A VARIATIONAL APPROACH FOR IMAGE FUSION VISUALIZATION

Gemma Piella

Dpt. of Information and Communication Technologies, Universitat Pompeu Fabra
Pg. Circumvalacio, 08003 Barcelona, Spain

ABSTRACT

We propose a variational framework to perform the fusion of an arbitrary number of images while preserving the salient information and enhancing the contrast for visualization. The assumption is that an optimal fused image has a geometry which approximates the geometry obtained from the inputs, while at the same time it maximizes the contrast without introducing over-saturation or contouring effects. Based on this assumption, we construct an energy functional whose minimization will give the perceptually enhanced fused image.

1. INTRODUCTION

Image fusion can be broadly defined as the process of combining multiple input images into a single one, which contains the 'relevant' information from the inputs, in order to enable a good understanding of the scene. Here, the word 'relevant' should be considered in the sense of 'relevant with respect to the task the output image will be subject to', in most cases high-level tasks such as interpretation or classification.

Herein, we are interested in image fusion for visualization purposes. In this context, the fusion algorithm requires to identify the perceptually salient information in the input images and transfer it into a single composite without introducing any artifacts or inconsistencies which could distract or mislead a human observer.

Since local geometric structures are likely to correspond to perceptually important features, most methods use high order information of the input images, such as first or second derivatives, or scale-space representations, which provide a convenient representation of such geometric structures [1, 2, 3]. For instance, a common strategy in multiscale-based fusion (see [4] for an overview) is to merge the multiscale transform coefficients from the different inputs and apply the inverse transform to the composite multiscale decomposition thus obtained.

In this paper, we propose to use the structure tensor [5, 6, 7] as the main tool to simultaneously describe the geometry of all the inputs. The basic idea is that the fused image should have a structure tensor which approximates the structure tensor obtained from the multiple inputs. At the same time, the fused image should appear 'natural' and 'sharp' to a human interpreter. We therefore propose to combine the geometry merging of the inputs with perceptual enhancement and intensity correction. In particular, the enhancement approach used in this paper is based on the perceptual color correction technique proposed in [8]. It is to be noted that the structure tensor approach is related to the first-order fusion approach presented in [9]. Another related work is proposed in [10] where the structure tensor is examined at multiple resolutions.

The paper is organized as follows. Section 2 briefly presents the perceptual contrast enhancement technique proposed by Bertalmío *et al.* in [8]. Section 3 first reviews the concept of structure tensor for multi-valued data and then proposes a functional minimization approach to find a single-valued image (the fused image) which has the closest geometry to the multi-valued one. In Section 4, a new functional is constructed in order to jointly perform fusion and enhancement. Through a variational approach, the Euler-Lagrange equation is derived, and a gradient descent method is employed to obtain the final output. Some experimental results are shown in Section 5. Finally, Section 6 ends with some conclusions and outlines directions for future development.

2. PERCEPTUAL CONTRAST ENHANCEMENT

Let $I:\Omega\rightarrow[0,1]$ be a gray level image, where $\Omega\subset\mathbb{R}^2$ denotes the image domain and $I(x)$ represents the gray level at $x\in\Omega$. Let $J:\mathbb{R}\rightarrow[0,\infty)$ be a convex even function and $w:\Omega\times\Omega\rightarrow\mathbb{R}^+$ a positive symmetric normalized weight function. Define the average local contrast of I as

$$C(I)=\int_\Omega\int_\Omega w(x,y)J\big(I(x)-I(y)\big)dxdy \qquad (1)$$

and the average quadratic local dispersion as

$$D(I)=\beta\int_\Omega\big(I(x)-\frac{1}{2}\big)^2dx+\gamma\int_\Omega\big(I(x)-I_0(x)\big)^2dx, \quad (2)$$

where $\beta,\gamma>0$ and I_0 is the original image we want to enhance. The first term in D measures the deviation with respect to its theoretical mean $1/2$ (i.e., the middle of the available dynamic range), while the second term in (2) contains an attachment to the original data I_0 to avoid departing too much from the original image.

Then, by minimizing the energy functional

$$E(I)=D(I)-C(I), \qquad (3)$$

we are increasing the contrast (measured by (1)) while controlling the local dispersion (measured by (2)) of the output image I. This formulation complies with basic properties of visual perception and allows to obtain a perceptually enhanced version of the original image I_0.

