

Restauración de imágenes usando el criterio de minimización de rizado en la imagen restaurada

Domingo Mery y Marcela López
Universidad de Santiago de Chile
Departamento de Ingeniería Informática
Av. Ecuador 3659, Santiago de Chile
eMail: dmery@ieee.org, mpli@manquehue.net
<http://www.diinf.usach.cl/>

Abstract

In this paper a methodology to restore images that have been blurred by a horizontal (or vertical) movement is presented. The blur process is defined by the equation $\mathbf{Hf} = \mathbf{g}$, where \mathbf{g} is the row (or column) of the blurred image, \mathbf{H} is the blur matrix and \mathbf{f} is the original image. The restoration problem estimates \mathbf{f} from \mathbf{H} and \mathbf{g} . Nevertheless, since matrix \mathbf{H} has more columns than rows, there are an infinite number of exact solutions for \mathbf{f} that satisfy $\mathbf{Hf} = \mathbf{g}$. For this reason, an additional constraint for \mathbf{f} is required. In order to obtain a sharp estimation, we propose to minimize the ripple of \mathbf{f} . This assumption means that the component of the higher frequencies of \mathbf{f} are not so significant in the wanted solution. Thus, the restoration process corresponds to an optimization problem in which \mathbf{f} is estimated by minimizing the ripple subject to $\mathbf{Hf} = \mathbf{g}$, where \mathbf{H} and \mathbf{g} are a priori known. We present several ripple criteria, that are assessed after restoration performance and computational time. In addition, we propose a new way to solve directly this optimization problem.

Keywords: image processing, image restoration, blurred images, simulation, optimization.

Resumen

En este artículo se presenta una metodología para restaurar imágenes degradadas por un movimiento horizontal (o vertical) puro. El proceso de degradación está definido por la ecuación $\mathbf{Hf} = \mathbf{g}$, donde \mathbf{g} es la fila (o columna) de la imagen degradada, \mathbf{H} es la matriz de degradación y \mathbf{f} es la fila (o columna) de la imagen original no degradada. El problema consiste en encontrar \mathbf{f} a partir de \mathbf{H} y \mathbf{g} . Sin embargo, debido a que la matriz de degradación posee más columnas que filas, existen infinitas soluciones para \mathbf{f} que satisfacen $\mathbf{Hf} = \mathbf{g}$. Por esta razón es necesario emplear una condición adicional para \mathbf{f} , de tal forma que la solución encontrada se asemeje a la señal original no degradada. Para restaurar adecuadamente la imagen original, se sugiere en este artículo el uso del criterio de *minimización de rizado* en \mathbf{f} . Este criterio fue establecido a partir de un conjunto de soluciones de $\mathbf{Hf} = \mathbf{g}$, en las que se observó que las soluciones no deseadas presentan grandes oscilaciones. De esta manera, el proceso de restauración puede ser planteado como un problema de optimización en el que se debe encontrar \mathbf{f} minimizando su rizado, sujeto a la restricción $\mathbf{Hf} = \mathbf{g}$, conociendo a priori \mathbf{H} y \mathbf{g} . Se presentan y evalúan distintos criterios de minimización de rizado según costo computacional y desempeño en la restauración de imágenes con y sin ruido. Además se presenta un nuevo método directo (no iterativo) que resuelve el problema de optimización.

Palabras claves: procesamiento de imágenes, restauración de imágenes, imágenes degradadas, simulación, optimización.

1 Introducción

La restauración de imágenes consiste en recuperar, de la mejor manera posible la imagen original, a partir de su correspondiente imagen degradada y del conocimiento a priori de las causas de la degradación. El conocimiento de la degradación puede ser obtenido ya sea a partir de un modelo analítico o bien a partir de un modelo empírico [1, 4]. El campo de la restauración de imágenes ha comenzado a tener importancia desde 1950 aproximadamente con los programas espaciales, para poder mejorar las imágenes obtenidas del espacio donde la opción de volver a obtener una foto del espacio llega a tener un costo elevadísimo [5]. Hoy en día las aplicaciones relacionadas con la restauración de imágenes varían desde nuevas facilidades incorporadas a cámaras digitales comerciales hasta la introducción de nuevos algoritmos en el campo de la tele-medicina en la que se realiza un diagnóstico de imágenes tomadas a grandes distancias y que sufren degradaciones tanto en la comunicación como en la compresión de información [3].

