MA3111: Mathematical Image Processing
Multi-Focus Image Fusion

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: December 01, 2021/Last updated: July 1, 2025

Outline of "multi-focus image fusion"

In this lecture, we will introduce multi-focus image fusion

- *using local standard deviations, and*
- *using the variational method with split Bregman iterations.*

The material of this lecture is based on

- K. He, J. Sun, and X. Tang, Guided image filtering, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35 (2013), pp. 1397-1409.
- F. Li and T. Zeng, Variational image fusion with first and second-order gradient information, *Journal of Computational Mathematics,* 34 (2016), pp. 200-222.
- S.-Y. Yang and C.-S. You, A simple and effective multi-focus image fusion method based on local standard deviations enhanced by the guided filter, *Displays,* 72 (2022), article102146.

Introduction to image fusion

- *Image fusion aims to integrate two or more source images of the same scene into a fused image with better visual quality than the source images.*
- Due to the limitation of depth-of-field in the imaging device, images probably cannot focus on all objects and miss partial details leading to blurring.
- Multi-focus image fusion is a technique that extends the depth of field of optical lenses by generating an all-in-focus image from a set of partially focused images.

(a) source image f_1

(b) source image f_2

(c) fused image.

The underlying ideas

- *The sharper pixels generally should have a comparatively higher local variance and hence higher local standard deviation, which is the square root of the local variance.*
- *The Laplacian is a second-order derivative operator, and it highlights sharp intensity transitions in an image and de-emphasizes regions of slowly varying intensities.*
- The sharper parts in the corresponding Laplacian images should come from the sharper parts in the source images.
- It is expected that a well-focused pixel should have a higher local standard deviation in the corresponding Laplacian image.

The underlying ideas (cont'd)

- *The guided filter is an edge-preserving smoothing technique. We use the guided filter to enhance the local standard deviation estimation.*
- We then employ the filtered local standard deviation of the Laplacian image as the focus measure to construct an initial decision map for pixel selection.
- *To make the selection more consistent and avoid pixel misclassification, we further improve the initial decision map using the small region removal strategy.*
- Combined with the small region removal strategy, we choose the pixel with the largest Laplacian-image local standard deviation from the set of partially focused source images.

The problem setting

- We focus on fusing two grayscale partially focused source images. *For color image fusion, we can employ the decision process for sharp pixels selection to their grayscale versions.*
- The pixel values of source and fused images are normalized into the interval $[0, 1]$.
- The image domain is a regular Cartesian grid of size $m_y \times n_x$, i.e., $\Omega_D = \{(i, j) : i = 1, 2, \cdots, m_y, j = 1, 2, \cdots, n_x\}$, where (i, j) denotes a pixel of the image.
- *Let $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ denote the two partially focused source images (i.e., two $m_y \times n_x$ matrices) of the same scene to be fused.*

(a) source image $\boldsymbol{f}_1$

(b) source image $\boldsymbol{f}_2$

The Laplacian images

- We introduce the following 3×3 Laplace kernel,

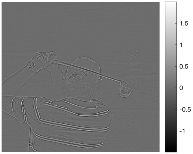
$$K_L = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

  Laplacian is a second-order derivative operator, it highlights sharp intensity transitions in an image and de-emphasizes regions of slowly varying intensities.

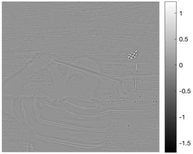
- Taking convolution of the 3×3 Laplace kernel K_L with each source image f_i, we obtain

$$L_i = K_L * f_i, \quad \text{for } i = 1, 2.$$

  We call L_i the Laplacian image of the source image f_i.



(c) Laplacian image L_1



(d) Laplacian image L_2

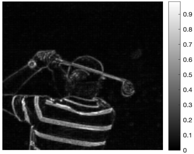
The local standard deviations

- Let $\boldsymbol{L}$ be a Laplacian image and $N(i,j)$ be a given neighborhood centered at (i,j). The local variance of pixel $(i,j) \in \Omega_D$ in the neighborhood $N(i,j)$ is defined as

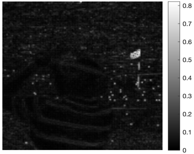
$$\sigma_{\boldsymbol{L}}^2(i,j) = \frac{1}{|N(i,j)|} \sum_{(m,n)\in N(i,j)} \left(\boldsymbol{L}(m,n) - \overline{\boldsymbol{L}}(i,j)\right)^2$$

  and then the local standard deviation of (i,j) is given by $\sigma_{\boldsymbol{L}}(i,j)$.
- $|N(i,j)|$ is the number of pixels that $N(i,j)$ contains; $\overline{\boldsymbol{L}}(i,j)$ is the local mean of pixel (i,j) in $N(i,j)$:

$$\overline{\boldsymbol{L}}(i,j) := \frac{1}{|N(i,j)|} \sum_{(m,n)\in N(i,j)} \boldsymbol{L}(m,n).$$