We remark that the contrast measure (1) encompasses several contrast models. Broadly speaking, function J accounts for the relative lightness appearance of the pixel, while function w weights the amount of local or global contribution. We shall use the same approach as in [8] and choose J such that J' is a sigmoid type function, e.g.,

$$J'(r)=k\arctan(\alpha r), \quad k>0,\alpha>1. \qquad (4)$$

Thus, if $r = I(x) - I(y)$, where x is a fixed pixel and y varies across the image, $J'(r)$ induces a nonlinear behaviour increasing contrast for small differences $|r|$ while saturating for larger $|r|$.

3. GEOMETRY-BASED IMAGE FUSION

3.1 Geometry in multi-valued images

The input images to be fused I_n, $n = 1,\ldots,N$, are considered to be the scalar components of a multi-valued image $I_M : \Omega \to [0,1]^N$. For every point $x = (x_1,x_2) \in \Omega$, $I_M(x) = \big(I_1(x),\ldots,I_N(x)\big)$, where $I_n(x)$ is the gray level of input image I_n at x.

As noticed in [5], the local geometry of a multi-valued image I_M can be represented for all $x \in \Omega$ by

$$G(x) = \sum_{n=1}^{N} \nabla I_n(x)(\nabla I_n(x))^T, \tag{5}$$

where $\nabla I_n = (\frac{\partial I_n}{\partial x_1}, \frac{\partial I_n}{\partial x_2})^T$. Equivalently,

$$G(x) = \begin{pmatrix} \sum_n \left(\frac{\partial I_n}{\partial x_1}\right)^2 & \sum_n \frac{\partial I_n}{\partial x_1}\frac{\partial I_n}{\partial x_2} \\ \\ \sum_n \frac{\partial I_n}{\partial x_1}\frac{\partial I_n}{\partial x_2} & \sum_n \left(\frac{\partial I_n}{\partial x_2}\right)^2 \end{pmatrix}. \tag{6}$$

In literature one refers to G as the structure tensor or second-moment matrix. Since $G(x)$ is a positive semi-definite matrix, its eigenvalues are both real and non-negative. The maximum λ^+ and minimum λ^- eigenvalues give the maximum and minimum rate of change of I_M at a given point x, and the corresponding eigenvectors θ^+, θ^- give the directions of maximum and minimum change.

For a single-valued image ($N = 1$), one can easily obtain $\lambda^+ = |\nabla I|^2$, $\lambda^- = 0$, and $\theta^+ = \nabla I/|\nabla I|$, $\theta^- = \nabla I^\perp/|\nabla I|$. Thus, for single-valued images the gradient is always perpendicular to the edges and we have

$$\nabla I = \sqrt{\lambda^+}\theta^+ \;\text{ and }\; G = \lambda^+\theta^+\theta^{+T}.$$

For multi-valued images,

$$G = \lambda^+\theta^+\theta^{+T} + \lambda^-\theta^-\theta^{-T}$$

and it is the value of λ^+ together with λ^- that discriminates different local geometries. This local geometry is not only due to the contribution of each input, but to the interaction between them. In other words, the structure tensor G gives a simultaneous description of directional information of all inputs involved.

3.2 Single-valued representation of multi-valued images

We can therefore attempt to reconstruct a single-valued image $\tilde{I}$ (the fused image) whose structure tensor $\tilde{G} = \tilde{\lambda}^+\tilde{\theta}^+\tilde{\theta}^{+T}$ (and hence its basic geometry) is similar to $G = \lambda^+\theta^+\theta^{+T} + \lambda^-\theta^-\theta^{-T}$. In the Frobenius norm sense, the best approximation is attained by

$$\tilde{\lambda}^+ = \lambda^+ \;\text{ and }\; \tilde{\theta}^+ = \theta^+,$$

and hence $\nabla\tilde{I} = \sqrt{\lambda^+}\theta^+$. We have thus translated the problem of fusion into the inverse problem of obtaining an image whose gradient is closest to $\sqrt{\lambda^+}\theta^+$. This can be formulated as a dispersion measure to minimize, i.e.,

$$D_V(I) = \int_\Omega |\nabla I(x) - \sqrt{\lambda^+}\theta^+|^2 dx, \tag{7}$$

subject to $0 \le I(x) \le 1$. In the sequel, we refer to $V(x) = \sqrt{\lambda^+}\theta^+$ as the target gradient.