En este artículo se considera que la degradación en toda la imagen ha sido causada por un mismo movimiento horizontal. Sin embargo, el enfoque también es válido para un movimiento vertical puro. Adicionalmente, empleando *matrices circulantes de bloque*, el método aquí explicado podría ser adaptado a degradaciones lineales e invariantes de posición que no se limitan a una sola fila o columna [5]. Sin embargo, debido al alto costo computacional que representa la utilización de estas matrices¹, en este artículo nos restringimos a una degradación uni-dimensional.

El proceso de degradación lineal puede ser modelado utilizando la teoría de sistemas lineales [8]. La entrada al sistema es la señal ideal f , agrupada en un vector $\mathbf{f} = [f_1 \dots f_M]^T$ de M elementos, mientras que la salida es la señal degradada g , agrupada en un vector $\mathbf{g} = [g_1 \dots g_N]^T$ de N elementos. Los vectores \mathbf{f} y \mathbf{g} representan entonces una de las filas de la imagen original y a su correspondiente fila de la imagen degradada. Este método es necesario repetirlo para cada una de las filas de la imagen. En esta teoría, la degradación se lleva a cabo mediante una convolución entre \mathbf{f} y \mathbf{h} , donde \mathbf{h} es un vector de n entradas que se define como la respuesta al impulso del proceso de degradación, o bien conocido como PSF (*point-spread function*) [5]. De esta manera, un elemento g_i del vector \mathbf{g} se calcula como la suma ponderada de n elementos de \mathbf{f} , es decir $g_i = h_1 f_i + h_2 f_{i+1} + \dots + h_n f_{i+n-1}$, para $i = 1, \dots, N$. Usando una matriz circulante, esta convolución puede ser escrita como $\mathbf{H}\mathbf{f} = \mathbf{g}$, o bien:

$$\mathbf{g} = \mathbf{f} * \mathbf{h} = \begin{bmatrix} h_1 & \dots & h_n & 0 & 0 & 0 & 0 \\ 0 & h_1 & \dots & h_n & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & h_1 & \dots & h_n \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{bmatrix} \quad (1)$$

Se aprecia en esta ecuación que $M = N + n - 1$. Este fenómeno, en que $M > N$, puede ser entendido más claramente empleando el siguiente ejemplo: si se toma una imagen con una cámara en movimiento horizontal, la cámara va a capturar información de una escena más ancha que la que hubiese tomado sin movimiento, entonces la imagen restaurada, que equivale al total de la escena, tiene más columnas que la imagen degradada.

Comúnmente, para una degradación correspondiente a un movimiento lineal uniforme de n píxeles, se definen los elementos de \mathbf{h} como constantes: $\mathbf{h} = [h_1 \dots h_n]^T$ donde $h_i = 1/n$. La Figura 1 muestra un ejemplo que ilustra este proceso de degradación. De esta forma, la imagen degradada corresponde a un

¹Para una imagen de $N \times N$ píxeles, es necesario definir matrices circulantes de bloque de $N^2 \times N^2$ elementos.

