

Enhancement of Sight Effectiveness by Dual Infrared System: Evaluation of Image Fusion Strategies

G. Corsini , M. Diani , A. Masini
Dipartimento di Ingegneria dell'Informazione
Università di Pisa , Italy
g.corsini@iet.unipi.it, m.diani@iet.unipi.it,
andrea.masini@iet.unipi.it

M. Cavallini
Galileo Avionica
Campi Bisenzio, Firenze, Italy
marco.cavallini@galileoavionica.it

Abstract - The problem of objective evaluation of multi-sensor image fusion strategies is analysed for the design of a dual infrared system. Such a system should be used to enhance the sight effectiveness in assisting a driver or a pilot in bad visibility conditions. Two *no-reference* indexes are used to quantify the performance of different image fusion methods. Numerical results are presented and discussed in terms of the quality of the fused images.

Index Terms – *Image/video processing. Data fusion. Image quality indexes.*

I. INTRODUCTION

Image fusion methods combine the visual information contained into different source images in a single composite one to enhance human perception and interpretation capabilities.

Such a tool is of paramount importance in many applications, where it is crucial to extend human vision to improve the operator performances. An interesting application can be easily found in the field of vehicle driving. In fact, the use of different infrared sensors can be very useful for the perception of objects or obstacles in bad visibility conditions (night, rain and fog) and, consequently, can aid a car driver or aircraft pilot. Multispectral infrared image sources have to be fused to synthesize the salient information collected in the different channels enabling a better scene interpretation and improving the situational awareness.

As a matter of fact, Image Fusion (IF) techniques are expected to achieve several objectives which can be summarized as follows:

a) integration of images from different sensors has to produce information that cannot be obtained by viewing the sensor outputs separately and consecutively; b) the information extracted from the input images must be salient with respect to the specific application and must improve the image semantic interpretation. Obviously, fusion methods should not discard any salient information from each source; c) an essential problem in merging images is *pattern conservation*: important details of the component images must be preserved in the resulting composite image. Therefore, the incomplete representation of objects in one image may be integrated by information from the other one (complementary information); d) the fusion process should not introduce any artefacts which

can distract or mislead a human observer; e) the merging operation shall harmonise the disparity between the images coming from the input sensors. For example, the sensor output images could not be equally reliable. Such disparities have to be taken into account when fusing the information from such sources; g) the fusion must be reliable, robust and has to have the capability to tolerate disturbances and errors (noise and misregistration); h) common but contrast reversal information must be treated in an appropriate way: there could be various objects and regions that occur in both images but with opposite contrast. Therefore, in this case, the direct approach of adding and averaging the source images is not satisfactory.

A fundamental issue of image fusion techniques is the process for evaluating the performances of a fusion scheme. In fact, the improvement depends on the particular scenario, the used sensors, the lighting conditions and, obviously, on the capabilities of the human observer. Then, it is very difficult to define general procedures to compare fusion results.

Traditionally, the quality of video sequences is evaluated subjectively by an appropriate number of human evaluators. This method has two main disadvantages: it requires an appropriate number of evaluators (thus it is time consuming and expensive), and it cannot be done in real time.

As a result, a considerable research effort has been addressed to the development of automatic objective methods for video quality measurement. Performance measures are essential for various reasons: 1) to ascertain the possible benefits of fusion; 2) to compare results obtained with different algorithms; 3) to obtain an optimal setting of parameters for tuning a specific fusion algorithm. A good quality index should extract all the important information from a perceptive point of view from the input images and measure the ability of the fusion process in transferring with the highest accuracy (that is minimising the number of artefacts or the amount of distortions) this information into the final image.

In this paper we propose some figures of merit for the evaluation and the comparison of fusion strategies in a dual infrared system, used to enhance the sight effectiveness in assisting a driver or a pilot in bad visibility conditions. Namely, we consider the two figures of merit, recently proposed in the literature by Xydeas and Petrovic [1] and

Wang-Bovik [2] and discuss their capability of assessing the performance of different fusion strategies applied to an experimental data set.

We refer to an experiment where, for the first time, two infrared cameras have been used in a fusion system in order to improve the car driving and the aircraft piloting. The first camera operates in the SWIR bands (0.8-2 μm), and the second in the LWIR (7-14 μm). These bands have been selected because of their visual complementary factor. In fact, the images resulting from the SWIR sensor show the overall details of the scene, while the LWIR shows the background details. The SWIR camera has been chosen also by reason of the good visibility conditions through the fog, and the LWIR camera has been considered for the further development of an un-cooled camera in these bands.

