In: I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XII*,
SPIE-5096 (2003), pp. 552-561. Bellingham, WA., USA: The International Society for Optical Engineering.

Performance Comparison of Different Graylevel Image Fusion Schemes through a Universal Image Quality Index

Alexander Toet* Maarten A. Hogervorst
TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands

ABSTRACT

We applied a recently introduced universal image quality index Q that quantifies the distortion of a processed image relative to its original version, to assess the performance of different graylevel image fusion schemes. The method is as follows. First, we adopt an original test image as the reference image. Second, we produce several distorted versions of this reference image. The distortions in the individual images are complementary, meaning that the same distortion should not occur at the same location in all images (it should be absent in at least one image). Thus, the information content of the overall set of distorted images should equal the information content of the original test image. Third, we apply the image fusion process to the set of distorted images. Fourth, we quantify the similarity of the fused image to the reference image by computing the universal image quality index Q. The method can also be used to optimize image fusion schemes for different types of distortions, by maximizing Q through repeated application of steps two and three for different parameter settings of the fusion scheme.

Keywords: Image fusion, image quality

1. INTRODUCTION

The increasing availability and the dramatic cost-reduction of co-registered multimodal imagery from different types of sensors has spurred the development of techniques for image fusion[1,9,13,21-23,25,28,29,31]. The goal of image fusion is to represent the visual information present in any number of input images in a single fused image, without the introduction of distortion and artifacts or loss of information.

The results of image fusion schemes are usually evaluated visually. Quantitatively assessing the performance is a complicated issue because the ideal composite image is normally not available[11,14,26,30]. One possible approach is to generate sets of distorted source images from a known reference image, and compare the fused image with the original reference image[9,15].

Most image fusion schemes currently in use take a hierarchical approach. They first decompose the input images into meaningful details over a range of spatial scales. Then they select the visually most relevant details at each spatial scale from the set of (decomposed) input images. Finally they construct a composite or fused image from the set of selected details. The result of hierarchical image fusion schemes depends on the number of levels (spatial scales) that are used in the image description. The minimal number of levels required to obtain a perceptually appreciable result depends on *(1)* the size of the relevant image details, *(2)* the nature of the multiresolution description, *(3)* the type of filters used in the analysis and synthesis phases of the fusion process, and *(4)* the position and orientation of the details in the source images[16].

2. THE UNIVERSAL IMAGE QUALITY INDEX

Wang[27] recently introduced a universal image quality index Q that quantifies the distortion of a processed image relative to its original version. The quality index correlates with the subjective evaluations of human observers for a wide variety

* E-mail: toet@tm.tno.nl

of distortions. It is defined as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Let $x = \{ x_i \mid i = 1, 2, \cdots, N \}$ and $y = \{ y_i \mid i = 1, 2, \cdots, N \}$ be the original and processed image signals respectively. The quality index Q is then given by:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2 \bar{x} \bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2 \sigma_x \sigma_y}{\sigma^2{}_x + \sigma^2{}_y} \tag{1}$$

where

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad , \quad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\sigma^2{}_x = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \quad , \quad \sigma^2{}_y = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

The dynamic range of Q is [-1,1]. The maximal value 1 only occurs when both images are identical, i.e. $y_i = x_i$ for all $i = 1, 2, \cdots, N$. The minimal value -1 occurs when $y_i = 2\bar{x} - x_i$ for all $i = 1, 2, \cdots, N$. The first component in Equation (1) is the correlation coefficient between x and y, which measures the degree of linear correlation between both images, and has a dynamic range of [-1,1]. The maximal value 1 is obtained when $y_i = a x_i + b$ for all $i = 1, 2, \cdots, N$, where a and b are constants and $a > 0$. Even if x and y are linearly related, there may still occur relative distortions between them. These are evaluated in the second and third components. The second component measures how close the mean luminance both images are, and ranges between [0,1]. It equals 1 when $\bar{x} = \bar{y}$. Since σ_x and σ_y can be regarded as estimates of the contrast of x and y, the third component measures how similar the contrasts of both images are. It also ranges from [0,1]. The highest value 1 is obtained if and only if $\sigma_x = \sigma_y$.