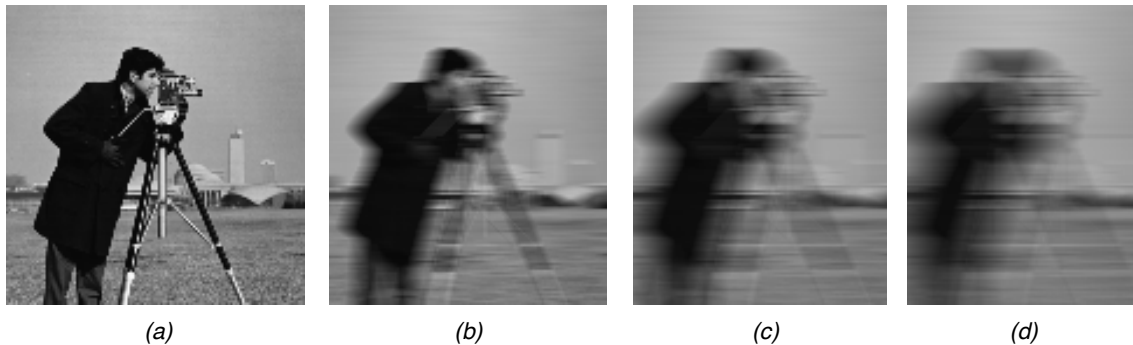


Figura 1: Degradación de la imagen *Cameraman* de 128×128 píxeles: a) imagen original y b), c) y d) imágenes degradadas por medio de un movimiento horizontal de $n = 10, 20$ y 30 píxeles respectivamente.

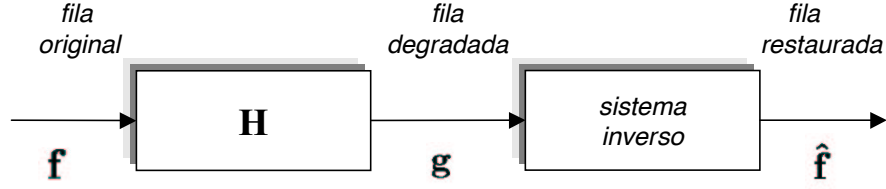


Figura 2: Modelo de solución.

promedio de n imágenes que han sido movidas cada una en un pixel. Generalmente, los elementos de \mathbf{h} deben cumplir con la condición $\sum h_i = 1$ para que la imagen degradada no sea ni más oscura ni más clara que la imagen original. La degradación para un movimiento horizontal puede ser simulada en MATLAB con el comando `G=conv2(F,h,'valid')`; donde \mathbf{G} y \mathbf{F} son las matrices de las imágenes degradada y original respectivamente, y \mathbf{h} es el vector de $1 \times n$ elementos que define la degradación. Para el movimiento lineal uniforme el vector de degradación es $\mathbf{h} = \text{ones}(1,n)/n$.

El problema de restauración consiste entonces en recuperar de la mejor manera posible el vector \mathbf{f} correspondiente a una fila de la imagen original, conociendo el vector \mathbf{g} correspondiente a la fila de la imagen degradada y el vector \mathbf{h} que define el proceso de degradación de la imagen. La idea general de solución se asemeja a la invertibilidad de sistemas lineales invariantes en el tiempo [8], donde al tener una señal degradada, producto de la convolución entre un efecto de distorsión y señal original, es posible restaurar la señal original realizando un sistema inverso sobre la señal degradada como se muestra en la Figura 2. La señal restaurada, para diferenciarla de la señal original, será denotada como $\hat{\mathbf{f}}$ y se almacenará en un vector $\hat{\mathbf{f}}$.

Recientemente se propuso en [7], un nuevo método de restauración basado en un criterio de similitud entre \mathbf{f} y \mathbf{g} . Este criterio se basa en una observación sencilla, que será explicada a continuación. De (1) se desprende que el sistema de ecuaciones $\mathbf{H}\mathbf{f} = \mathbf{g}$ tiene M variables y N ecuaciones. Dado que $M > N$, el sistema es subdeterminado y existen infinitas soluciones $\hat{\mathbf{f}}$ que satisfacen $\mathbf{H}\hat{\mathbf{f}} = \mathbf{g}$. Aún cuando estas soluciones son matemáticamente correctas, no corresponden en la gran mayoría de casos a la fila original \mathbf{f} . La Figura 3 muestra un ejemplo en el que se aprecia una señal \mathbf{f} de 50 muestras y su correspondiente degradación \mathbf{g} obtenida con $\mathbf{h} = [0.5 \ 0.5]^T$. En esta misma figura se grafican cuatro señales $\hat{\mathbf{f}}^1, \dots, \hat{\mathbf{f}}^4$, que a pesar de que satisfacen $\mathbf{H}\hat{\mathbf{f}}^k = \mathbf{g}$ no corresponden a la solución buscada. La observación mencionada es que las soluciones no deseadas presentan fuertes oscilaciones (*artefactos*), y la solución buscada no. Es por esta razón que una solución adecuada para $\hat{\mathbf{f}}$ sería aquella en que su rizado sea mínimo. Se introduce así una restricción de suavizado debido a que la gran mayoría de los píxeles de la imagen original no presentan gran variación en su intensidad con respecto a sus vecinos y esto quiere decir que la información de las imágenes está contenida en las frecuencias bajas y medias [3, 5].

