

La Fusión de Imágenes

Por: Fray León Osorio Rivera¹

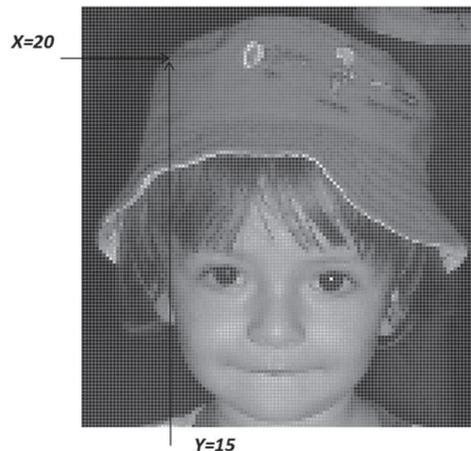
Operaciones con imágenes

En el procesamiento digital de imágenes, muchas soluciones involucran en lugar de procesar una sola imagen, varias de ellas (generalmente de una misma escena) sobre todo para hallar diferencias o fundirlas en una sola. En el primer caso, una de las mayores utilidades comerciales es la detección de movimiento. En el segundo, el enfoque automático de las cámaras fotográficas o los microscopios electrónicos.

Para poder formular una operación con imágenes, en primer lugar cada una se define como una función de la intensidad (tono o cantidad de un color) de la siguiente manera:

$$I(x, y)$$

Para todo $0 \leq x < \text{ancho}$ y $0 \leq y < \text{alto}$. Lo cual significa que dadas 2 coordenadas, en el cruce de ellas se tiene una intensidad determinada: $I(20,15)=155$



¹ Ingeniero de sistemas egresado de la Universidad de Antioquia, maestrante de Automatización en el Instituto Tecnológico Metropolitano, Decano y docente de la Facultad de Ingeniería en la Corporación Universitaria de Sabaneta, docente de cátedra en la Universidad de San Buenaventura y en la Universidad de Medellín

En este sentido, el almacenamiento de una imagen es una matriz de números enteros, cuyos valores van desde 0 hasta 255. Las dimensiones de la matriz son por tanto, el ancho y alto de la imagen.

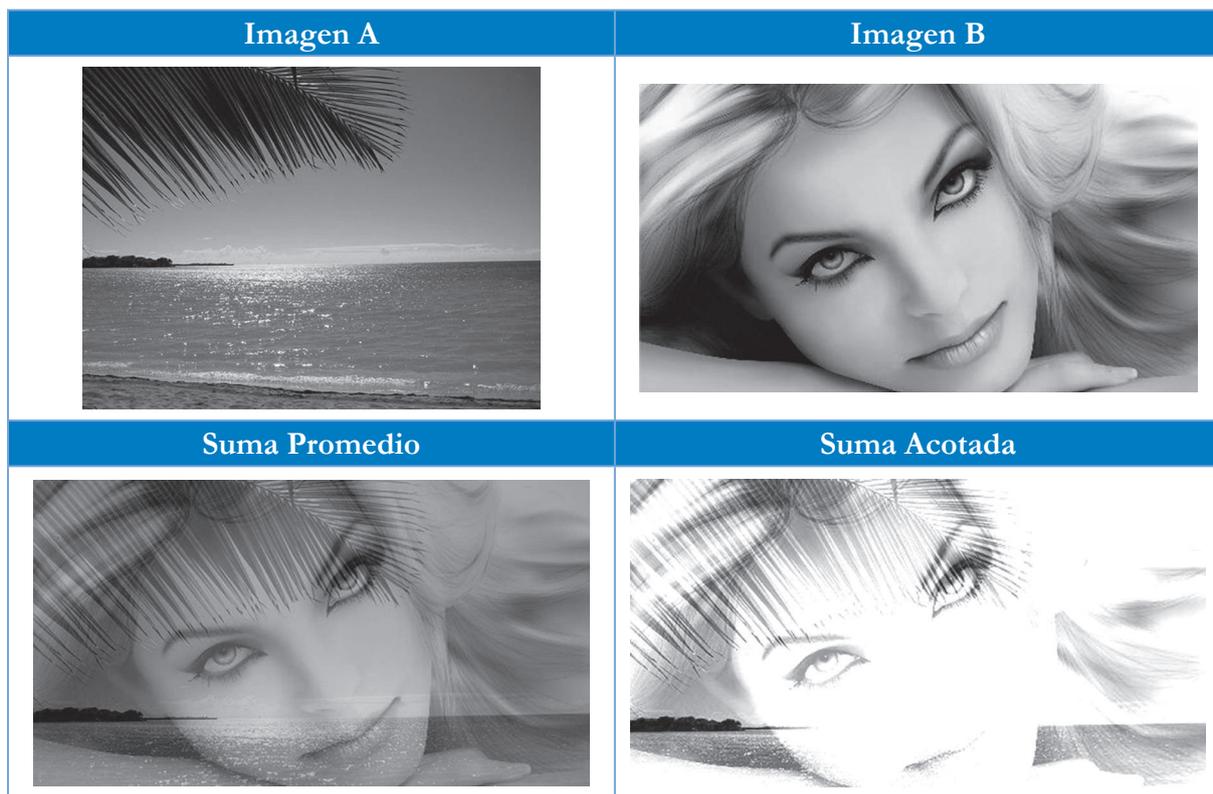
La primera operación a considerar es la **suma** de imágenes, cuya limitante es la saturación de la imagen resultante debido a que los valores sumados pueden superar el valor máximo de 255. Hay 2 alternativas:

- **La media de las intensidades**

$$I(x, y) = \frac{I_A(x, y) + I_B(x, y)}{2}$$

- **Acotamiento**

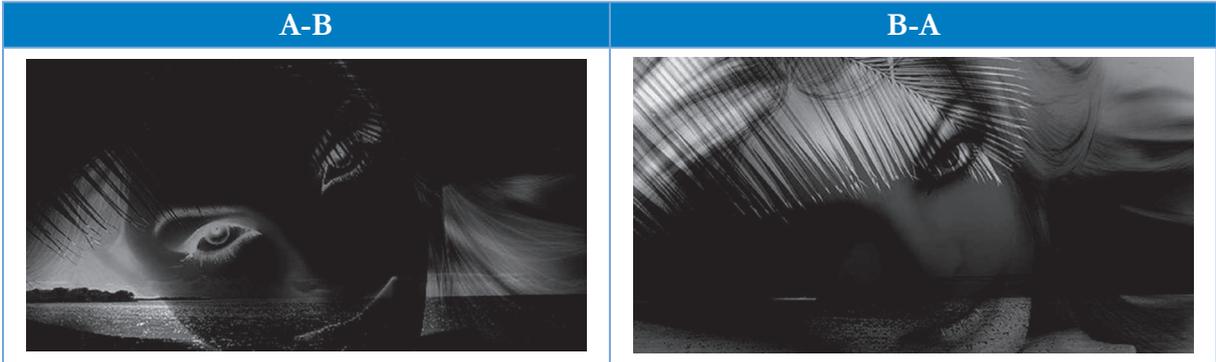
$$I(x, y) = \begin{cases} I_A(x, y) + I_B(x, y) & I_A(x, y) + I_B(x, y) \leq 255 \\ 255 & I_A(x, y) + I_B(x, y) > 255 \end{cases}$$



En la **resta** entre imágenes se debe tener en cuenta que se pueden obtener valores negativos. Una primera forma de resolver este problema es la resta acotada:

$$I(x, y) = \begin{cases} I_A(x, y) - I_B(x, y) & I_A(x, y) \geq I_B(x, y) \\ 0 & I_A(x, y) < I_B(x, y) \end{cases}$$

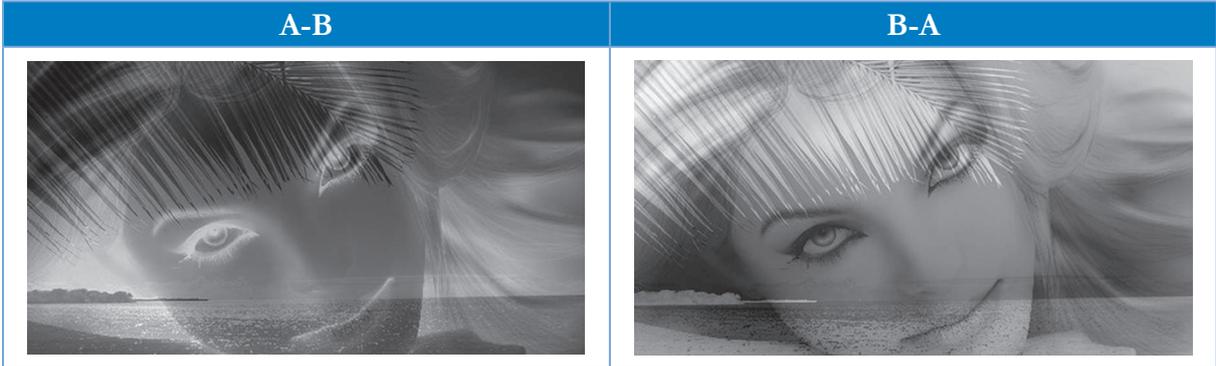
La operación no es conmutativa:



Otra forma de resta es el ajuste de la media, cuya fórmula sería:

$$I(x, y) = \frac{255}{2} + \frac{I_A(x, y) - I_B(x, y)}{2}$$

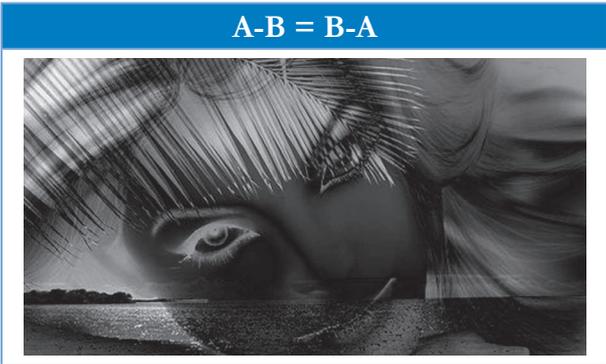
La operación tampoco es conmutativa:



Una última forma de resta es la basada en el valor absoluto:

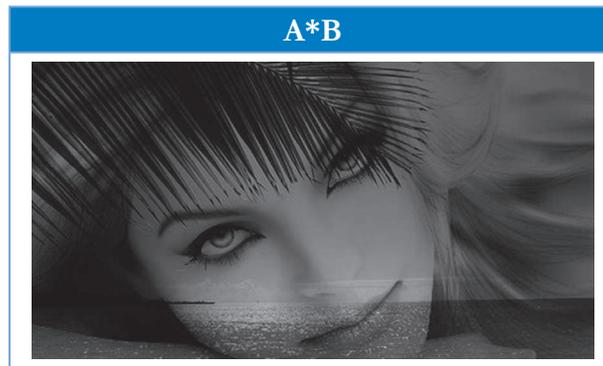
$$I(x, y) = |I_A(x, y) - I_B(x, y)|$$

La operación es conmutativa:



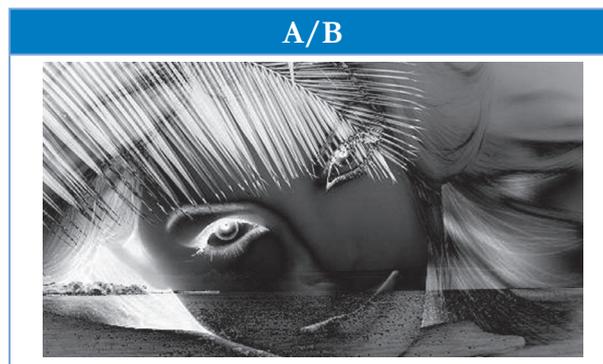
La **multiplicación** de imágenes se realiza mediante la siguiente fórmula:

$$I(x, y) = \frac{1}{255} I_A(x, y) I_B(x, y)$$



La **división** entre imágenes se realiza mediante la siguiente fórmula:

$$I(x, y) = 255 - \frac{255 \min(I_A(x, y), I_B(x, y))}{\max(I_A(x, y), I_B(x, y)) + 1}$$



¿Para qué la fusión de imágenes?

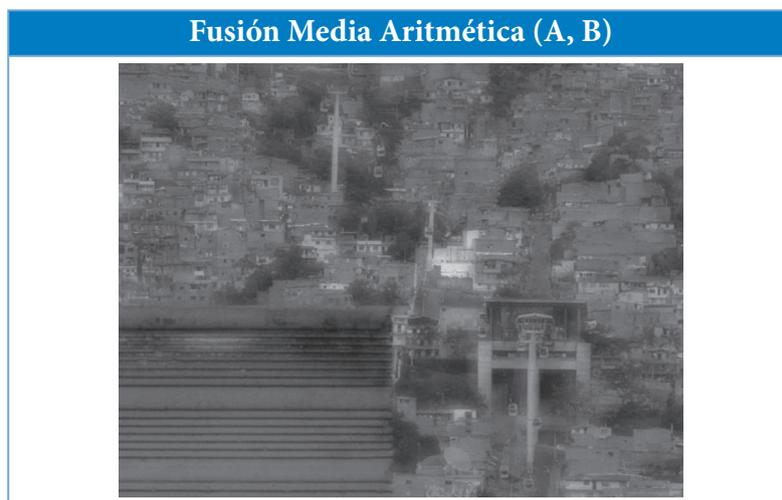
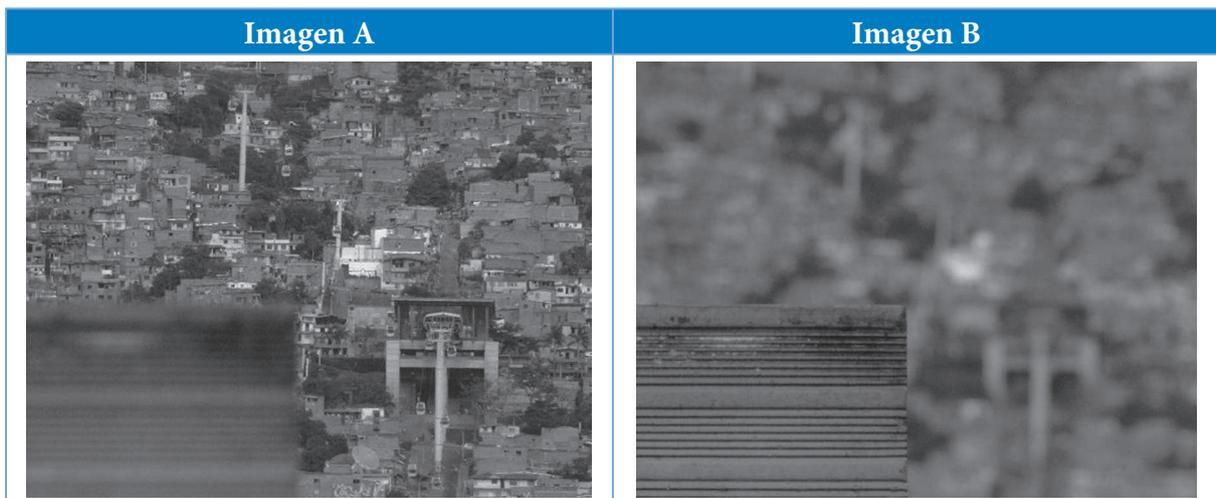
Debido al dinamismo en las necesidades y técnicas de adquisición de imágenes, la fusión de imágenes se ha convertido en una herramienta bastante útil. Por ejemplo, en materia de *sensado remoto*, en (Zhenhua Li, 2004) se introduce un algoritmo que une la imagen obtenida de un sensor infrarrojo con otra imagen del espectro visible para detectar objetivos militares usando DWFT (*Discrete wavelet frame transform*) igual que sucede en (Haitao Li) donde se define un nuevo esquema de fusión de imágenes que combina la descomposición wavelet de multiescala con la técnica SVR (*Synthetic Variable Ratio*) con el objetivo de conservar la información espacial de la imagen pancromática y reproducir las características espectrales de la imagen multiespectral. En materia de imágenes médicas, en (France Laliberté, 2002) se introduce un nuevo esquema de fusión de imágenes y angiografías de la retina evaluando algunos métodos piramidales y basados en DWT.