(e) local standard deviation σ_{L_1}



(f) local standard deviation σ_{L_2}

The guided filter

- *The guided filter is an edge-preserving smoothing operation.*
- Let $\boldsymbol{p}$ be the image to be filtered (*in the context here, $\boldsymbol{p}$ should be the "image" of local standard deviations, $\boldsymbol{\sigma}_{\boldsymbol{L}_1}$ and $\boldsymbol{\sigma}_{\boldsymbol{L}_2}$*). We assume that the relation between the specified guidance image $\boldsymbol{I}$ and the filtering output $\boldsymbol{q}$ (also denoted as $\boldsymbol{Gp}$) is locally linear.
- Let ω_k be a window centered at the pixel (i_k, j_k) with radius r_k, the local linear models of the guided filter in ω_k is given by

$$\begin{aligned} \boldsymbol{q}(i,j) &= \boldsymbol{p}(i,j) - \boldsymbol{n}(i,j), \quad \forall (i,j) \in \omega_k, \\ \boldsymbol{q}(i,j) &= a_k \boldsymbol{I}(i,j) + b_k, \quad \forall (i,j) \in \omega_k, \end{aligned}$$

  where $\boldsymbol{n}$ denotes some unwanted components such as the noise or textures, a_k and b_k are the linear coefficients assumed to be constant within the window ω_k.

The guided filter (cont'd)

- The process of solving the filtering result is to minimize the unwanted noise or textures $\boldsymbol{n}$ in the neighborhood ω_k,

$$\boldsymbol{n}(i,j) = \boldsymbol{p}(i,j) - \left(a_k \boldsymbol{I}(i,j) + b_k\right), \quad \forall (i,j) \in \omega_k.$$

- The minimization problem with regularization reads:

$$\min_{a_k, b_k} \boldsymbol{E}_k(a_k, b_k),$$

  where the objective function $\boldsymbol{E}_k(a_k, b_k)$ is given by

$$\boldsymbol{E}_k(a_k, b_k) = \sum_{(i,j)\in\omega_k} \left(\left(a_k \boldsymbol{I}(i,j) + b_k - \boldsymbol{p}(i,j)\right)^2 + \varepsilon a_k^2 \right)$$

  and $\varepsilon > 0$ is a regularization parameter.

The guided filter (cont'd)

- The unique solution can be directly attained by setting the objective function's gradient to zero,

$$\frac{\partial \boldsymbol{E}_k}{\partial a_k} = 0 = \frac{\partial \boldsymbol{E}_k}{\partial b_k}.$$

- By direct computations, we have

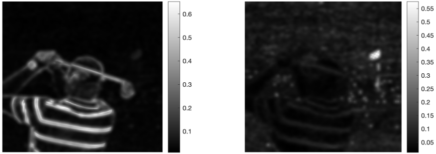
$$\begin{aligned} a_k &= \left(\frac{1}{|\omega_k|}\sum_{(i,j)\in\omega_k} \boldsymbol{I}(i,j)\boldsymbol{p}(i,j) - \mu_k \bar{\boldsymbol{p}}_k\right)\left(\sigma_k^2 + \varepsilon\right)^{-1}, \\ b_k &= \bar{\boldsymbol{p}}_k - a_k \mu_k, \end{aligned}$$

  where $|\omega_k|$ is the total number of pixels, μ_k and σ_k^2 are the mean and variance of the guidance image $\boldsymbol{I}$, respectively, and $\bar{\boldsymbol{p}}_k$ is the mean of the filtering input, all of which are calculated in ω_k.

- We consider all windows ω_k disjoint and $\Omega_D = \cup_k \omega_k$.

Numerical results

- $N(i,j)$ in the calculations of local standard deviation is chosen as a 5×5 pixel array centered at the pixel (i,j).
- In the guided filter, we take $r_k = 2$ for all windows ω_k, i.e., ω_k is a 5×5 pixel array for all k.
- The regularization parameter is chosen as $\varepsilon = 0.1$.