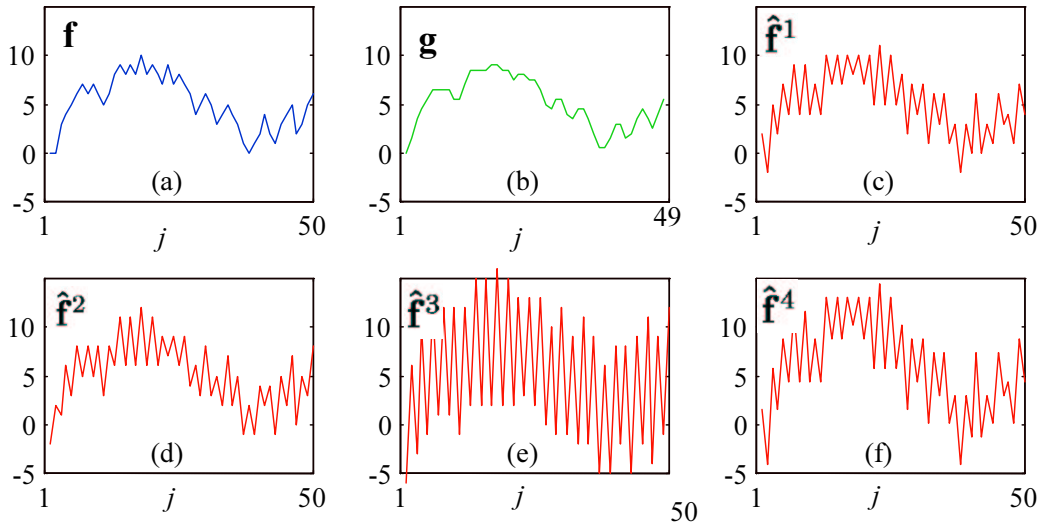


Figura 3: Restauración de una fila \mathbf{f} : a) fila original, b) fila degradada con $n = 2$, c), d) e) y f) cuatro posibles soluciones que satisfacen $\mathbf{H}\hat{\mathbf{f}} = \mathbf{g}$ [7].

En [7], esta condición de suavizado se obtiene forzando una similitud entre $\hat{\mathbf{f}}$ y \mathbf{g} . Sin embargo, $\hat{\mathbf{f}}$ y \mathbf{g} no tienen el mismo número de elementos. El criterio empleado es entonces la minimización de la norma $\|\hat{\mathbf{f}}_N - \mathbf{g}\|$, donde $\hat{\mathbf{f}}_N$ es definido como los primeros N elementos de $\hat{\mathbf{f}}$. Este método señala entonces que la imagen restaurada $\hat{\mathbf{f}}$ es aquella que minimice $\|\hat{\mathbf{f}}_N - \mathbf{g}\|$ sujeta a la restricción $\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\| = 0$. Esto no quiere decir que $\hat{\mathbf{f}}_N$ y \mathbf{g} sean iguales ya que esta solución no cumpliría necesariamente la restricción $\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\| = 0$. Mediante este criterio de similitud se escoge una de las infinitas soluciones de tal forma que se garantice la disminución del rizado. Matemáticamente, este problema de optimización se plantea usando la función objetivo

$$V(\hat{\mathbf{f}}) = \lambda \|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\|^2 + \|\mathbf{P}\hat{\mathbf{f}} - \mathbf{g}\|^2 \rightarrow \min, \quad (2)$$

donde λ es un número muy grande conocido como multiplicador de Lagrange [1], y la matriz \mathbf{P} de $N \times M$ es definida de tal forma que $\mathbf{P}\hat{\mathbf{f}} = \hat{\mathbf{f}}_N$, es decir:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ & & \vdots & & & \vdots & 0 \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix}. \quad (3)$$