En materia comercial, se podría pensar por ejemplo en una cámara profesional que capture varias imágenes con diferentes enfoques (lo cual implica diferentes profundidades de campo) y con mínimas diferencias de tiempo (al punto de que puedan considerarse simultáneas). Esto significa que todo el escenario estará enfocado por partes. La supuesta cámara incluiría un software que permita obtener una imagen nítida a partir de la fusión de las partes enfocadas de cada una de las imágenes.

Algoritmos para la fusión de imágenes

El primer algoritmo que se tiene en cuenta para fusionar imágenes es el que calcula la nueva imagen a partir de la media de las dos imágenes de entrada, realizado píxel a píxel. Se le conoce como método de la **media aritmética**. Dadas dos imágenes I_A e I_B , la imagen fusionada se calcula como:

$$I(x, y) = \frac{I_A(x, y) + I_B(x, y)}{2}$$



Este método tiene el problema de no distinguir que partes de las imágenes están nítidas o borrosas. Una solución sería tratar de asignar más peso al pixel más nítido. Se trata de la **media ponderada**. En este caso se definen dos matrices de pesos W_A y W_B tales que:

$$\begin{aligned} 0 \leq w_A(x, y), w_B(x, y) \leq 1 \\ w_A(x, y) + w_B(x, y) = 1 \end{aligned}$$

Donde $w_{A1}(x, y)$ es un elemento de W_A y $w_B(x, y)$ es un elemento de W_B . Luego la imagen resultante se obtendría mediante la siguiente ecuación:

$$I(x, y) = w_A(x, y)I_A(x, y) + w_B(x, y)I_B(x, y)$$

Para asignar los pesos, se puede utilizar la estrategia en (Haeberli, 1994) la cual se basa en los bordes de los objetos de las imágenes los cuales se obtienen aplicando un filtro de paso alto, debido a que las altas frecuencias en una imagen son aquellas partes que cambian rápidamente, lo cual son los bordes de la imagen.

Aplicando como máscara del filtro paso alto la siguiente matriz:

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

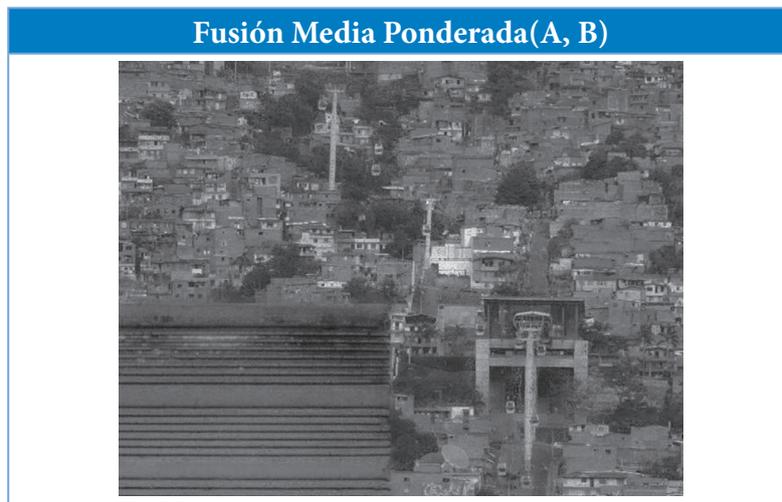
Se puede obtener mediante convolución la matriz:

$$H(x, y) = |h * I(x, y)|$$

Que correspondiente a la información de los bordes. En consecuencia, las matrices de pesos serían:

$$\begin{aligned} W_A(x, y) &= \frac{H_A(x, y)}{H_A(x, y) + H_B(x, y)} \\ W_B(x, y) &= 1 - W_A(x, y) \end{aligned}$$

El resultado final obtenido sería el siguiente:



Según (Mertens, 2007), el inconveniente con la fusión por Media Ponderada es que existen áreas donde los pesos cambian rápidamente y esto afecta la imagen resultante con efectos “fantasmales” no deseados como alrededor del ladrillo o alrededor de la estación de metrocable en la imagen anterior.

El anterior método ha tratado de obtener detalles de la imagen. Esa misma estrategia es la que utiliza la **pirámide laplaciana** ó pirámide de diferencia de paso bajo (DOLP) (Sadjadi). Este método obtiene recursivamente una serie de imágenes aplicando un filtro paso bajo y una reducción de tamaño. Cada imagen pierde detalle y dimensión a partir de dichas operaciones (coeficientes de aproximación o pirámide gaussiana) y se conserva la diferencia entre esas imágenes y las originales (coeficientes de detalle o pirámide Laplaciana) como se observa en (Burt, 1993).