We should point out that the fusion model in (7) coincides with the fusion technique proposed by Socolinsky *et al.* in [9].

Using G as in (6), all input images contribute equally to the geometry description. However, we would like to determine to what extent the local geometry from one input is more relevant than that from another, and use this knowledge to decide on an appropriate weighting. To this aim, for each input I_n, $n = 1,\ldots,N$, of I_M and each pixel $x = (x_1,x_2) \in \Omega$, we assign a normalized weight $s_n(x)$ which represents the local saliency of I_n in a neighborhood of x. We compute the weighted structure tensor as

$$G_s(x) = \begin{pmatrix} \sum_n \left(s_n(x)\frac{\partial I_n}{\partial x_1}\right)^2 & \sum_n s_n^2(x)\frac{\partial I_n}{\partial x_1}\frac{\partial I_n}{\partial x_2} \\ \\ \sum_n s_n^2(x)\frac{\partial I_n}{\partial x_1}\frac{\partial I_n}{\partial x_2} & \sum_n \left(s_n(x)\frac{\partial I_n}{\partial x_2}\right)^2 \end{pmatrix}. \tag{8}$$

The target gradient V is then constructed from the spectral elements of G_s as described above. That is, the magnitude of $V(x)$ is the square root of the maximum eigenvalue of $G_s(x)$ and the direction is the corresponding eigenvector. Note however that the diagonalization does not uniquely specify the sign of the eigenvectors. A simple solution to avoid this ambiguity is to make the orientation of V to agree with that of the gradient of the average of all inputs. The target gradient is thus obtained as

$$V(x) = \sqrt{\lambda^+}\theta^+ \text{sign}\left(\theta^+ \cdot \sum_{n=1}^{N} s_n(x)\nabla I_n(x)\right). \tag{9}$$

There are several possible choices for s_n. In particular, for our experiments in Section 5 we have simply used

$$s_n(x) = \frac{|\nabla I_n(x)|}{\left(\sum_{i=1}^{N} |\nabla I_i(x)|^2\right)^{1/2}}, \tag{10}$$

which is a rough indicator of important areas in the images.

4. PERCEPTUAL GEOMETRY-BASED FUSION

Given a multi-valued image $I_M : \Omega \to [0,1]^N$, our purpose is to produce a fused image $I : \Omega \to [0,1]$ which contains the local geometry from the different bands while perceptually enhancing it and maintaining image coherence. To this aim, we combine the 'enhancing-purpose' functional in (3) with the 'geometry-merging-purpose' functional in (7). The new functional to minimize is

$$\begin{aligned} \tilde{E}(I) &= \eta D_V(I) + E(I) \qquad (11)\\ &= \eta\int_\Omega |\nabla I(x) - V(x)|^2 dx + \beta\int_\Omega \big(I(x) - \frac{1}{2}\big)^2 dx \\ &\quad + \gamma\int_\Omega \big(I(x) - I_0(x)\big)^2 dx \\ &\quad - \int_\Omega\int_\Omega w(x,y)J\big(I(x) - I(y)\big)dxdy, \end{aligned}$$

subject to $0 \le I(x) \le 1$. Here, I_0 is a linear combination (e.g. an average) of the elements in I_M.

We reformulate the functional $\tilde{E}$ in (11) in a discrete framework by replacing the integrals by sums (w.r.t. indexes in a discrete domain Ω) and the gradient by a discrete approximation using forward difference derivatives. To keep expressions simple, we have avoided to use a different notation. The discrete formulation of (11) can be written as

$$\begin{aligned}\tilde{E}(I) &= \eta \sum_{x\in\Omega} |\nabla I(x) - V(x)|^2 + \beta \sum_{x\in\Omega} \big(I(x) - \frac{1}{2}\big)^2 \\ &\quad + \gamma \sum_{x\in\Omega} \big(I(x) - I_0(x)\big)^2 \\ &\quad - \sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y) J\big(I(x) - I(y)\big) \end{aligned} \tag{12}$$

subject to $0 \le I(x) \le 1$.