Esta metodología será denominada en este trabajo como **minio**, ya que minimiza la norma de la diferencia entre la señal de entrada (*input*) y la señal de salida (*output*). La solución de (2) se obtiene derivando V con respecto a $\hat{\mathbf{f}}$ e igualando el resultado a cero [7]:

$$\hat{\mathbf{f}} = [\lambda \mathbf{H}^T \mathbf{H} + \mathbf{P}^T \mathbf{P}]^{-1} [\lambda \mathbf{H} + \mathbf{P}] \mathbf{g} \quad (4)$$

En el presente artículo se presentan otras restricciones que pueden ser empleadas para minimizar el rizado de $\hat{\mathbf{f}}$, algunas de ellas se basan en la similitud de $\hat{\mathbf{f}}_N$ y \mathbf{g} , mientras que otras se fundamentan en la minimización de los armónicos de $\hat{\mathbf{f}}$. Además se presenta una solución directa para el segundo caso que no depende del multiplicador de Lagrange. Este trabajo se ha estructurado de la siguiente manera. La Sección 2 presentan los criterios de similitud. En la Sección 3 se explican los criterios de frecuencia. Los experimentos y resultados se exponen en la Sección 4. Finalmente, se presentan las conclusiones en la Sección 5.

2 Criterios de similitud

En esta Sección se establecen otros criterios de similitud entre \mathbf{g} y $\hat{\mathbf{f}}_N$, donde $\hat{\mathbf{f}}_N$ se define como los primeros N elementos de $\hat{\mathbf{f}}$. Estos criterios pueden ser utilizados para restaurar una imagen que ha sido degradada por $\mathbf{H}\mathbf{f} = \mathbf{g}$. Los criterios de similitud se basan en la idea de que dos señales son similares si están correlacionadas entre sí. Existen varias métricas que sirven para evaluar la similitud entre dos señales, tres de ellas son la correlación, la covarianza y el coeficiente de correlación [8]. En nuestro caso, estas tres métricas tienen valores altos cuando \mathbf{g} y $\hat{\mathbf{f}}_N$ son similares.

La solución buscada es una señal $\hat{\mathbf{f}}_N$ que maximice la métrica de similitud y que al mismo tiempo minimice $\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\|$. Este problema puede plantearse como la minimización de una función objetivo cuyo numerador y denominador sean respectivamente $\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\|$ y la métrica de similitud. A continuación se definen las tres funciones objetivo correspondientes a las métricas de similitud mencionadas:

- Criterio de Similitud 1: Correlación

$$V(\hat{\mathbf{f}}) = \left[\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\| \right] \left[\|\mathbf{g}^T \hat{\mathbf{f}}_N\| \right]^{-1} \rightarrow \min \quad (5)$$

- Criterio de Similitud 2: Covarianza

$$V(\hat{\mathbf{f}}) = \left[\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\| \right] \left[\|\mathbf{g} - \mu_g\|^T [\hat{\mathbf{f}}_N - \mu_{f_N}] \right]^{-1} \rightarrow \min \quad (6)$$

donde μ_g y μ_{f_N} son los promedios de \mathbf{g} y \mathbf{f}_N respectivamente.

- Criterio de Similitud 3: Coeficiente de correlación

$$V(\hat{\mathbf{f}}) = \left[\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\| \right] \left[\frac{\|\mathbf{g} - \mu_g\|^T [\hat{\mathbf{f}}_N - \mu_{f_N}]}{\sigma_g \sigma_{f_N}} \right]^{-1} \rightarrow \min \quad (7)$$

donde σ_g y σ_{f_N} son las desviaciones estándares de \mathbf{g} y \mathbf{f}_N respectivamente.

Estas funciones objetivos, al ser no lineales, lamentablemente deben ser minimizadas con métodos iterativos. Esto significa, como se mostrará en los resultados, que su costo computacional es muy elevado².