The paper is organised as follows. In section II we review the figures of merit used to assess the performance of the image fusion process. In section III we discuss the results obtained on the experimental data set and show the agreement between the numerical values of the indexes and the results of a perceptual experiment conducted by different human evaluators on the same data set. Finally, the work is concluded in section IV.

II. EVALUATION OF THE IMAGE FUSION PROCESS

An *objective* image quality index can automatically predict perceived image quality and can play an important role in a broad range of applications. It can be used for image fusion to examine the quality of the output image in order to control and aid a human operator in his decisions. It can assist the parameters' setting of image fusion systems in the choice of multiscale decomposition algorithm and in the choice of a pixel-based or region-based approach.

A quality index can also be used to compare different image processing systems and algorithms.

Objective image quality indexes can be classified according to the availability of a target image to be compared with the processed image. According to the Bovik nomenclature, if a complete reference image (target) is known, the approaches are called *full-reference*. These approaches are the most recurrent in the literature. When a complete reference image is not available, the approaches are called *no-reference*, and if the target is partially known the approaches are called *reduced-reference*.

Before proceeding to the fusion of two images it is necessary to ascertain if the fusion action is significant in some sense. It is obvious that the two images must have visual complementary informative contents or represent different objects. The definition of entropy and mutual information permits the measurement of the informative content.

In the literature, two objective fusion performance indexes have been proposed where the knowledge of ideal fused image is not assumed. Herein, these two objective no-reference figures of merit for fused images are

presented: they utilize local measures to estimate the level of the salient information transferred from the input image into the fused one.

The first figure of merit is based on Xydeas and Petrovic index and the second one on an image quality index recently introduced by Wang and Bovik [4]. The Xydeas-Petrovic approach is based on the observation that the human visual system is particularly sensitive to the edges in the image. Therefore, from the evaluation of the quantity of information associated to the edges which is transferred from the input images to the fused one, it is possible to obtain an index that measures the performance of the fusion method. In the method proposed by Xydeas and Petrovic [1], first the edge information is extracted from the input images by the application of the Sobel edge operator, then the edge strength and orientation are calculated. These features are subsequently used for the index evaluation. The edge strength $g_A[i, j]$ and orientation $\alpha_A[i, j]$ are defined as:

$$\alpha_A[i, j] = \tan^{-1} \left(\frac{s_A^y[i, j]}{s_A^x[i, j]} \right) \quad (2.1)$$

$$g_A[i, j] = \sqrt{s_A^x[i, j]^2 + s_A^y[i, j]^2} \quad (2.2)$$

where $s_A^x[i, j]$ and $s_A^y[i, j]$ are the output images obtained from the horizontal and vertical Sobel operators applied to the input image A . The same operations are applied to the image B and to the fused image F . Then, we calculate the relative values of strength $G^{AF}[i, j]$ and orientation $A^{AF}[i, j]$ of the input image A with respect to F as:

$$G^{AF}[i, j] = \begin{cases} \frac{g_F[i, j]}{g_A[i, j]} & \text{if } g_A[i, j] \geq g_F[i, j] \\ \frac{g_A[i, j]}{g_F[i, j]} & \text{elsewhere} \end{cases} \quad (2.3)$$

$$A^{AF}[i, j] = 1 - \frac{|\alpha_A[i, j] - \alpha_F[i, j]|}{\frac{\pi}{2}} \quad (2.4)$$

These values are used to derive the edge strength and orientation of the input image (A), which are preserved in the fused one:

$$Q_g^{AF}[i, j] = \frac{\Gamma_g}{1 + e^{k_g(G^{AF}[i, j] - \sigma_g)}} \quad (2.5)$$

$$Q_\alpha^{AF}[i, j] = \frac{\Gamma_\alpha}{1 + e^{k_\alpha(A^{AF}[i, j] - \sigma_\alpha)}} \quad (2.6)$$

$Q_g^{AF}[i, j]$ and $Q_\alpha^{AF}[i, j]$ model the loss of information in the fused image F , in terms of strength and orientation, that could be perceived by human operators.

The parameters Γ_g , k_g , σ_g , Γ_α , k_α , σ_α determine the exact shape of the sigmoid functions used in (2.5) and (2.6). In [1], Xydeas and Petrovic suggest the best values for these parameters.