Before minimizing the functional $\tilde{E}(I)$, let us derive its first variation which is shown[1] to be

$$\begin{aligned}\delta\tilde{E}(I) &= 2\eta\big(\mathrm{div}\, V(x) - \Delta I(x)\big) + 2\beta\big(I(x) - \frac{1}{2}\big) \\ &\quad + 2\gamma\big(I(x) - I_0(x)\big) \\ &\quad - 2\sum_{y\in\Omega} w(x,y) J'\big(I(x) - I(y)\big)\,. \end{aligned}$$

Now, the minimizer of (12) can be found by a gradient descent procedure:

$$\frac{\partial I}{\partial t} = -\delta\tilde{E}(I)\,. \tag{13}$$

This evolution equation can be explicitly discretized (13) by the following Euler scheme:

$$\begin{aligned}\frac{I^{k+1}(x) - I^k(x)}{\Delta t} &= 2\eta\big(\Delta I^k(x) - \mathrm{div}\, V(x)\big) + 2\beta\big(\frac{1}{2} - I^k(x)\big) \\ &\quad + 2\gamma\big(I_0(x) - I^k(x)\big) + \frac{R_I^k(x)}{2M}\,, \end{aligned} \tag{14}$$

where $\Delta t > 0$ is the time increment, $\frac{R_I^k(x)}{2M} = 2\sum_{y\in\Omega} w(x,y) J'\big(I^k(x) - I^k(y)\big)$ and $M = max_{x\in\Omega}\{R_I^0(x)\}$.

We may rewrite (14) as

$$\begin{aligned}I^{k+1}(x) &= I^k(x)\big(1 - 2(\beta+\gamma)\Delta t\big) + 2\eta\big(\Delta I^k(x) - \mathrm{div}\, V(x)\big)\Delta t \\ &\quad + \big(\beta + 2\gamma I_0(x) + \frac{R_I^k(x)}{2M}\big)\Delta t\,. \end{aligned} \tag{15}$$

The constraint that $I^k(x)$ should remain in the range between 0 and 1 can be imposed after each iteration by truncation.

For the initialization image I^0, we use the weighted combination $\sum_n s_n(x) I_n(x)$ with weights given by (10). For the construction of V, the gradients ∇I_n are approximated by forward differences. The divergence $\mathrm{div}\, V$ is then approximated by backward difference of the components of V. The Laplacian operator is realized by $\Delta I^k(x_1,x_2) = I^k(x_1+1,x_2) + I^k(x_1-1,x_2) + I^k(x_1,x_2+1) + I^k(x_1,x_2-1) - 4I^k(x_1,x_2)$.

[1] See [8] for the derivation of δE. The computation of δD_V is straightforward.

5. SIMULATIONS

Based on the image fusion scheme proposed above, we present some experimental results. Unless otherwise stated, we have used $\eta = 0.1$, $\beta = 0.5$, $\gamma = 0.3$, a time increment $\Delta t = 0.15$, a kernel w of Gaussian shape and a slope $\alpha = 10$ in (4). In all cases, we get to a steady state[2] after 12-15 iterations (for images of size 512×512 or below).

Recall that we iteratively minimize (12) using the gradient descent equation in (15). The limit cases (*i*) $\eta = 0$ and (*ii*) $\beta = \gamma = 0$ with null weight function w correspond to functionals E and D_V respectively.

Let us start by comparing the results of simultaneously performing the geometry merging and enhancement (through the minimization of $\tilde{E}$) with those obtained by the geometry merging alone (through the minimization of D_V) or by cascading the merging and enhancement stages. We take as input images the complementary pair shown in Fig. 1(a)-(b). They have been created by blurring the original 'Cameraman' image of size 256×256 with a disk of diameter of 11 pixels. The images are complementary in the sense that the blurring occurs at the left half and the right half, respectively. Fig. 1(c) shows the modulus of the target gradient V. Fig. 1(d) shows the composite image obtained from minimizing functional D_V in (7), while Fig. 1(e) shows the perceptual enhanced version of Fig. 1(d). That is, Fig. 1(e) is the result of minimizing functional E taking Fig. 1(d) as the input image I_0. It can be seen that, in both cases, fluctuations occur around the contours of the man. This 'halo' artifact is avoided in Fig. 1(f), which has been obtained by minimizing $\tilde{E} = E + \eta D_V$. As Fig. 1(f) shows, the proposed technique results in a more 'natural' look, producing sharper contours and enhanced contrast.