(g) guided filtered local standard deviation $\boldsymbol{G\sigma_{L_1}}$; *(h) guided filtered local standard deviation* $\boldsymbol{G\sigma_{L_2}}$

The decision maps for pixel selection

- For each pixel $(i,j) \in \Omega_D$, we set

$$\boldsymbol{B}(i,j) = \begin{cases} 1, & \text{if } \boldsymbol{G\sigma}_{\boldsymbol{L}_1}(i,j) > \boldsymbol{G\sigma}_{\boldsymbol{L}_2}(i,j), \\ 0, & \text{otherwise,} \end{cases}$$

  where we use the source images $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ as the guidance images for computing $\boldsymbol{G\sigma}_{\boldsymbol{L}_1}$ and $\boldsymbol{G\sigma}_{\boldsymbol{L}_2}$, respectively.

- We further improve the initial decision map $\boldsymbol{B}$ to $\boldsymbol{B}_S$ by the small region removal strategy: *a region which is smaller than an area threshold* $R(= 0.01)$ *is reversed in the map* $\boldsymbol{B}$.

- The binary decision map $\boldsymbol{B}_S$ can be further modified as a non-binary weight map $\boldsymbol{GB}_S$ by using the guided filter one more time with the guidance image $\boldsymbol{B}_S$ itself:

$$0 \leq \boldsymbol{GB}_S(i,j) \leq 1, \quad \forall\, (i,j) \in \Omega_D.$$

LSDGF1 and LSDGF2 multi-focus image fusion

With the help of the decision maps $\boldsymbol{B}_S$ and the weight map $\boldsymbol{GB}_S$ We can form a fused image $\boldsymbol{f}$ in two different ways:

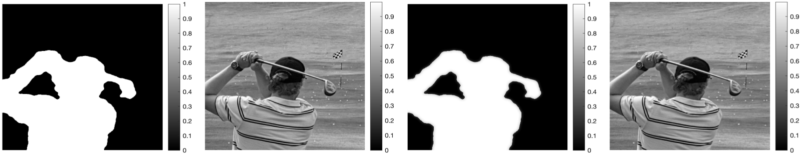
- **LSDGF1:**

$$\boldsymbol{f}(i,j) = \boldsymbol{B}_S(i,j)\boldsymbol{f}_1(i,j) + \big(1 - \boldsymbol{B}_S(i,j)\big)\boldsymbol{f}_2(i,j), \quad \forall\,(i,j) \in \Omega_D.$$

- **LSDGF2:**

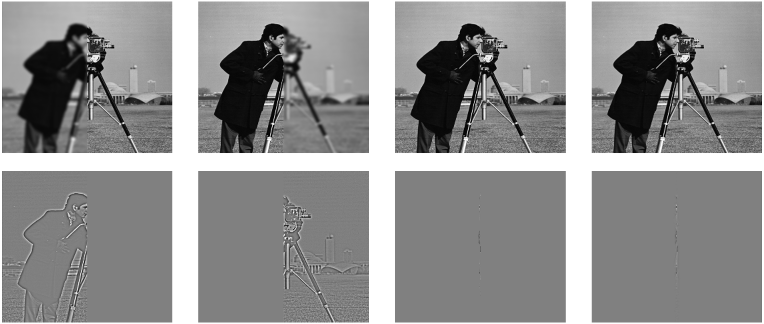
$$\boldsymbol{f}(i,j) = \boldsymbol{GB}_S(i,j)\boldsymbol{f}_1(i,j) + \big(1 - \boldsymbol{GB}_S(i,j)\big)\boldsymbol{f}_2(i,j), \quad \forall\,(i,j) \in \Omega_D.$$

  The fused image $\boldsymbol{f}$ is a weighted combination of $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$.



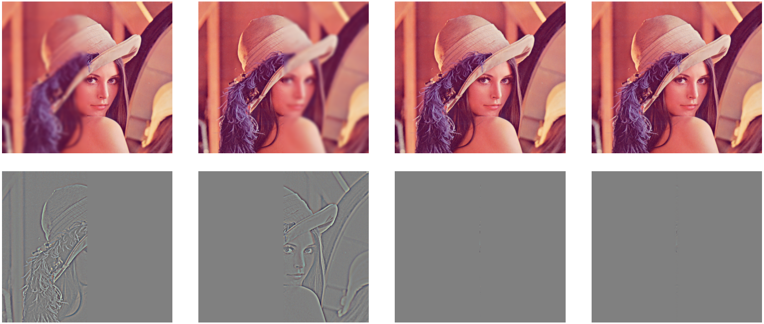
(i) $\boldsymbol{B}_S$; (j) fused image by LSDGF1; (k) $\boldsymbol{GB}_S$; (l) fused image by LSDGF2.