These values have been obtained from the subjective scores assigned by a proper number of human evaluators. Finally, the edge information preservation function is defined as:

$$Q^{AF}[i,j] = Q_g^{AF}[i,j] \cdot Q_\alpha^{AF}[i,j] \quad (2.7)$$

Obviously, Q^{AF} assumes values in the interval $[0,1]$. Q^{AF} equal to zero corresponds to the complete loss of edge information, at $[i,j]$ location, i.e. no information is transferred from A to F in that location. Q^{AF} equal to one indicates a full transfer of edge information from A to F at $[i,j]$ location.

Once that $Q^{AF}[i,j]$ and $Q^{BF}[i,j]$ have been evaluated, the performance metric $Q_p^{AB/F}$, for the fused image F of size $M \times N$ is obtained by computing the weighted mean over the full image as follows:

$$Q_p^{AB/F} = \frac{\sum_{i=1}^M \sum_{j=1}^N Q^{AF}[i,j] \cdot \omega^A[i,j] + Q^{BF}[i,j] \cdot \omega^B[i,j]}{\sum_{i=1}^M \sum_{j=1}^N \omega^A[i,j] + \omega^B[i,j]} \quad (2.8)$$

The edge preservation functions $Q^{AF}[i,j]$ and $Q^{BF}[i,j]$ are weighted with the coefficients $\omega^A[i,j]$ and $\omega^B[i,j]$, respectively. Since a pixel with high edge strength should transfer more edge information to the fused image than one of relatively low edge strength, the coefficients $\omega^A[i,j]$ and $\omega^B[i,j]$ are given by:

$$\omega^A[i,j] = (g^A[i,j])^P \quad \text{and} \quad \omega^B[i,j] = (g^B[i,j])^P$$

where P is a constant, and therefore $0 \leq Q_p^{AB/F} \leq 1$.

A general quality index is proposed in [3], [4], where Wang and Bovik construct a structural similarity quality measure by considering an image formation model. Following their approach two images are compared by using three parameters: luminance, contrast and structure. Given an $M \times N$ image X and its reference image Y , the quality index proposed by Wang, Bovik [3] is calculated as:

$$Q(X,Y) = \frac{(2\bar{X}\bar{Y} + C_1) \cdot (2\sigma_{XY} + C_2)}{(\sigma_X^2 + \sigma_Y^2 + C_2) \cdot (\bar{X}^2 + \bar{Y}^2 + C_1)} \quad (2.9)$$

where:

$$C_1 = (k_1 \cdot L)^2 \quad C_2 = (k_2 \cdot L)^2$$

\bar{X} and \bar{Y} indicate the means of the two images, σ_X and σ_Y are the standard deviations of X and Y , respectively, σ_{XY} represents the covariance between the two images, L is the dynamic range for the image pixel values, $k_1 \ll 1$ and $k_2 \ll 1$ are two constants that are chosen equal to 0.01 and

0.03, respectively. These values are somewhat arbitrary, but in our experiments it has been noted that the quality index is fairly insensitive to variations of k_1 and k_2 . C_1 and C_2 are introduced in order to stabilize the measure because the denominator approaches zero in the flat regions.

Note that $-1 \leq Q \leq 1$. The value one is achieved when the two images X and Y are the same.

This index can be written as a three factor product:

$$Q(X,Y) = f(l(X,Y), c(X,Y), s(X,Y)) = \frac{\sigma_{XY} + \frac{C_2}{2}}{\sigma_X \sigma_Y + \frac{C_2}{2}} \cdot \frac{2\bar{X}\bar{Y} + C_1}{\bar{X}^2 + \bar{Y}^2 + C_1} \cdot \frac{2\sigma_X \sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (2.10)$$

The first component represents the correlation coefficient between X and Y and measures the degree of correlation between the two images; it varies between -1 and 1. The highest value is obtained when X is a scaled and shifted version of Y . A distortion between X and Y is measured through the second and third index components. The second component, that varies between zero and one, measures the similarity between the mean luminance values of the two images; the third component measures the contrast similarity and varies between zero and one.

An important point is that the three components are relatively independent.

For image quality assessment, it is useful to apply the quality index locally rather than globally. Image statistical features are usually highly spatially non-stationary. Image distortions, which may depend on the local image statistics, may also be space-variant. At typical viewing distances, only a local area in the image can be perceived with high resolution by the human observer at one time instance (cause of the foveal character of the HVS).

Finally, localized quality measurements can provide a spatially varying quality-map, which delivers more information about the quality degradation of the image and may be useful in some applications.