In practice we usually want to characterize an entire image using a single overall image quality measure. However, image quality is often spatially variant, meaning that different image regions may have different types of distortions. It is therefore more appropriate to measure statistical properties locally and combine them into a single measure. Following Wang[27] we therefore compute the image quality index Q over local image regions using a sliding window approach. Starting from the top-left corner of the image, a sliding window of size 8x8 moves pixel by pixel horizontally and vertically through all the rows and columns of the image until the bottom-right corner is reached. At the j^{th} step in this procedure the local quality index Q_j is computed over the area of the 8x8 sliding window. If the total number of steps is equal to *M*, the overall image quality index is given by

$$Q = \frac{1}{M} \sum_{j=1}^{M} Q_j \tag{2}$$

3. IMAGE FUSION METHODS

In this section we give a concise overview of the different grayscale image fusion methods we investigated in this study.

The most straightforward way to fuse graylevel images is to take a (weighted) pixelwise average of the source images. Simply taking the mean of the source images significantly reduces the contrast of the details represented in the individual source images. A better approach is to determine the optimal weighting coefficients by a principal component analysis (PCA) of all input values[7,12,18]. In that case the weighting coefficients correspond to the eigenvector of the covariance matrix of the input values that has the largest eigenvalue.

Hierarchical or multi-scale image fusion algorithms are widely used in image fusion, since they can preserve the perceptually relevant contrast details from the original source images and they do not introduce many artifacts in the fusion process. These methods first decompose the source images in contrast details at different spatial scales. Then they select the perceptually most relevant (combination of) details from these individual decompositions and use these to construct a multi-scale representation for the fused image. Finally, they reconstruct the fused image from its multi-scale representation. The multi-scale image fusion techniques we investigated in this study are respectively: the Difference of Low-Pass (DoLP) or Laplacian pyramid[4], the Ratio of Low-Pass pyramid[21], the contrast pyramid[24], the filter-subtract-decimate Laplacian pyramid[2,6], the gradient pyramid[3,6], the morphological pyramid[20], the discrete wavelet transform[8-10,19], and a shift invariant extension of the discrete wavelet transform[8,15-17]. Each of these methods will be discussed briefly in the folowing sections.

3.1. Laplacian pyramid fusion

An image pyramid is a collection of images at different spatial scales that together represent the original source image. Such a multi-resolution image representation can be obtained through a recursive *reduction* of the input image, i.e. a combination of low-pass or band-pass filtering and decimation. The popular Laplacian pyramid (also known as Difference of Low-Pass pyramid), introduced by Burt and Adelson[5], is a sequence of images in which each image is Laplacian filtered and subsampled copy of its predecessor. The construction of this pyramid is as follows. First, a Gaussian or low-pass pyramid is constructed. The original image is adopted as the bottom or zero-level of the Gaussian pyramid. Each node of pyramid level i ($1 \le i \le N$, where N is the index of the top level of the pyramid, i.e. the lowest resolution level) is obtained as a (Gaussian) weighted average of the nodes at level $i-1$. Because of the reduction in spatial frequency content each image in the sequence can be represented by an array that has only half the dimensions of its predecessor. The process which generates each image in the sequence from its predecessor is called a REDUCE operation, since both the sample density and the resolution are decreased. A set of band-pass filtered images that correspond to Laplacian or difference of low-pass filtered images, is obtained by taking the difference of successive levels of the Gaussian pyramid. Since these levels differ in sample density it is necessary to interpolate new values between the given values of the lower frequency image before it can be subtracted from the higher frequency image. Interpolation is achieved simply by defining the EXPAND operation as the inverse of the REDUCE operation. Thus, every level in the Laplacian pyramid is a difference of two levels in the Gaussian pyramid, making it equivalent to a convolution with a Laplacian-like band-pass filter.

