicates the absence of relevant information.

The methods that achieves the best results are the Laplacian Pyramid with four levels, followed by Wavelets with one and two levels of decomposition using the 1^{st} fusion rule (see Fig. 15). The Laplacian Pyramid with four levels obtains the highest results because it preserves the contours of images and contributes to a better description of the fused image.

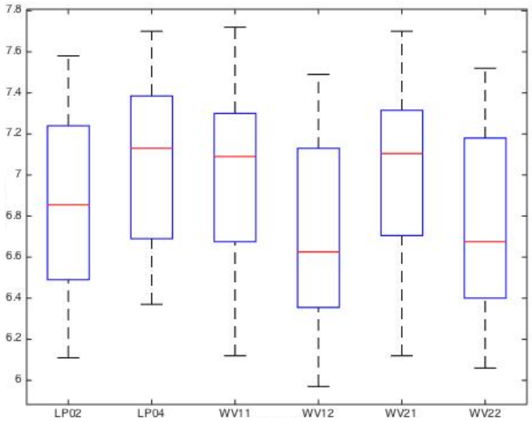

Fig. 15. Entropy

As in the previous metric, the Wavelet with two-level decomposition presents results very close to only one level of decomposition. However, as shown in the qualitative analysis, with two levels of decomposition the image is degraded (see Fig. 8). That is, although the entropy is high, the fusion is not increasing the information content, but is rather deteriorating the image through the introduction of noise [9].

The Entropy is an appropriate metric for comparing the quality of fused images, so providing a relative assessment of the implemented methods.

3) Spatial Frequency

The Spatial Frequency indicates the overall activity level of an image, that is, the greater its value more information has the image, except in cases where there is the introduction of noise in the fusion process.

Spatial Frequency of the images obtained with the method using the PCA has values close to zero. The method which achieves better results is the Laplacian Pyramid with four levels of decomposition, because it preserves the outlines of images and thus contribute with a greater detail to the fused image, followed by Wavelets with two levels of decomposition, which have a performance better than those with one decomposition level, contrary to what was expected (see Fig. 16).

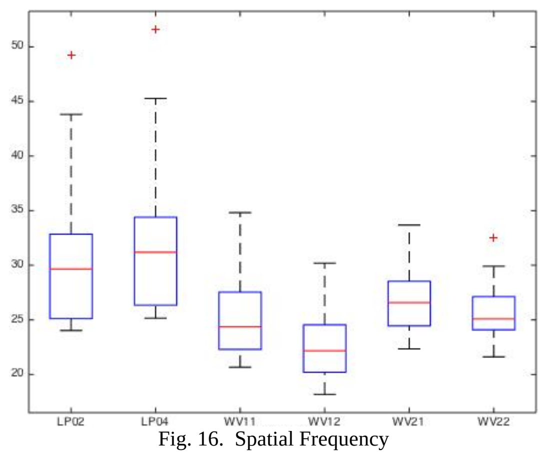

Fig. 16. Spatial Frequency

When comparing the methods using Wavelets, it can be deduced that the distortion introduced by the fusion process is the reason why wavelets with two levels of decomposition achieve better results. The poor quality of the fused image will contribute to increased Spatial Frequency [12].

4) Mutual Information

The Mutual Information measures the degree of dependence between two images. The value of the Mutual Information is the sum of the Mutual Information of each input image with the fused image, so the greater its value the greater is the dependence of the input images and the fused image.

The best results are obtained when using wavelets with the 1^{st} fusion rule, both with one and two levels of decomposition; results with PCA are the worst.

The highest values depend on the fact that the pixels with the highest intensity being selected, information transfer is increased. On the other hand, in the Laplacian Pyramid and Wavelets with the 2^{nd} fusion rule is used the average of the pixels, and therefore the mutual information values are lower.