It is more appropriate to calculate this quality index locally and then to combine the results of different measures together. In order to evaluate the quality index on regions, an 8x8 moving window has been used.

The local statistics \bar{X} , σ_X and σ_{XY} are computed using the data collected by the local 8x8 moving window.

To measure the overall image quality the mean quality index has been computed as follows:

$$Q(X,Y) = \frac{1}{M} \sum_{i=1}^M Q_i(X_i, Y_i) \quad (2.11)$$

where X and Y are the reference and the distorted images, respectively; X_i and Y_i are the image contents at the i -th local window, and M is the number of samples in the quality-map.

This figure of merit must be modified to evaluate image fusion methods. As a first requirement it is desirable

to remove the dependence of the method from the selection of a reference image. Then, it is necessary to adapt the index to the fusion evaluation problem. A modified version of the Wang-Bovik index has been proposed by G. Piella in [5]. In order to define the figure of merit a local weight λ_i is assumed. λ_i varies between 0 and 1 and it indicates the relative importance of the image A with respect to B . Therefore, the larger λ_i is, the more importance will be given to the image A in the index calculation, and vice versa [5].

$$\lambda_i = \frac{s_i(A)}{s_i(A) + s_i(B)} \quad (2.12)$$

where $s_i(A)$ and $s_i(B)$ are the features of interest extracted from the input images inside the i -th window.

As a possible feature, related to the edge of the images, the one that is obtained by applying the Laplacian operator can be considered.

The quality index to compare the result of the image fusion is given by :

$$Q(AB, F) = \frac{1}{M} \sum_{i=1}^M [\lambda_i \cdot Q_i(A, F) + (1 - \lambda_i) \cdot Q_i(B, F)] \quad (2.13)$$

where $Q_i(\cdot, \cdot)$ is the Wang-Bovik quality index for a couple of images as defined in (2.10).

Many simulations have been performed which showed that this index is in compliance with subjective evaluations, and can therefore be used to compare different image fusion methods.

III. NUMERICAL RESULTS

In this section the performances of various fusion techniques are discussed by computing the figures of merit described in section II. This analysis allows a rational evaluation of the implemented fusion techniques. The fusion techniques, here considered, are based on multi-resolution or multi-scale source image decomposition and are: Laplacian pyramid [6,11], FSD (Filter-Subtract-Decimate) pyramid [7], RoLP (Ratio of Low Pass) pyramid [8,12,13], Gradient transform [9], Morphological Decomposition through low pass morphological filter [15], Discrete Wavelet Transform [14,16], and Region Fusion based on Histogram Segmentation (RFHS) [10].

A database of images has been utilized in this work to test the indexes discussed in this paper.

The images (240 rows by 320 columns pixels) represent a real scenario, and are taken from the system developed in the SEE project. This system is composed of a SWIR (Short Wave IR, 0.8-2 μm) camera and a LWIR (Long Wave IR, 7-14 μm) camera, and is mounted on a car. The two cameras are developed by Zeiss in the framework of the same project. The collected frames give a view of the road-scenario. An example of two database images, to be fused, is shown in Fig. 1.

All the images, on which the fusion algorithms are tested, are interesting for automotive applications.

In Table 1, the entropy and the mutual information, for two considered sources, are reported. As discussed, the mutual information gives a measure of the similarity between the two images to be fused.

The MI value in the table shows that the two images are different and a fusion process can integrate the full information into a single image.



Fig. 1 Source images:
a) LWIR image, b) SWIR image

TABLE 1.

Entropy, and mutual information of the two images in Fig.1

Entropy LWIR	5.71
Entropy SWIR	5.85
Mutual Information	1.01

The evaluated quality indexes, for the example images presented in Fig. 2, are summarized in Table 2 and are consistent with the result of the other database images. It is possible to note that the Bovik index spans from 0.666 to 0.756, assumed respectively by the RoLP fusion technique and by the Laplacian one. Good results are given by Gradient (0.736) and FSD (0.745) fusion techniques. As already discussed, the Bovik index measures the luminance and the contrast similarity between the input and the fused images. The observation of the images in Fig. 2, shows that the index value reflects the visual perception of the images, in fact, the image fused by using the Morphological Decomposition technique appears darker with low contrast, while the FSD, Gradient and Laplacian fused images have a good luminance and contrast.