²Si bien es cierto que la función de correlación (5) puede ser implementada de manera no iterativa con un multiplicador de Lagrange: $V(\hat{\mathbf{f}}) = \lambda \|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\|^2 - \|\mathbf{g}^T \mathbf{P}\hat{\mathbf{f}}\|^2 \rightarrow \min$, se optó por implementar la forma no iterativa debido a la dificultad que representa determinar las condiciones en las que la función objetivo es convexa.

3 Criterios de frecuencia

En la primera parte de esta Sección, se establecen algunos criterios que tienen como objeto reducir el rizado de la señal restaurada a partir de la disminución de sus armónicos. En la segunda parte, se propone una nueva forma directa de resolver el problema de restauración.

Existen diversas formas de plantear este problema, la forma más sencilla sin embargo es minimizando una señal $\hat{\mathbf{s}}$ obtenida a partir de un filtro pasa altos de $\hat{\mathbf{f}}$, es decir $\|\hat{\mathbf{s}}\| = \|\hat{\mathbf{f}} * \mathbf{w}\| = \|\mathbf{W}\hat{\mathbf{f}}\| \rightarrow \min$, donde \mathbf{w} es el vector que representa la respuesta al impulso del filtro pasa altos, y \mathbf{W} es la matriz circulante respectiva definida a partir de \mathbf{w} de la misma manera como \mathbf{H} fue definido a partir de \mathbf{h} en (1). De esta forma la función objetivo a minimizar es:

$$V(\hat{\mathbf{f}}) = \lambda \|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\|^2 + \|\mathbf{W}\hat{\mathbf{f}}\|^2 \rightarrow \min \quad (8)$$

En este planteamiento se utiliza el multiplicador de Lagrange λ conocido en la literatura también como el parámetro de regularización. El fin de este parámetro es mantener un equilibrio entre la fidelidad de la información expresada como $\|\mathbf{H}\hat{\mathbf{f}} - \mathbf{g}\|$ y el suavizado de la solución expresado como $\|\mathbf{W}\hat{\mathbf{f}}\|$ [3]. La minimización de $V(\hat{\mathbf{f}})$ se obtiene derivando V con respecto a $\hat{\mathbf{f}}$ e igualando su resultado a cero, es decir:

$$\hat{\mathbf{f}} = [\lambda \mathbf{H}^T \mathbf{H} + \mathbf{W}^T \mathbf{W}]^{-1} \mathbf{H}^T \mathbf{g}. \quad (9)$$

En este trabajo se investigan distintas formas de minimizar el rizado de $\hat{\mathbf{f}}$. La primera de ellas, que llamaremos **mindft**, es incluyendo una restricción que minimiza los armónicos de alta frecuencia en el dominio de la frecuencia. La transformada discreta de Fourier (DFT) de $\hat{\mathbf{f}}$ se puede obtener mediante $\hat{\mathbf{F}}_{DFT} = \mathbf{W}_{DFT} \hat{\mathbf{f}}$, donde \mathbf{W}_{DFT} es una matriz de tamaño $M \times M$ cuyo elemento (i, k) está definido por $\exp(-2\pi i k j / M)$ con $j = \sqrt{-1}$ [5]. La señal $\hat{\mathbf{f}}$ puede ser filtrada ponderando los elementos $\hat{\mathbf{F}}_{DFT}$, es decir multiplicando $\hat{\mathbf{F}}_{DFT}$ por una matriz diagonal \mathbf{Q} de $M \times M$ cuyo elemento (i, i) es el factor de ponderación correspondiente a la frecuencia discreta i . Lo que se pretende minimizar en este planteamiento es la energía de la señal filtrada, esto es $\|\mathbf{Q}\mathbf{W}_{DFT} \hat{\mathbf{f}}\| \rightarrow \min$. Esto se consigue utilizando (9) con $\mathbf{W} = \mathbf{Q}\mathbf{W}_{DFT}$. Con el objeto de evitar el efecto *ringing*, los coeficientes $Q(i, i)$ deben corresponder a una ventana de cambios suaves [5]. Una posibilidad es escoger una ventana Gaussiana de desviación estándar σ_F cuyo máximo corresponda a la frecuencia más alta de $\hat{\mathbf{F}}_{DFT}$, es decir $\mu_F = M/2 + 1$.