The Laplacian pyramid is a complete representation of the input image. The input image can be recovered exactly by reversing the steps used in the construction of the pyramid. The Laplacian image fusion scheme is a three step procedure[5,21,22]. First, a Laplacian pyramid is constructed for each of the source images. Next, a Laplacian pyramid is constructed for the composite image by selecting from corresponding nodes in the component pyramids those that have a maximum absolute value. Finally, the composite or fused image is recovered from its pyramid representation through the EXPAND and add reconstruction procedure.

3.2. Ratio of Low-Pass pyramid fusion

The construction of a Ratio of Low-Pass or RoLP pyramid[21] is very similar to that of a Difference of Low-Pass (DoLP) or Laplacian pyramid. First, a Gaussian pyramid is created for the input source image. Instead of taking the difference

between successive layers of the Gaussian pyramid, as is done in the construction of the DoLP pyramid, we take the ratio of two successive layers. The rationale for this approach is that global luminance changes should have no influence on the multiresolution representation of the image.

The RoLP image fusion scheme is essentially identical to the Laplacian scheme. First, a RoLP pyramid is constructed for each of the source images. Then, the RoLP pyramid of the composite image is composed by comparing corresponding nodes in the individual pyramids and selecting those that have a maximum absolute contrast value. Finally the fused image is recovered from its pyramid representation through the EXPAND and multiply reconstruction procedure.

3.3. Contrast pyramid fusion

A contrast pyramid[21] is defined as the difference of the RoLP and the identity pyramid. The rationale for this image representation is the fact that the human visual system is sensitive to luminance ratios or contrast. The fusion scheme is identical to the DoLP and RoLP fusion schemes.

3.4. Filter-Subtract-Decimate pyramid fusion

The filter-subtract-decimate (FSD) process reduces the computational complexity of the Laplacian pyramid construction process[2,6]. However, the resulting image representation does not allow an exact reconstruction of the input image. For binomial filters of small extent significant reconstruction errors may result[16]. The fusion scheme is identical to the DoLP and RoLP fusion schemes.

3.5. Gradient pyramid fusion

In addition to extracting edge information via the Laplacian or RoLP pyramids, an image can also be represented as a series of gradient pyramids[3]. The gradient pyramid can be generated by applying gradient operators to each level of the Gaussian pyramid. This produces horizontal, vertical, and diagonal pyramid sets for each level in the Gaussian pyramid. To reconstruct an image from its gradient pyramid is somewhat more complicated than reconstruction from either a Laplacian or a RoLP pyramid, because there are several intermediate steps involved. The fusion scheme is identical to the other pyramid fusion schemes.

3.6. Morphological pyramid fusion

A morphological low-pass pyramid can be constructed by successive application of an alternating iterative morphological filter followed by sampling[20]. Linear filters alter object intensities and therefore the estimated location of their contours. In contrast, morphological filters remove image details without adding a grayscale bias. They are therefore well suited for hierarchical pattern decomposition. A morphological band-pass pyramid is obtained by subtracting the successive layers of a morphological low-pass pyramid, using the morphological (dilation) equivalent of the linear EXPAND operation.

3.7. Discrete wavelet transform fusion

A method similar to the pyramid image fusion schemes is based on the discrete wavelet transform (DWT). The main difference is that while image pyramids are overcomplete signal representations, the wavelet transform results in a nonredundant image representation. A drawback of the DWT is the fact that it is a shift-variant signal representation, i.e. a shift of the input signal yields a nontrivial modification of the transformation coefficients. When applied to pixel-level image fusion, this results in a shift dependent fusion scheme. This problem can be avoided by using a shift invariant extension of the DWT (SIDWT), which yields an overcomplete and thus shift invariant multiresolution image representation[15-17]. This technique is especially useful for the fusion of motion sequences, where it outperforms other methods with respect to temporal stability and consistency[8,15-17].

3.8. Node selection rule

All multiscale fusion schemes investigated in this study were applied with a node-wise maximum value selection rule, in combination with taking the average of the lowest resolution image representations (pyramid top levels). In practice, there is of course a multitude of selection and combination rules available. The actual choice of the rules will depend on the intended application.