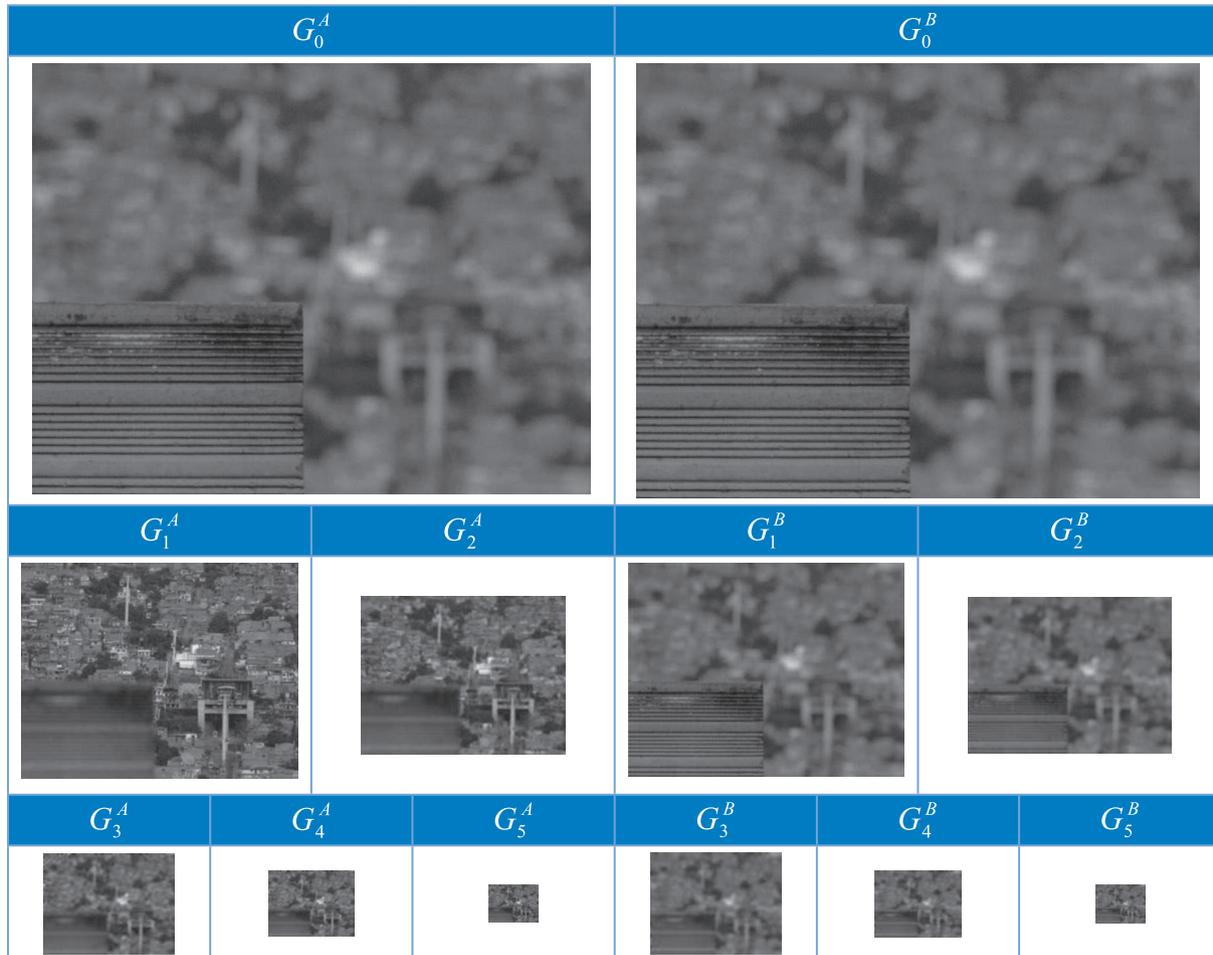
La versión de la **pirámide laplaciana RE** (Reducción-Expansión) calcula la pirámide gaussiana mediante la fórmula:

$$G_k = [w * G_{k-1}]_{\downarrow 2}$$

Donde:

- G_k es la imagen correspondiente al nivel k ($\forall_{k>0}$), siendo G_0 la imagen original.
- w es el núcleo para la convolución con la gaussiana obtenida en el nivel anterior.
- $[\dots]_{\downarrow n}$ Es una operación de muestreo reductivo, en la que por cada fila se eliminan las siguientes $n-1$ filas, y $n-1$ columnas por cada columna.