Alternativamente, con el objeto de evitar el uso de números complejos, se puede emplear en vez de la DFT, la transformada discreta de cosenos (DCT) [5]. Este método será denominado **mindct**. Como se verá en la Sección 4, los resultados son muy similares pero el costo computacional disminuye.

Otra posibilidad es no realizar el filtro pasa altos en el dominio de la frecuencia, sino en el dominio del espacio. Una alternativa es definir \mathbf{w} como una ventana de Laplace, que como bien se sabe realiza un filtro pasa altos. Existen diversas variantes de esta ventana, en este trabajo se utilizarán dos: la primera, denominada **minL3**, es una ventana simple $\mathbf{w} = (-1, 2, -1)^T$ y la segunda, que llamaremos **minLoG**, es una ventana que corresponde al filtro Laplacian-of-Gauss (LoG) [5] definido como:

$$w_i = \left(1 - \frac{i^2}{2\sigma^2}\right) \exp\left(-\frac{i^2}{2\sigma^2}\right). \quad (10)$$

Por otra parte, se observa que el criterio de similitud propuesto en [7] –y planteado en (2)–, en el que se pretende minimizar la norma de la diferencia existente entre $\hat{\mathbf{f}}_N$ y \mathbf{g} , corresponde a definir \mathbf{W} en la ecuación (9) como $\mathbf{W} = \mathbf{P} - \mathbf{H}$, ya que $\hat{\mathbf{f}}_N - \mathbf{g} = \mathbf{P}\hat{\mathbf{f}} - \mathbf{H}\hat{\mathbf{f}}$. Esto quiere decir que la respuesta al impulso de este filtro pasa altos sería $\mathbf{w} = [(1 - h_1) \ (-h_2) \ \dots \ (-h_n)]^T$. Este método será denominado **minio2**.

A continuación se discutirá un último criterio para disminuir el rizado de $\hat{\mathbf{f}}$. Debido a que las fuertes oscilaciones en $\hat{\mathbf{f}}$ incrementan su norma, en este método simplemente se minimiza $\|\hat{\mathbf{f}}\|$. Esto correspondería a usar un filtro *pasa todo*, con el que no se filtra $\hat{\mathbf{f}}$, sino que se fuerza a que la solución tenga una norma mínima, y así se hace disminuir el rizado. Este método corresponde a minimizar (8) con $\mathbf{W} = \mathbf{I}$, donde \mathbf{I} es la matriz identidad de $M \times M$. Esta solución es conocida como la pseudo-inversa $\hat{\mathbf{f}} = \mathbf{H}^+ \mathbf{g}$ [5], ya que $\mathbf{H}\mathbf{H}^+ = \mathbf{I}$. MATLAB calcula esta pseudo-inversa a partir de los valores propios de $\hat{\mathbf{f}}$ utilizando la fórmula de Monroe. Este método lo llamaremos **pinv**. Sin embargo, como se verá en la Sección 4, el uso de (9) con $\mathbf{W}^T \mathbf{W} = \mathbf{I}$:

$$\hat{\mathbf{f}} = [\lambda \mathbf{H}^T \mathbf{H} + \mathbf{I}]^{-1} \mathbf{H}^T \mathbf{g} \quad (11)$$

disminuye considerablemente el costo computacional. Este método se denotará como **pinv2**.

Finalmente, se propone en esta sección una nueva forma de resolver el problema de restauración. Se observa que la solución directa (9) depende del parámetro de regularización λ . Escoger este parámetro no es trivial [3] y existen varias investigaciones al respecto (ver por ejemplo [2]). A continuación se presenta una