Image “Cameraman”



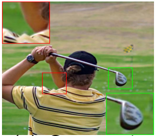
Source images of left half blurred and right half blurred, respectively, by the Gaussian blur with $\mu = 0$ and $\sigma = 4$; fused images by LSDGF1 and LSDGF2, respectively. The residual is defined as $\boldsymbol{f}_{exact} - \boldsymbol{I} + 0.5$

Image "Lena"



Source images of left half blurred and right half blurred, respectively, by the Gaussian blur with $\mu = 0$ and $\sigma = 4$; fused images by LSDGF1 and LSDGF2, respectively. The residual is defined as $\boldsymbol{f}_{exact} - \boldsymbol{I} + 0.5$

Image “Golf”


(a) Source 1


(b) Source 2

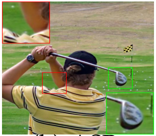

(c) DWT

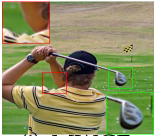

(d) DTCWT


(e) NSCT


(f) GFF


(g) SR


(h) ASR


(i) MWGF


(j) ICA

Image “Golf” (cont’d)

(k) NSCT-SR

(l) SSSDI

(m) Quadtree

(n) DSIFT

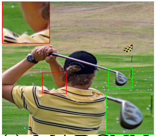
(o) SRCF

(p) GFDF

(q) BRW

(r) MISF

(s) CNN

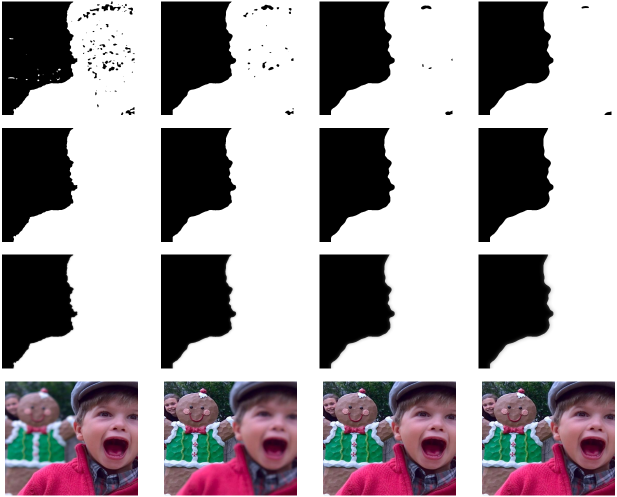


(t) MADCNN

(u) LSDGF1

(v) LSDGF2

Image “Child”



From left to right: 3×3, 5×5, 7×7, and 9×9 for sizes of $N(i,j)$ and ω_k, and $R = 0.01$; From top to bottom: $\boldsymbol{B}$, $\boldsymbol{B}_S$, and $\boldsymbol{GB}_S$; Source images $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, and fused images by LSDGF1 and LSDGF2

Average running time of different fusion methods

Type	Method	Time in second
Transform domain	DWT	0.1348
	DTCWT	0.5557
	NSCT	7.3610
	SR	164.3228
	MWGF	4.5028
	NSCT-SR	109.5024
Spatial domain	QUADTREE	1.6868
	DSIFT	3.3900
	GFDF	0.1468
	BRW	0.9659
	MISF	0.1224
Deep learning	CNN	112.5151
	MADCNN*	0.2164
	DRPL*	0.1530
	SESF*	0.7391
	GACN*	0.2318
Present	LSDGF1	0.1126
	LSDGF2	0.1553

Comparisons with deep learning methods

(a) Source1 (b) Source2 (c) LSDGF1 (d) LSDGF2

(e) MADCNN (f) DRPL (g) SESF (h) GACN

Image "Child": (a) source image f_1; (b) source image f_2; (c)-(h) fused images by different fusion methods

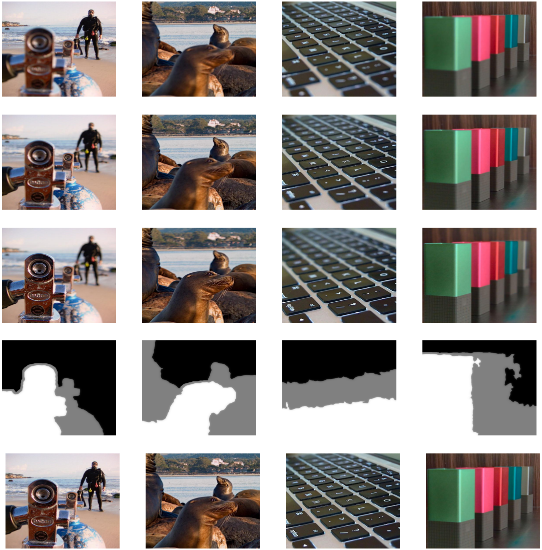
Comparisons with deep learning methods (cont'd)