En el ejemplo de fusión de imágenes en cuestión, las siguientes serían las imágenes de dichas pirámides:

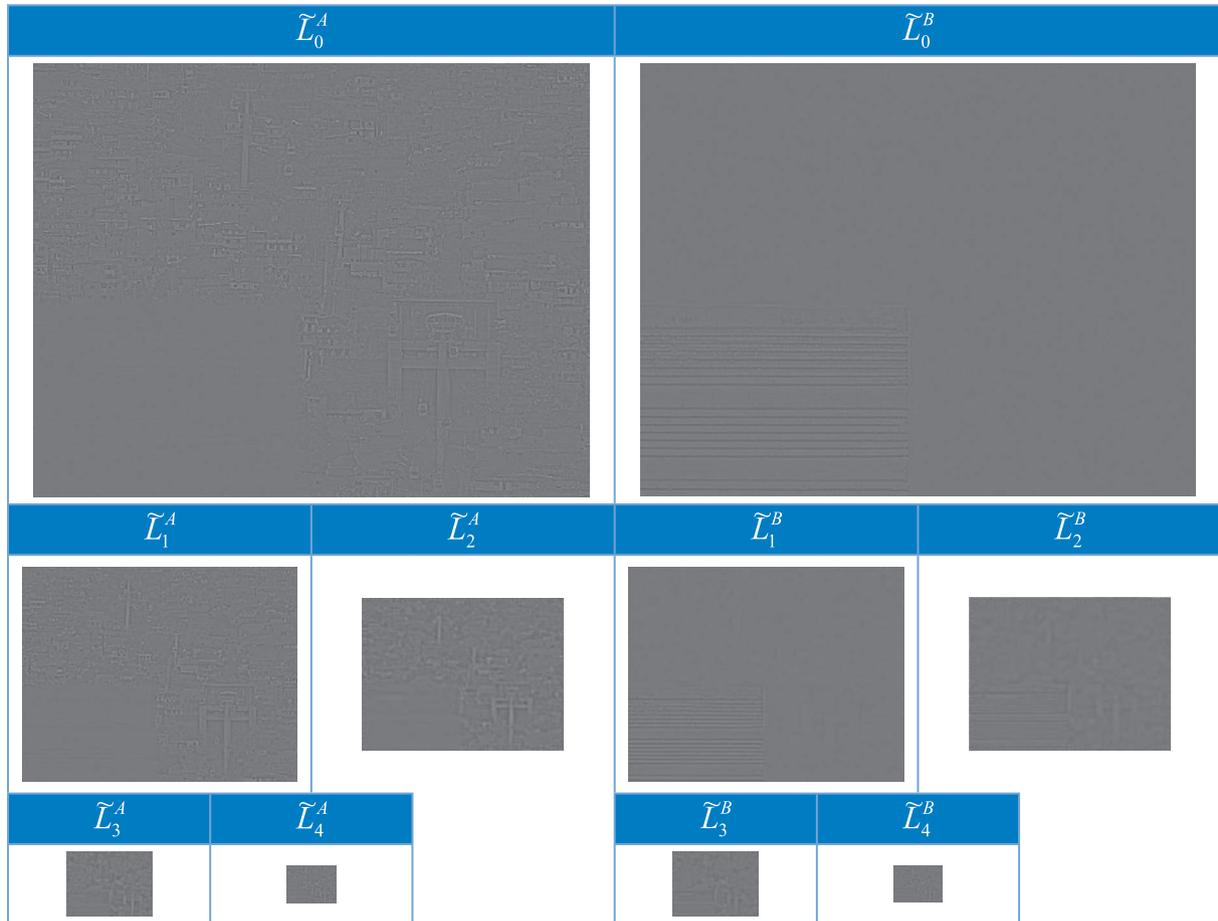


La pirámide laplaciana se calcula mediante la fórmula:

$$\tilde{L}_k = G_k - 4w^*[G_{k+1}]_{\uparrow 2}$$

Donde:

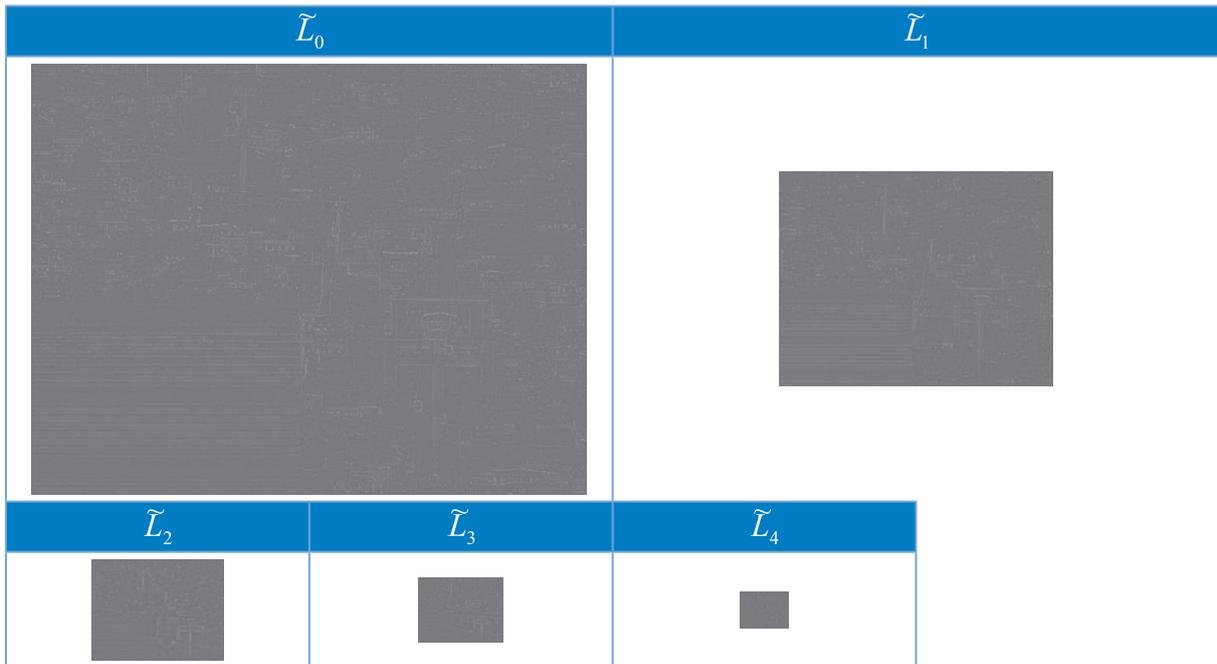
- \tilde{L}_k es la diferencia entre las gaussianas correspondientes al nivel k
- $[\dots]_{\uparrow n}$ Es una operación de muestreo expansivo en la que se insertan $n - 1$ filas entre cada fila y $n - 1$ columnas entre cada columna.