4. EXPERIMENTS

4.1. Creation of complementary image pairs

Figure 1a shows the original 512x512 image of Lena. We repeatedly applied the blur tool in Photoshop 7, using a soft brush with a diameter of 230 pixels, to eliminate high spatial frequency details from Lena's face and from the feathers on her hat. We applied the eraser tool, using a soft brush with a diameter of 300 pixels and with the opacity set to 40%, to create the impression of a highlight and a shadow. Figure 1b shows the result of eliminating the high spatial frequency details from Lena's face, and the addition of a highlight to the upper right corner of the image. Figure 1c shows the result of eliminating the high spatial frequency details from the feathers on Lena's hat, and the addition of a shadow to the lower right corner of the image.

4.2. Fusion of complementary image pairs

The complementary image pairs from Figure 1b and 1c were fused using the different image fusion procedures described in Section 3. An example of a fused result is shown in Figure 1d. The multiscale fusion methods (i.e. all methods presented in Section 3 except the pixelwise average and the principal component method) were applied with the number of scales ranging from 1 to 7.

4.3. Image fusion quality

The quality of the fused result was determined relative to the original Lena image in Figure 1a by computing the quality measure Q (Equation 2) with both the fused image and the original Lena image as input. Figure 2 shows the quality of the fusion result as a function of the number of levels used in the fusion process, for all multiscale fusion methods investigated in this study. For comparison we also show the result of simply averaging the input images. The result of the principal component analysis is not shown here since it coincides with the result of the averaging procedure. Figure 2 shows that the Laplacian image fusion scheme has the best overall performance. The contrast pyramid has a comparable performance which is only slightly below that of the Laplacian pyramid. The discrete wavelet transform and the shift invariant discrete wavelet transform perform less well. When up to 4 levels are used in the fusion process the results of these two methods are comparable. For more than 4 levels the discrete wavelet transform outperforms the shift invariant discrete wavelet transform. When 7 levels are used in the fusion process the result of the discrete wavelet transform even approaches that of the Laplacian pyramid scheme. In contrast, the quality of the fusion result of the shift invariant discrete wavelet transform decreases with an increasing number of fusion levels. Identical and of still lower quality is the result of the filter-subtract-decimate and gradient pyramid fusion processes. In contrast to the other methods the result of these fusion schemes is nearly independent of the number of resolution levels or scales used in the fusion process. The quality of the morphological fusion scheme steadily decreases with the number of fusion levels. When more than 5 levels are used the quality of the fused result is even below that of a simple average. Finally, the ratio pyramid shows the overall lowest performance, which is always below the quality obtained through simple pixelwise averaging.

Most hierarchical image fusion procedures appear to have a maximum level beyond which the quality of the fused result does not increase, and beyond which it sometimes even decreases. The reason is that oversegmentation occurs at these levels, resulting in incorrect local mean image values.

Figure 3 shows the quality of the different levels of a 7 level Laplacian pyramid representation of respectively the two input images from Figure 1b and 1c, and an example of a fused result in Figure 1d. The quality of each level is computed relative to the corresponding level of the Laplacian pyramid representation of the original Lena image as shown in Figure 1a. This figure shows that the quality of the Laplacian pyramid layers of the fused image is always larger than the quality of each of the individual input image layers. This implies that the fused image provides a more complete representation of the image details at all levels of resolution. Note that the quality of the lower levels of the image in which Lena's hat was blurred is less than the quality of corresponding layers of the image in which her face was blurred. This is a result of the fact that the Lena's hat contains a larger amount of small scale detail than her face. Hence, blurring the hat removes a larger amount of detail than blurring her face. As a result the quality of the lower levels of the blurred-hat image is less than those of the blurred-face image.