Next, we compare our method with some multiscale-based fusion methods. First, we consider the fusion of a magnetic resonance image (MRI) and a computer tomography (CT) image; see Fig. 2(a)-(b). Fig. 2(c) shows the result obtained by the proposed method of minimizing $\tilde{E}$ while the bottom row displays the results obtained through multiscale-based methods. In Fig. 2(d) we have used the Laplacian pyramid fusion scheme [12]. Fig. 2(e) shows the fused image obtained by the wavelet fusion scheme proposed by Li *et al.* in [11]. Fig. 2(e) illustrates the gradient pyramid fusion algorithm proposed by Burt and Kolczynski in [1]. We can observe that the multiscale-based fused images suffer from a loss of contrast. Notice also the ringing artifacts of the wavelet fused image at the bottom part. The fused image obtained by the proposed scheme, however, introduces no noticeable artifact and exhibits an increased contrast and detail information.

Another example is shown in Fig. 3 where images with different focus are fused. Again we can observe that our method (Fig. 2(c)) produces enhanced contrast and crisper edges while avoiding artifacts.

6. CONCLUSIONS

In this paper we have presented a variational model for combining multiple images into a single most informative gray level image for the purpose of visualization. The main features of the model are (1) the use of the structure tensor to si-

[2] Less than 0.5% of root-mean-squared difference from one iteration to the next.

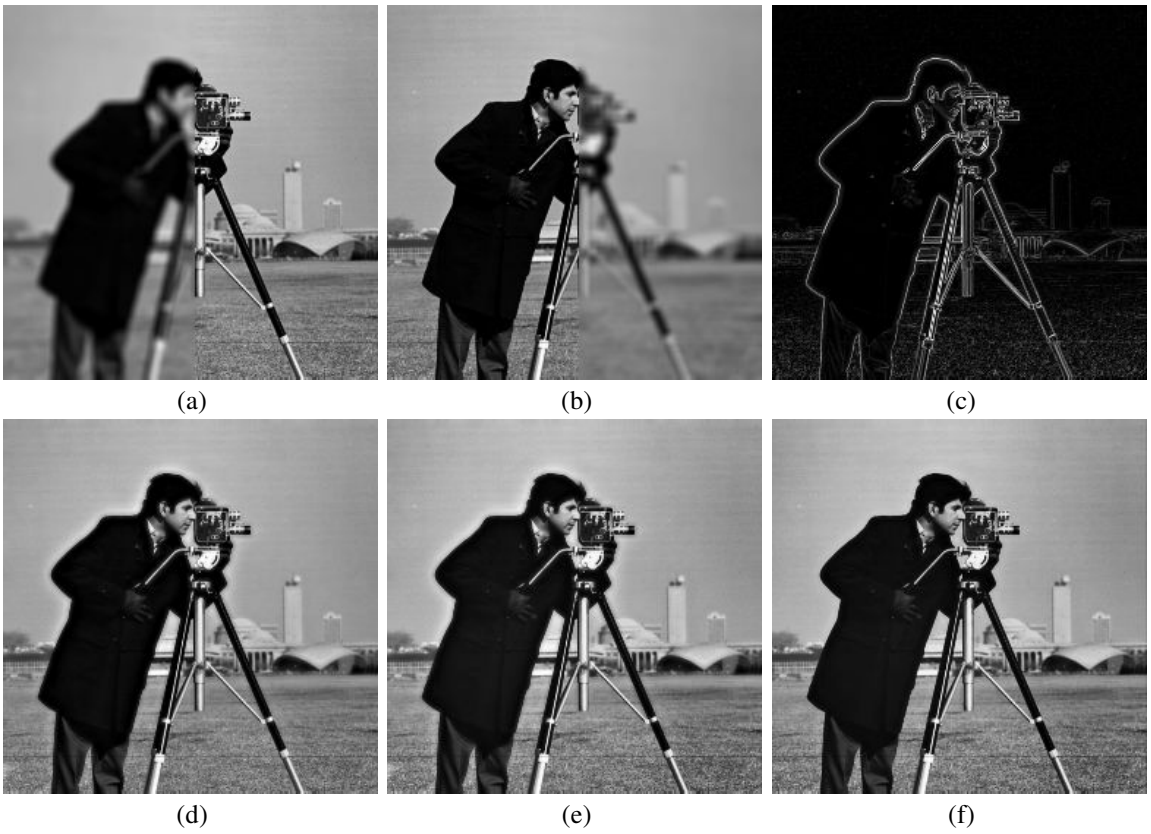

Figure 1: *Top: (a)-(b) input images; (c) target gradient V. Bottom: fused images obtained by minimizing (d) D_V; (e) E with $I_0 = argmin_I D_V$ and (f) $\tilde{E}$.*