La reconstrucción de la imagen se realiza a partir de las pirámides laplacianas de cada una de las imágenes a fusionar, con el fin de obtener una sola.

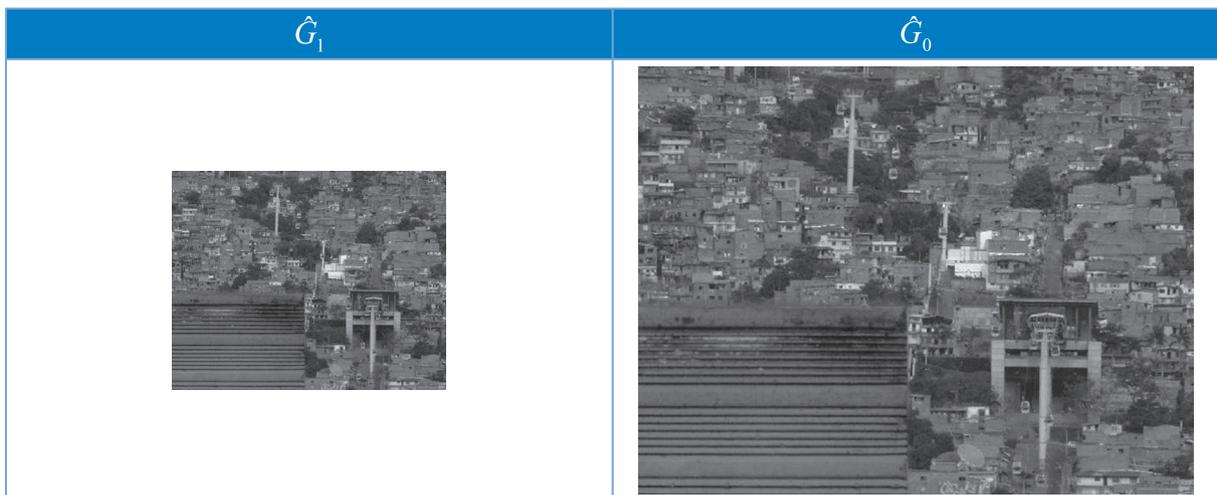
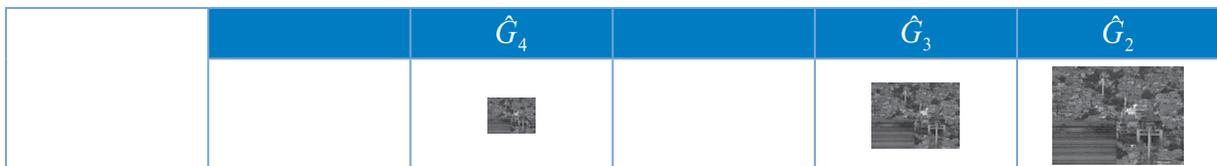
Para fusionar las laplacianas se utiliza la fórmula:

$$\tilde{L}_k^C(x, y) = \begin{cases} \tilde{L}_k^A(x, y) & \text{si } |\tilde{L}_k^A(x, y)| > |\tilde{L}_k^B(x, y)| \\ \tilde{L}_k^B(x, y) & \text{en otro caso} \end{cases}$$



Siendo \hat{G}_k la imagen reconstruida en el nivel k ($\forall_{k < N}$) y haciendo $\hat{G}_N = G_N^C = \frac{G_N^A + G_N^B}{2}$, la fórmula para obtener la imagen reconstruida por cada nivel sería:

$$\hat{G}_k = \tilde{L}_k^C + 4w * [\hat{G}_{k+1}]_{\uparrow 2}$$



Este método es el que ofrece los mejores resultados pero computacionalmente no es muy eficiente, razón por la cual se han tratado de crear otras metodologías que mejoren este aspectos y que se analizan a continuación.

Tenemos en primer lugar el método de la **pirámide laplaciana FSD** (Filtrar-Substraer-Diezmar) que mejora la carga computacional del anterior.