5. CONCLUDING REMARKS

We showed that a recently introduced universal image quality index Q that quantifies the distortion of a processed image relative to its original version[27], can be used to assess the performance of graylevel image fusion schemes. The method involves the use of an original test image as the reference image, and two or more versions of this reference image that have (partly) complementary distortions. The fusion schemes to be evaluated are applied to the distorted images, and the image quality index Q is computed with the fused result and the original reference images as input. The method can in principle be used to optimize image fusion schemes for different types of distortions, by maximizing the image quality index Q for different parameter settings of the fusion scheme.

REFERENCES

1. Bogoni, L. and Hansen, M., Pattern-selective color image fusion, *Pattern Recognition,* 34*(8)* ,pp. 1515-1526, 2001.
2. Burt, P.J., Smart sensing with a pyramid vision machine, *Proceedings IEEE,* 76*(8)* ,pp. 1006-1015, 1988.
3. Burt, P.J., A gradient pyramid basis for pattern-selective image fusion, In: *Proceedings SID International Symposium 1992*, pp. 467-470, Society for Information Display, Playa del Rey, CA, 1992.
4. Burt, P.J. and Adelson, E.H., The Laplacian pyramid as a compact image code, *IEEE Transactions on Communications,* COM-31*(4)* ,pp. 532-540, 1983.
5. Burt, P.J. and Adelson, E.H., Merging images through pattern decomposition, In: A.G. Tescher (Ed.), *Applications of Digital Image Processing VIII*, pp. 173-181, The International Society for Optical Engineering, Bellingham, WA, 1985.
6. Burt, P.J. and Kolczynski, R.J., Enhanced image capture through fusion, In: *Proceedings Fourth International Conference on Computer Vision*, pp. 173-182, IEEE Computer Society Press, Washington, USA, 1993.
7. Gonzalez, R.C. and Wintz, P., *Digital Image Processing, 2nd ed.* Addison-Wesley, Reading, MA, 1987.
8. Lemeshewsky, G.P., Multispectral multisensor image fusion using wavelet transforms, In: S.J. Park & R.D. Juday (Ed.), *Visual Information Processing VIII*, pp. 214-222, The International Society for Optical Engineering, Bellingham, WA, 1999.
9. Li, H., Manjunath, B.S. and Mitra, S.K., Multisensor image fusion using the wavelet transform, *Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing,* 57*(3)* ,pp. 235-245, 1995.
10. Li, S., Kwok, J.T. and Wang, Y., Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images, *Information Fusion,* 3*(1)* ,pp. 17-23, 2002.
11. Petrovic, V.S. and Xydeas, C.S., Optimizing multiresolution pixel-level image fusion, In: B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications Proc. SPIE Vol. 4385, p. 96-107, Sensor Fusion: Architectures, Algorithms, and Applications V*, pp. 96-107, The International Society for Optical Engineering, Bellingham, WA, 2001.
12. Pratt, W.K., *Digital image processing, 2nd edition* , Wiley, New York, USA, 1991.
13. Pu, T. and Ni, G., Contrast-based image fusion using the discrete wavelet transform, *Optical Engineering,* 39*(8)* ,pp. 2075-2082, 2000.
14. Qu, G.H., Zhang, D.L. and Yan, P.F., Information measure for performance of image fusion, *Electronics Letters,* 38*(7)* ,pp. 313-315, 2002.
15. Rockinger, O., Image sequence fusion using a shift-invariant wavelet transform, In: *Proceedings of the IEEE International Conference on Image Processing*, pp. 288-291,1997.
16. Rockinger, O. (1999). Multiresolution-Verfahren zur Fusion dynamischer Bildfolge. Berlin, GE: Technische Universität Berlin.
17. Rockinger, O. and Fechner, T., Pixel-level image fusion: the case of image sequences, In: I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition VII*, pp. 378-388, The International Society for Optical Engineering, Bellingham, WA, 1998.
18. Scheunders, P., Local mapping for multispectral image visualization, *Image and Vision Computing,* 19*(13)* ,pp. 971-978, 2001.
19. Scheunders, P. and De Backer, S., Fusion and merging of multispectral images using multiscale fundamental forms, *Journal of the Optical Society of America A,* 18*(10)* ,pp. 2468-2477, 2001.
20. Toet, A., A morphological pyramidal image decomposition, *Pattern Recognition Letters,* 9 ,pp. 255-261, 1989.
21. Toet, A., Image fusion by a ratio of low-pass pyramid, *Pattern Recognition Letters,* 9 ,pp. 245-253, 1989.
22. Toet, A., Hierarchical image fusion, *Machine Vision and Applications,* 3 ,pp. 1-11, 1990.
23. Toet, A., Multi-scale contrast enhancement with applications to image fusion, *Optical Engineering,* 31*(5)* ,pp. 1026-1031, 1992.
24. Toet, A., Ruyven, J.J. and Valeton, J.M., Merging thermal and visual images by a contrast pyramid, *Optical Engineering,* 28 ,pp. 789-792, 1989.
25. Toet, A. and Walraven, J., New false colour mapping for image fusion, *Optical Engineering,* 35*(3)* ,pp. 650-658, 1996.