multaneously describe the geometry of the inputs, (2) the assumption that the fused image should have a geometry which strongly resembles the inputs geometry, and (3) the combination of geometry merging with perceptual enhancement.

The results obtained show that the proposed method preserves important local perceptual cues while avoiding traditional artifacts such as blurring, ringing or haloing.

Further research is necessary to study the influence of the different parameters and how to select them in order to optimize the results. Another topic which deserves further study is the incorporation of multiscale techniques in the proposed framework.

Acknowledgement

The author would like to thank M. Bertalmío for kindly providing the software implementation used in [8].

REFERENCES

[1] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion", *Proceedings of the 4th International Conference on Computer Vision*, Berlin, Germany, May 1993, pp. 173–182.

[2] G. Harikumar and Y. Bresler, "Feature extraction for exploratory visualization of vector valued imagery", *IEEE Transactions on Image Processing*, 5(9):1324–1334, 1996.

[3] S. K. Rogers, T. A. Wilson, and M. Kabrisky, "Perceptual-based image for hyperspectral data", *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):1007–1017, July 1997.

[4] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions", *Information Fusion*, 9:259–280, December 2003.

[5] S. Di Zenzo, "A note on the gradient of multi-image," *Computer Vision, Graphics, and Image Processing*, 33:116–125, 1986.

[6] A. Cumani, "Edge detection in multispectral images", *CVGIP: Graphical Models and Image Processing*, 53(1):40–51, 1991.

[7] J. Weickert, editor, *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, Stuttgart, 1998.

[8] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi, "Perceptual color correction through variational techniques", *IEEE Transactions on Image Processing*, 16(4):1058–1072, April, 2007.

[9] D. A. Socolinsky and L. B. Wolff, "Multispectral image visualization through first-order fusion", *IEEE Transactions on Image Processing*, 11(8):923–931, August 2002.

[10] P. Scheunders and S. De Backer, "Fusion and merging of multispectral images using multiscale fundamental forms", *Journal of the Optical Society of America A*, 18(10):2468–2477, 2001.

[11] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform", *Graphical Models and Image Processing*, 57(3):235–245, May 1995.

[12] P. J. Burt, "The pyramid as a structure for efficient computation", in A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pages 6–35. Springer-Verlag, Berlin, 1984.