La pirámide laplaciana para cada imagen se calcula mediante la fórmula:

$$\tilde{L}_k = G_k - w * G_k$$

La cual no requiere muestreo expansivo. Esto hace cambiar la fórmula de reconstrucción a:

$$\hat{G}_k = \tilde{L}_k^C + w * (\tilde{L}_k^C + [4\hat{G}_{k+1}]_{\uparrow 2})$$

Otro método piramidal no se basa en las diferencias sino en las razones tratando de simular la sensibilidad del ojo humano a las variaciones en el contraste de una imagen más que a las variaciones en el valor absoluto. Por tal motivo, este método, denominado **pirámide del contraste**, utiliza divisiones en lugar de restas. En un nivel k determinado, la razón sería:

$$\tilde{R}_k = \frac{G_k}{4w * [G_{k+1}]_{\uparrow 2}}$$

La fórmula de reconstrucción sería:

$$\hat{G}_k = \tilde{R}_k^C \cdot 4w * [\hat{G}_{k+1}]_{\uparrow 2}$$

Los otros métodos piramidales se incluyen en una tabla final de resumen de algoritmos.

Otro método de fusión de imágenes es la **transformada discreta de wavelets** (DWI) (Pajares, 2004) donde no se requiere la pirámide gaussiana sino cuatro pirámides cada una correspondiente a los coeficientes de aproximación (bajas frecuencias de la imagen) y los coeficientes de detalle horizontales, verticales y diagonales (altas frecuencias en cada uno de los ejes o en ambos).

Por cada nivel k de la pirámide se deben calcular las siguientes matrices:

Matriz	Descripción
$LL_k = [w_L * L_k]_{1\downarrow 2}$	Esta es una subimagen lisa correspondiente a la banda de bajas frecuencias y puede ser considerado como una versión reducida de la original
$LH_k = [w_H * L_k]_{1\downarrow 2}$	Subimagen reducida de los detalles horizontales
$HL_k = [w_H * H_k]_{1\downarrow 2}$	Subimagen reducida de los detalles verticales
$HH_k = [w_H * H_k]_{1\downarrow 2}$	Subimagen reducida de los detalles diagonales

Donde:

- L_k es un proceso intermedio correspondiente a las matrices de coeficientes que se calcula como $L_k = [w_L * I]_{2 \downarrow 1}$
- H_k es un proceso intermedio que se calcula como $H_k = [w_H * I]_{2 \downarrow 1}$
- $[\dots]_{1 \downarrow 2}$ Es una operación de muestreo reductivo en el que se eliminan la mitad de las filas de la imagen, dejando intacta la cantidad de columnas.
- $[\dots]_{2 \downarrow 1}$ Es una operación de muestreo reductivo en el que se eliminan la mitad de las columnas de la imagen, dejando intacta la cantidad de filas.
- w_L Y w_H son los núcleos de paso bajo y paso alto equivalentes a $w_L = \frac{1}{\sqrt{2}} [1 \quad 1]$ y $w_H = \frac{1}{\sqrt{2}} [1 \quad -1]$

La reconstrucción de la imagen se hace con la siguiente ecuación:

$$\hat{I}_k = LL_k^{-1} + LH_k^{-1} + HL_k^{-1} + HH_k^{-1}$$

Finalmente, el siguiente es un cuadro comparativo de los métodos más representativos de la fusión de imágenes, algunos de los cuales fueron descritos en el presente documento:

Método	Categoría	Descripción	Desventajas
Media Aritmética	Simple	Suma las imágenes	No tiene en cuenta las áreas desenfocadas
Media Ponderada	Simple	Tiene en cuenta las áreas enfocadas	Produce “artefactos” en la imagen resultante
Pirámide Laplaciana ER	Piramidal	Elimina artefactos. Mejor calidad de imagen	Carga computacional alta
Pirámide Laplaciana FSD	Piramidal	Computacionalmente más eficiente	Incluye errores en la imagen fusionada
Pirámide del Contraste	Piramidal	No importa la eficiencia sino imitar el comportamiento del ojo humano	
Pirámide del gradiente	Piramidal	Utiliza además de las gaussianas, otras cuatro matrices que registran detalles en cada una de las orientaciones (horizontal, vertical, diagonal principal y secundaria)	No funciona bien con todas las imágenes
Pirámide morfológica	Piramidal	Se basa en operadores morfológicos	Sus resultados casi nunca son correctos
Transformada wavelets	Basado en transformadas	Computacionalmente eficiente	Sensible a desplazamientos de píxeles
Transformada wavelets invariante a cambios	Basado en transformadas	Corrige desplazamientos de píxeles en las imágenes	

Bibliografía

- Burt, P. J. (1993). Enhanced Image Capture Through Fusion. *IEEE* .
- France Laliberté, L. G. (2002). Registration and Fusion of Retinal Images: A Comparative Study. *International Conference on Pattern Recognition 2002*, 715-718.
- Haeberli, P. (1994). A Multifocus Method for Controlling Depth of Field.
- Haitao Li, J. Y. (n.d.). An image fusion scheme based on wavelet decomposition and SVR.
- Mertens, T. K. (2007). Exposure Fusion.
- Pajares, G. y. (2004). A wavelet-based image fusion tutorial. *Pattern recognition* , 1855-1872.
- Sadjadi, F. (n.d.). Comparative Image Fusion Analaysais. *IEEE* .
- Zhenhua Li, Z. J. (2004). Multi-sensor image fusion using discrete wavelet frame transform . *Chinese Optics Letters* , 578.