26. Ulug, M.E. and Claire, L., A quantitative metric for comparison of night vision fusion algorithms, In: B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications IV*, pp. 80-88, The International Society for Optical Engineering, Bellingham, WA, 2000.
27. Wang, Z. and Bovik, A.C., A universal image quality index, *IEEE Signal Processing Letters, 9(3)* ,pp. 81-84, 2002.
28. Wilson, T.A., Rogers, S.K. and Kabrisky, M., Perceptual-based image fusion for hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing,* 35*(4)* ,pp. 1007-1017, 1997.
29. Wilson, T.A., Rogers, S.K. and Myers, L.R., Perceptual-based hyperspectral image fusion using multiresolution analysis, *Optical Engineering,* 34*(11)* ,pp. 3154-3164, 1995.
30. Xydeas, C.S. and Petrovic, V.S., Objective pixel-level image fusion performance measure, In: B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications IV*, pp. 89-98, The International Society for Optical Engineering, Bellingham, WA, 2000.
31. Zhang, Z.-L., Sun, S.-H. and Zheng, F.-C., Image fusion based on median filters and SOFM neural network, *Signal Processing,* 81*(6)* ,pp. 1325-1330, 2001.

(a) (b) (c) (d)

Figure 1 (a) Original image of Lena. (b) Result of the application of a local blur operator to Lena's face and the superposition of a bright glare spot to the upper right corner of image (a); Q_b=0.7854. (c) Result of the application of a local blur operator to the feathers on Lena's hat and the superposition of a dark shadow to the lower right corner of image (a); Q_c=0.6751. (d) Result of the fusion of images (b) and (c) with a 7 level Laplacian pyramid, in combination with a maximal node selection rule and the adoption of the average of the mean intensities (top-levels) of (b) and (c) as the mean intensity of (d). Q_d= 0.8986.