The Xydeas and Petrovic index spans from 0.491 to 0.615, assumed by RoLP and the Laplacian techniques, respectively. Good results are given by the RFHS (0.568), Morphological (0.587) and Gradient (0.567) techniques. In the previous section it has been observed that the Xydeas and Petrovic index evaluates the quantity of information associated to the edges which is transferred from the input images to the fused one.

Once again, Fig. 2 confirms that visual perception is in accordance with the index values. In fact, in the image fused by using the RoLP technique the edges are scarcely perceivable, while in the images merged by the fused region based on histogram, the Gradient and the Laplacian

techniques, the edges are more evident and the objects in the scene are easily identifiable.

A global evaluation of the two indexes shows that the algorithms which give good performance are the Laplacian, the Gradient, and the RFHS. On the contrary, the algorithms which give the worst results are the RoLP, and DWT fusion techniques.

In particular, considering both indexes, the best technique is the Laplacian one. The visual impression, obtained by observing the images in Fig. 2, matches the global index evaluation, in fact, the images fused with the RFHS, Laplacian and Gradient techniques have a good contrast and luminance, and it is possible to discriminate the object shape in the scene.

To validate the indexes' results, the database images have been analysed by three human experts and two experts of infrared imaging in the Risø National Laboratory, Denmark. The applied method is the following: each person was asked to analyse on his own the images from the point of view of being a driver guided by such images. The criteria were to evaluate the quality of the image details, such as traffic signs, road signs, people, other vehicles, houses, trees, etc. This may include contrast, light intensity, and possible artefacts/noise. Each set of images was presented on a screen simultaneously in order to facilitate their comparison.

The fusion methods analysed were the following: Discrete Wavelet Transform, FSD, Gradient, Histogram Area, Laplacian, and RoLP fusion techniques. The morphological fusion method was not analysed because, for the most scenarios, it clearly gives the worst visual result. This study agrees with the indexes' results. In fact, the Laplacian, the Gradient and the Histogram Area fusion techniques were judged as the best ones. On the contrary, the other fusion techniques seem to have worse quality than the above-mentioned criteria.

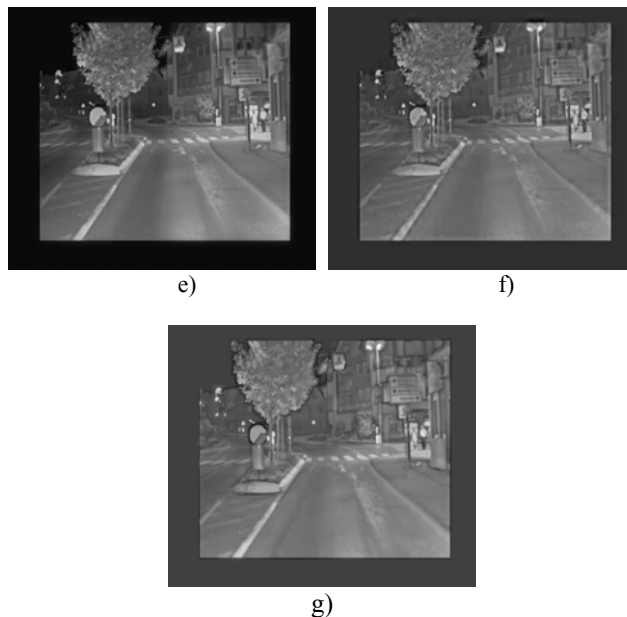
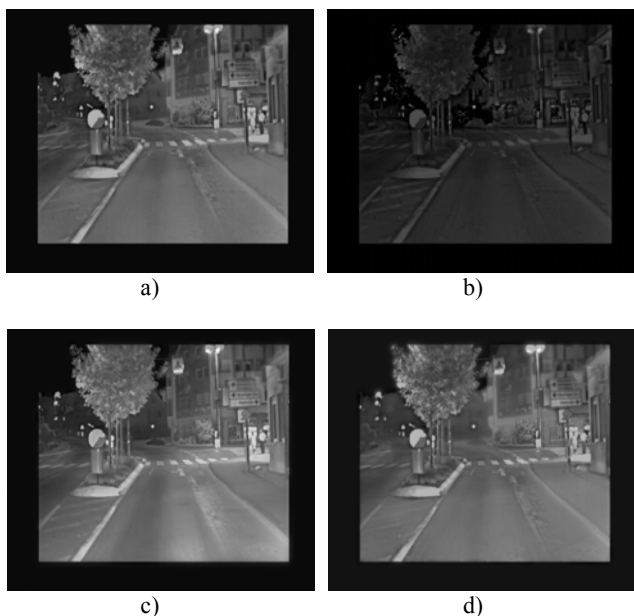


Fig. 2 Fused images:
a) Laplacian, b) RoLP, c) Gradient, d) RFHS, e) FSD, f) DWT, g) morphological.