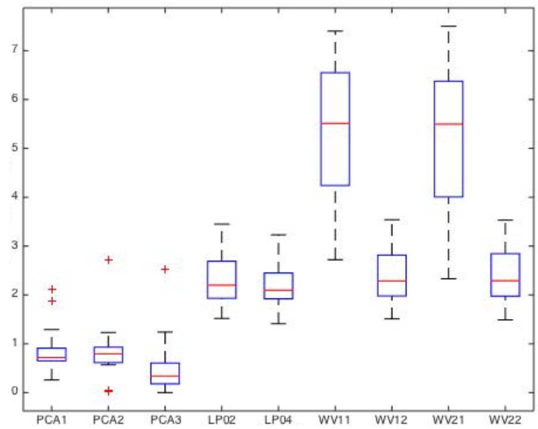

Fig. 17. Mutual Information

5) Fusion Quality Index

The Fusion Quality Index measures the similarity between the fused image and the input images. It takes values between zero and one, and the higher the similarity the nearest to one is the value.

The best results are obtained when using the method based on PCA, followed by the method that uses the Laplacian Pyramid (see Fig. 18).

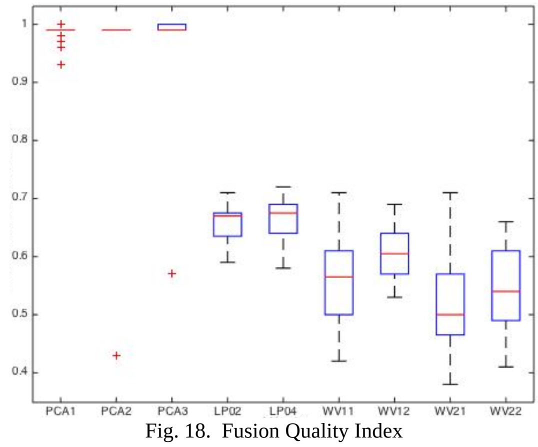

Fig. 18. Fusion Quality Index

Contrary to what has been observed in the previous metrics, methods based on PCA achieves the highest values. This can be explained by the fact that in this method, a weighted average of the input images is made for the fused image and thus both input images contribute to the final image in the same way. That is, while in methods that use Wavelets there are coefficients that are selected from only one of the input images (having detail coefficients of one of the input images that do not contribute to the fused image), in methods based on PCA that does not happen.

This metric is considered to be the most reliable for fused images that do not have reference image [5]. However, the results for PCA methods are misleading.

6) Structural Similarity Index

The Structural Similarity Index measures the similarity of the structure of information on the images to be compared; is the sum of the Structural Similarity Index of each input image with the fused image. The greater the value obtained, the higher the similarity between images.

The values obtained for the three variants of the PCA were close to zero.

As can be seen in Fig. 19, unlike what happened so far, the Laplacian Pyramid with two levels obtained better results, as well as Wavelets which used the 2nd fusion rule.

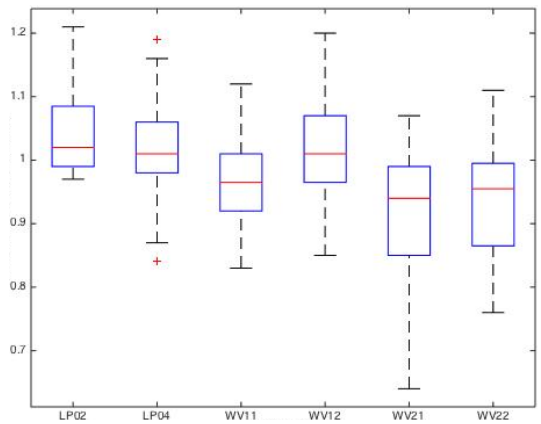

Fig. 19. Structural Similarity Index

These results are explained by the lower level of decomposition of Laplacian Pyramid which makes the fused image more similar to input images. And as the 2nd fusion rule of the Wavelets uses the mean of the approximation coefficients the fused image will also be closer to input images.

V. CONCLUSIONS

This work arises from the need to implement a method enabling the detection of camouflage people while they are dissimulated in their environment, making their camouflage inefficient. The objective is to implement pixel level image fusion methods, which intend to merge visible images with thermal images getting a richer fused image. There were implemented three methods: PCA, Laplacian Pyramid and Wavelets.