(a) LSDGF1 (b) LSDGF2

(c) MADCNN (d) DRPL (e) SESF (f) GACN

Image "Child": decision maps of different fusion methods

Image fusion of the 3-focus images (LSDGF2)



The bounded variation space $BV(\Omega)$

Let Ω be an open subset of $\mathbb{R}^2$. The space of functions of bounded variation $BV(\Omega)$ is defined as the space of real-valued function $u \in L^1(\Omega)$ such that the total variation is finite, i.e.,

$$BV(\Omega) = \{u \in L^1(\Omega) : \|u\|_{TV(\Omega)} < \infty\},$$

where

$$\|u\|_{TV(\Omega)} := \sup\left\{ \int_\Omega u(\nabla \cdot \varphi)\, d\boldsymbol{x} : \varphi \in C_c^1(\Omega, \mathbb{R}^2), \|\varphi\|_{(L^\infty(\Omega))^2} \leq 1 \right\},$$

$C_c^1(\Omega, \mathbb{R}^2)$ is the space of continuously differentiable vector functions with compact support in Ω, $L^1(\Omega)$ and $L^\infty(\Omega)$ are the usual $L^p(\Omega)$ space for $p = 1$ and $p = \infty$, respectively, equipped with the $\|\cdot\|_{L^p(\Omega)}$ norm. *For a sufficiently smooth function u, we have* $\|u\|_{TV(\Omega)} = \int_\Omega |\nabla u| d\boldsymbol{x}$.

Then $BV(\Omega)$ is a Banach space with the norm,

$$\|u\|_{BV(\Omega)} := \|u\|_{L^1(\Omega)} + \|u\|_{TV(\Omega)}.$$

Variational method for multi-focus image fusion

We consider a variational method for multi-focus image fusion which only uses *the first-order gradient information.*

Given the gradient information V and the data function u_0, we solve the minimization problem,

$$\min_{u \in BV(\Omega) \cap L^2(\Omega)} \left\{ \int_\Omega |\nabla u - V| \, d\boldsymbol{x} + \frac{\lambda}{2} \int_\Omega (u - u_0)^2 \, d\boldsymbol{x} \right\},$$

where $\lambda > 0$ is a model parameter.

Some remarks on the variational model

- *In general, the given data function u_0 is chosen as one of the source images or their average (e.g. $u_0 = (f_1 + f_2)/2$ for 2 source images).*
- The first term in the model plays a data-fidelity term which forces the gradient of the fused image u matching with the *feature target V*. Therefore, the target gradient information V is more crucial, needs to be further designed.
- The second term acts not only for the data fidelity, but also somewhat for the regularization which ensures the uniqueness of the minimizer u of minimization problem.
- **Theorem [LZ 2016]:** *Assume that $u_0 \in BV(\Omega) \cap L^2(\Omega)$ and V is a finite vector-valued Radon measure, then the minimization problem has a unique minimizer $u^* \in BV(\Omega) \cap L^2(\Omega)$.*

The split Bregman iterative scheme

The minimization problem can be solved efficiently by *the split Bregman iterative scheme.*

- We reformulate the model as the following constrained minimization problem:

$$\min_{u,d} \int_\Omega |d|\, d\boldsymbol{x} + \frac{\lambda}{2}\int_\Omega |u-u_0|^2\, d\boldsymbol{x}$$
$$\text{subject to} \quad d = \nabla u - V,$$

  where d is an induced variable related to the iterative scheme.

- Given the auxiliary variable $b^{(k)}$, we define the energy functional

$$\begin{aligned} E_r(u,d) \quad := \quad & \int_\Omega |d|\, d\boldsymbol{x} + \frac{\lambda}{2}\int_\Omega |u-u_0|^2\, d\boldsymbol{x} \\ & + \frac{\mu}{2}\int_\Omega |\nabla u - V - d + b^{(k)}|^2\, d\boldsymbol{x}, \end{aligned}$$

  where $\lambda > 0$ and $\mu > 0$ are two penalty parameters.