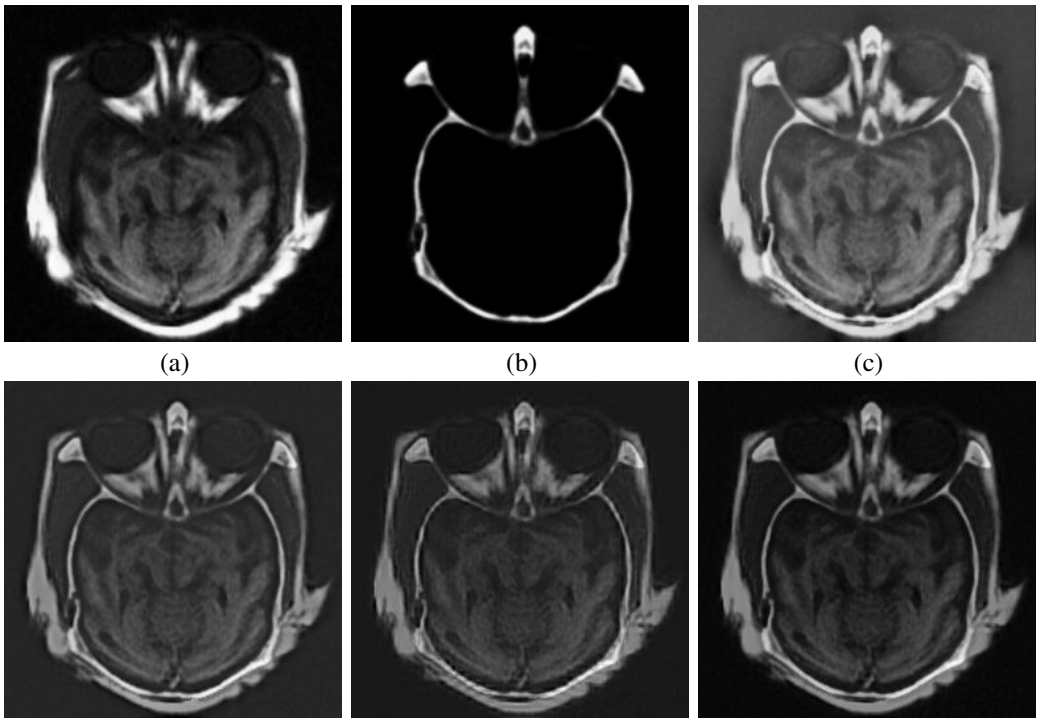

Figure 2: *Top: (a) MRI and (b) CT input images; (c) fused imaged using $\tilde{E}$. Bottom: fused images using (d) the Laplacian pyramid; (e) the DWT and (f) the gradient pyramid.*

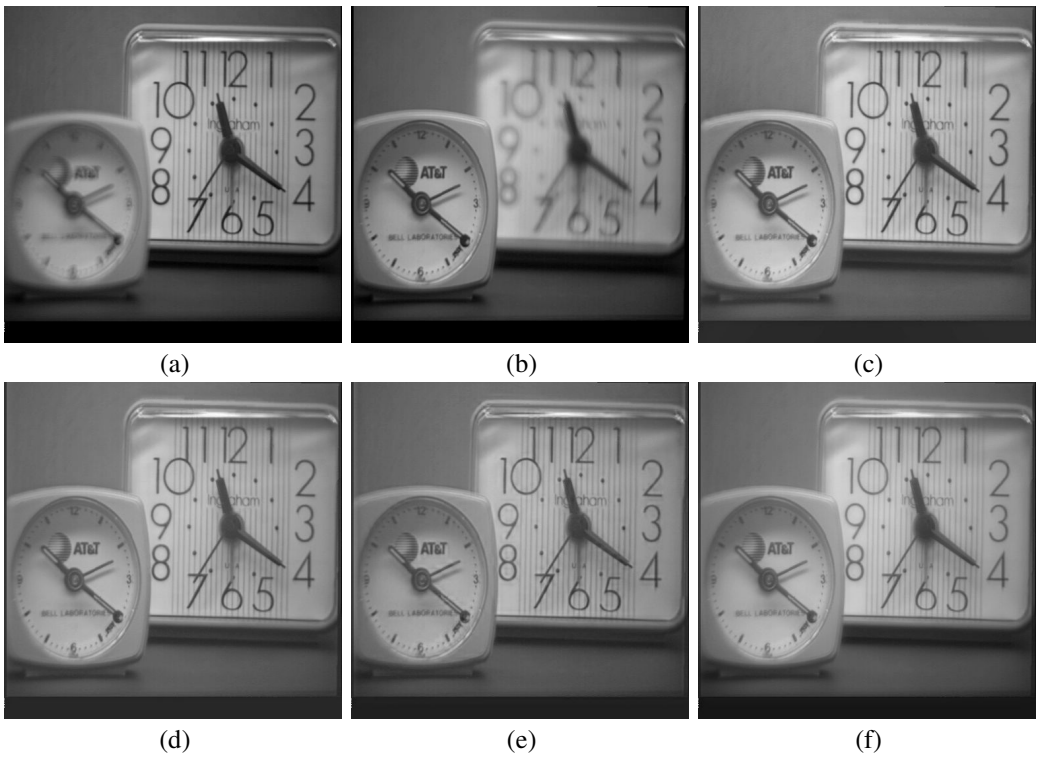

Figure 3: *Top: (a)-(b) input images; (c) fused imaged using $\tilde{E}$. Bottom: fused images using (d) the Laplacian pyramid; (e) the DWT and (f) the gradient pyramid.*