It was done a qualitative analysis of the fused images based on four sets of three images (third variant of the PCA, Laplacian Pyramid with four levels and Wavelets with one decomposition level and the 1st fusion rule). From this analysis, it is concluded that the method which achieves better results uses Wavelets followed by the method using Laplacian Pyramid and finally using the PCA.

A quantitative analysis was performed using six performance metrics: Standard Deviation, Entropy, Spatial Frequency, Mutual Information, Fusion Quality Index and Structural Similarity Index. This analysis was done based on 20 images obtained for each of the methods and their variants, making a total of nine methods. The results obtained for each of the metrics are shown in boxplot graphs, which provides a good insight of the set of results. By observing the graphs obtained, the values for the three variants of the PCA method are those that stand out the negative, because they are well below of those obtained for the remaining methods, which only with a visual inspection was not noticeable. The values obtained for the methods using the Laplacian Pyramid and Wavelets are within the ranges obtained by other authors, and support the conclusions drawn from the qualitative analysis.

Of the six performance metrics implemented, it is concluded that the Standard Deviation, Entropy and Spatial Frequency metrics are suitable for making a relative comparison among implemented methods (for measuring the quality of the fused image), however, these metrics do not take into account the relationship between the fused image and the input images. Thus, the Mutual Information, Fusion Quality Index and the Structural Similarity Index take into account this relationship and therefore are considered the most appropriate metrics to measure the quality of image fusion.

Comparing both qualitative and quantitative results, it is noted that the methods considered as the best in qualitative assessment are in fact the best quantitative results, that is, the method that uses the Wavelets with one decomposition level with the 1st fusion rule and the method using the Laplacian Pyramid with four levels. Despite of the qualitative analysis provides a good evaluation of methods, it is required a quantitative analysis that supports the conclusions drawn by visual inspection, and that allows us to identify the most appropriate method for the intended purpose.

VI. FUTURE WORK

The work was effective in detecting camouflaged people in most cases. However, there were pictures where it did not happen clearly and fused images have little information. Thus, one of the possibilities for future work is to do image fusion at feature level, and then at the decision level.

Feature level image fusion is the next level of processing where image fusion can take place. Fusion at this level requires extraction of features of the input images. Feature level methods have the ability to produce images with better subjective quality. The most common algorithms for image fusion at this level include the methods of edges detection and classifiers.

Decision level image fusion methods are the highest level of processing where image fusion can be performed. The fusion at this level takes feature level methods a step further by declaring identities to recognized objects in the individual input images. Some common algorithms used in decision level methods include Fuzzy Logic, Fusion-based rules and Bayesian Networks.

ACKNOWLEDGMENTS

The author would like to thank to her family, namely her parents and sister for all the patient and support. Also would like to thank to Professor José Bioucas-Dias and Professor José Silva for all the dedication and support through the elaboration of this work.

REFERENCES

[1] Aparna Akula, Ripul Ghosh, and H K Sardana, "Thermal Imaging And Its Application In Defence Systems," in *Optics: Phenomena, Materials, Devices, and Characterization*, 2011, pp. 333-335.

[2] 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.

[3] Yufeng Zheng, "Multi-scale Fusion Algorithm Comparisons: Pyramid, DWT and Iterative DWT," in *12th International Conference on Information Fusion*, Seatle, 2009, pp. 1060-1067.

[4] 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 Autonoma de Barcelona, Barcelona, 1577-5097, 2011.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

Nádia Alexandra Ferreira Bento was born in Cascais, Portugal, on May 30, 1991. She lived there until she graduated from high school.

She joined the Military Academy in September of 2009 at the age of eighteen.

In the Military Academy she concluded the first four years of a seven-year course, and then she attended two years of Electrical and Computer Engineering in Instituto Superior Técnico (Lisbon).

She is currently attending her final year of the course in Mafra, after which she will become a Signals Officer of the Portuguese Army.