The split Bregman iterative scheme (cont'd)

- u**-subproblem:**

$$u^{(k+1)} = \arg\min_{u} E_r(u, d^{(k)}).$$

- d**-subproblem:**

$$d^{(k+1)} = \arg\min_{d} E_r(u^{(k+1)}, d).$$

- **The auxiliary variable** b**:**

$$b^{(k+1)} = b^{(k)} + \nabla u^{(k+1)} - V - d^{(k+1)}.$$

Construction of gradient V

Let f be the fused image by LSDGF1 or LSDGF2, the discrete version of the first-order gradient V can readily be attained by

$$V(i,j) = \nabla f(i,j), \quad \forall\, (i,j) \in \Omega_D,$$

which is expected to be close to the target image gradient.

The first four are the source images and the fifth is the fused image

Another feature selection

Suppose that two source images f_1 and f_2 are defined in Ω_D.

- The corresponding image features are given by

$$M_1(i,j) := \frac{\partial f_1}{\partial x_1}(i,j) \quad \text{and} \quad M_2(i,j) := \frac{\partial f_2}{\partial x_1}(i,j).$$

- For $\boldsymbol{x}, \boldsymbol{y} \in \Omega_D$, the feature selection will be operated through a convolution kernel K which is defined as follows:

$$K(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \frac{1}{|\omega_{\boldsymbol{x}}|}, & \text{if } \boldsymbol{y} \in \omega_{\boldsymbol{x}}, \\ 0, & \text{otherwise}, \end{cases}$$

  where $\omega_{\boldsymbol{x}}$ denotes a bounded neighborhood of the pixel $\boldsymbol{x}$ with area $|\omega_{\boldsymbol{x}}|$ and K is an average kernel that establishes the selection criterion.

Feature selection (cont'd)

- We take M_1^2 and M_2^2 as salience measure. The main purpose is to eliminate the negative sign at the pixel that makes the feature selection wrong. The convolution K operates as below

$$B(i,j) = \begin{cases} 1, & \text{if } (K * M_1^2)(i,j) > (K * M_2^2)(i,j), \\ 0, & \text{otherwise,} \end{cases}$$

  where $*$ is the convolution symbol.

- Then we operate the convolution kernel K to B

$$\widetilde{B}(i,j) = \begin{cases} 1, & \text{if } (K * B)(i,j) > 0.5, \\ 0, & \text{otherwise,} \end{cases}$$

  where $\widetilde{B}$ is a binary mask. *The main goal of this step is to eliminate isolated points.*

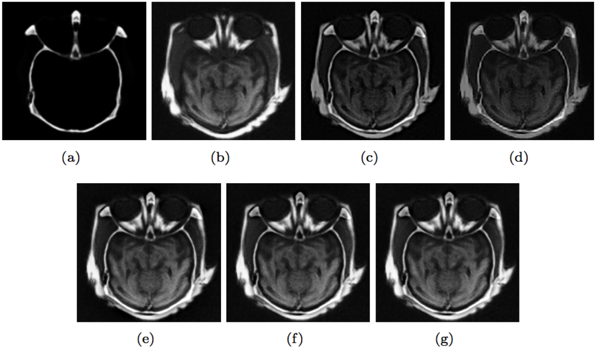
Feature selection (cont'd)

- When the pixel value of $\widetilde{B}$ at (i, j) is 1, it means that the first source image $\boldsymbol{f}_1$ has more salience feature than the second source image $\boldsymbol{f}_2$ at that pixel (i, j).
- The first component of the feature target $V = (V_1, V_2)$ can be determined as follows:

$$V_1 = M_1 \circ \widetilde{B} + M_2 \circ (\mathbf{1} - \widetilde{B}),$$

  where $\circ$ means entrywise product (i.e., Hadamard product).
- We do the same procedure for the feature component V_2.
- *The feature selection will go wrong at the blurred edge because the image feature may contain both blurred and clear pixels.*

Numerical results



(a) shows the CT image in which the structure of bone is better visualized; (b) is the MR image in which the pathological soft tissues are better visualized; (e) fused image by the above method; (c)(d)(f)(g) fused images by other methods.

Open problems

Image fusion aims to integrate information from several source images into a fused image with better visual quality than the source images.

Despite the remarkable progress that has been achieved in recent years, there remain several challenges that need further improvements:

1. *Transition regions between focused and defocused ones.*
2. *Intersection of defocused regions of the source images is nonempty.*
3. *Source images are mis-registered.*