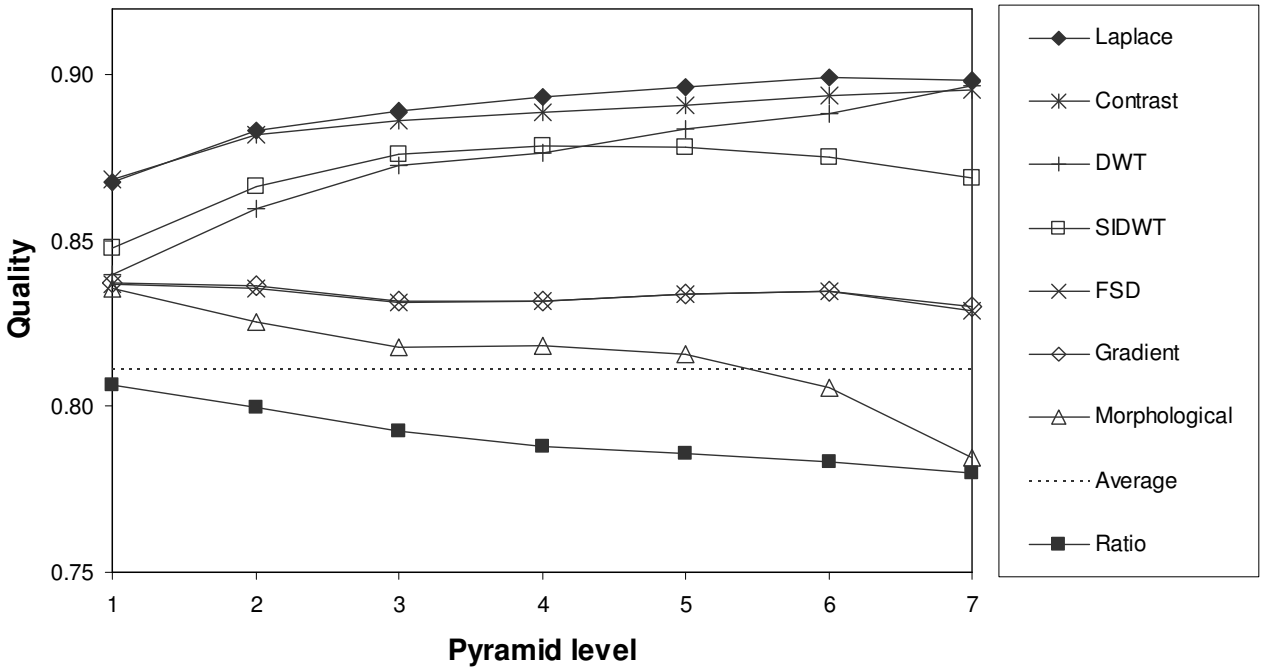

Figure 2 Quality of the fusion result as a function of the number of levels used in the fusion process, for all multiscale fusion methods investigated in this study. The result of simple pixelwise averaging is also shown for comparison. The result of the principal component analysis is not shown here since it coincided with that of the averaging process.

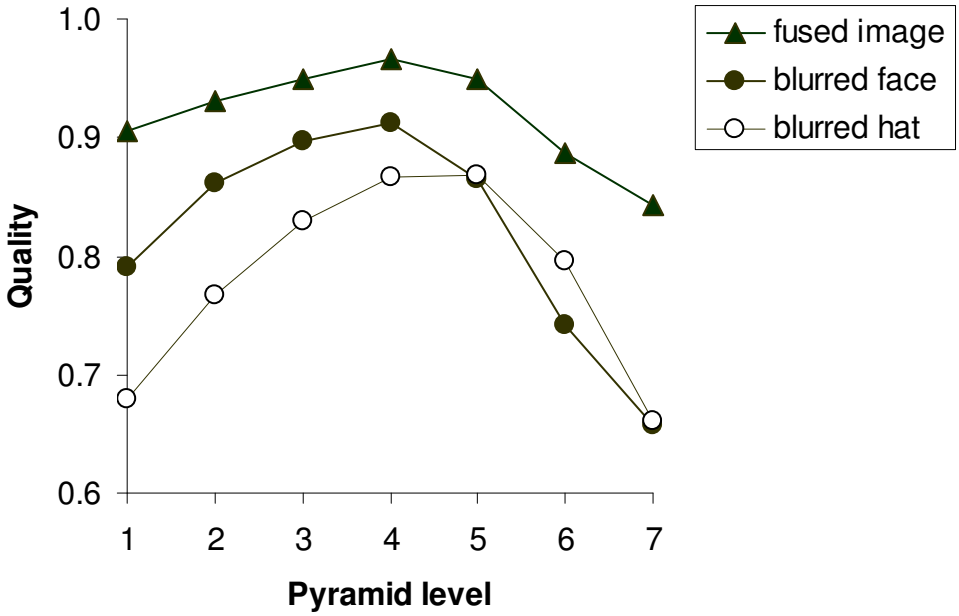

Figure 3 Quality of the different levels of a 7 level Laplacian pyramid representation of respectively the two input images from Figure 1b and 1c, and the fused result shown in Figure 1d. The quality of each level is computed relative to the corresponding level of the Laplacian pyramid representation of the original Lena image as shown in Figure 1a.