TABLE 2.
Quality indexes

	Bovik index	Petrovic index
Laplacian	0.756	0.615
Gradient	0.736	0.567
RFHS	0.719	0.568
FSD	0.745	0.552
Morphologica l	0.677	0.587
DWT	0.693	0.567
RoLP	0.666	0.491

IV. CONCLUSION

In this paper two *no-reference* methods for the comparison and the analysis of different image fusion algorithms are applied to study the effectiveness of the fusion process in the condition of two single image sources in the SWIR and LWIR bands. This case is directly referred to a prototypal dual infrared system to enhance the sight effectiveness in assisting a driver in bad visibility conditions. Many multi-scale fusion methods have been analysed to determine, in this particular case, the algorithms that perform better.

The analysis shows that the best performances are reached by the Laplacian, Region Fusion based on Histogram Segmentation, and Gradient techniques, while RoLP, and DWT fusion techniques are poorer. These results are in compliance with the perceptual experiment conducted by the Risø National Laboratory, Denmark. It is worth noting that these results do not necessarily imply that

the fusion strategies in the first group obtain the best performance in all the scenarios. So, the use of figures of merit represents a low cost and effective tool to select the appropriate fusion method.

ACKNOWLEDGEMENT

This work is funded by the European Commission in the framework of the 5th PCRD Contract IST-2001-38228 SEE (Sight Effectiveness Enhancement) Project.

REFERENCES

- [1] C. Xydeas, V. Petrovic, "Objective pixel-level image fusion performance measure," IEE Electronics Letters, vol. 36, no. 4, pp. 308-309, 17th February 2000.
- [2] Z. Wang, A.C. Bovik, "A universal image quality index," IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002.
- [3] Z. Wang, A.C. Bovik, "Image Quality Assessment: From Error Measurement to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, no. 1, pp. 600-612, January 2004.
- [4] Z. Wang, L. Lu, A.C. Bovik, "Video quality assessment based on structural distortion measurement," Signal processing: Image communication, vol. 9, no.1, pp. 121-132, January 2004.
- [5] G. Piella, H. Heijmans, "A new quality metric for image fusion," http://homepages.cwi.nl/~henkh/publications/icip03_fusion.pdf, 2003.
- [6] P. J. Burt and E. H Adelson., "The Laplacian Pyramid as Compact Image Code," IEEE Trans. on Com., vol. COM-31, Apr. 1983.
- [7] C. H. Anderson, "A Filter-Subtract-Decimate Hierarchical Pyramid Signal Analysing and Synthesising Technique," United States Patent 4, 718, 104, 1987.
- [8] A. Toet., L. J. Van Ruyven and J. M. Valetton, "Merging thermal and visual images by a contrast pyramid," Optical Engineering, , vol. 28, n. 7, pp. 789-792, July 1989.
- [9] P. J. Burt, "A gradient pyramid basis for pattern selective image fusion," in Proceedings of the Society for Information Display Conference (1992).
- [10]G. Piella, "A General Frameworks of Mutiresolution Image Fusion: From Pixels to Regions," PNA Report, Centrum voor Wiskunde en Informatica.
- [11]P.J. Burt and R.J. Kolczinski,"Enhanced image capture through fusion," in Proceeding of the 4th International Conference on Computer Vision, Berlin, Germany, pp. 173-182, May 1993.
- [12]A. Toet, "Image fusion by a ratio of low-pass pyramid," Pattern Recognition 9, pp. 245-253, 1989.
- [13]A. Toet, "Multiscale contrast enhancement with application to image fusion," Optical Engineering 31, pp. 1026-1031, May 1992.
- [14]O. Rockinger, "Pixel-level fusion of image sequences using wavelet frames," in Proceeding of the 16th Leeds Applied Shape Research Workshop 1996, Leeds University Press.
- [15]A. Toet, "Hierarchical image fusion," in: Machine Vision and Applications, Vol. 3, No. 1, pp. 1-11, 1990.
- [16]O. Rockinger, "Image sequence fusion using a shift invariant wavelet transform," in: Proc. IEEE Intl. Conference on Image Processing, 1997, pp. 288-291