Table 1: Métodos de restauración usando criterios de frecuencia

Método	Solución	Observaciones
<code>minio</code>	Ecuación (2)	
<code>minio2</code>	Ecuación (9) ó (16)	$\mathbf{W} = \mathbf{P} - \mathbf{H}$
<code>mindft</code>	Ecuación (9) ó (16)	$\mathbf{W} = \mathbf{Q}\mathbf{W}_{DFT}$
<code>mindct</code>	Ecuación (9) ó (16)	$\mathbf{W} = \mathbf{Q}\mathbf{W}_{DCT}$
<code>minL3</code>	Ecuación (9) ó (16)	\mathbf{W} circulante definida con $\mathbf{w} = (-1, 2, -1)^\top$
<code>minLoG</code>	Ecuación (9) ó (16)	\mathbf{W} circulante definida con $\mathbf{w} = [w_i]$ de ecuación (10)
<code>pinv</code>	$\hat{\mathbf{f}} = \text{pinv}(\mathbf{H})\mathbf{g}$	<code>pinv</code> es un comando MATLAB
<code>pinv2</code>	Ecuación (9) ó (16)	$\mathbf{W}^\top \mathbf{W} = \mathbf{I}$

solución directa que es independiente de λ . La solución parte de la idea de que los últimos $n - 1$ elementos de \mathbf{f} son iguales a las variables $\theta_1, \dots, \theta_{n-1}$. De esta manera, (1) puede ser planteada de la siguiente manera:

$$\begin{bmatrix} h_1 & h_2 & \dots & h_n & 0 & 0 & 0 & 0 & 0 \\ 0 & h_1 & h_2 & \dots & h_n & 0 & 0 & 0 & 0 \\ & \vdots & & & \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & h_1 & h_2 & \dots & h_n & 0 \\ 0 & 0 & 0 & \dots & 0 & h_1 & h_2 & \dots & h_n \\ \hline 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ & \vdots & & & \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \\ f_{N+1} \\ f_{N+2} \\ \vdots \\ f_M \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{N-1} \\ g_N \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n-1} \end{bmatrix} \quad (12)$$

o bien matricialmente como $\mathbf{R}\mathbf{f} = [\mathbf{g}^\top \theta^\top]^\top$. Así se puede escribir la solución para \mathbf{f} como:

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_N \\ \theta \end{bmatrix} = \mathbf{R}^{-1} \begin{bmatrix} \mathbf{g} \\ \theta \end{bmatrix} = \mathbf{S}\mathbf{g} + \mathbf{T}\theta \quad (13)$$

donde \mathbf{S} y \mathbf{T} son submatrices de \mathbf{R}^{-1} definidas como las primeras N columnas y las últimas $(n-1)$ columnas de \mathbf{R}^{-1} respectivamente. La búsqueda de $\hat{\mathbf{f}}$ que minimice $\|\mathbf{W}\hat{\mathbf{f}}\|$, se puede plantear ahora a partir de (13) definiendo una nueva función objetivo:

$$V(\hat{\mathbf{f}}) = \|\mathbf{W}\hat{\mathbf{f}}\|^2 = V(\hat{\theta}) = \|\mathbf{W}\mathbf{S}\mathbf{g} + \mathbf{W}\mathbf{T}\hat{\theta}\|^2 \rightarrow \min \quad (14)$$

En este planteamiento está implícita la restricción $\mathbf{H}\hat{\mathbf{f}} = \mathbf{g}$. Esto quiere decir que el problema de restauración se reduce a encontrar $\hat{\theta}$ tal que se minimice $\|\mathbf{W}\mathbf{S}\mathbf{g} + \mathbf{W}\mathbf{T}\hat{\theta}\|^2$ siendo \mathbf{W} , \mathbf{S} , \mathbf{T} y \mathbf{g} conocidos. Este problema se resuelve por medio de la técnica de mínimos cuadrados. En esta técnica se encuentra $\hat{\theta}$ tal que $\|\mathbf{X}\theta - \mathbf{y}\|^2 \rightarrow \min$ conociendo \mathbf{X} e \mathbf{y} . En este caso la solución es $\hat{\theta} = [\mathbf{X}^\top \mathbf{X}]^{-1} \mathbf{X}^\top \mathbf{y}$ [5], en nuestro caso la solución quedaría

$$\hat{\theta} = -[\mathbf{T}^\top \mathbf{W}^\top \mathbf{W} \mathbf{T}]^{-1} \mathbf{T}^\top \mathbf{W}^\top \mathbf{W} \mathbf{S} \mathbf{g}. \quad (15)$$

Conociendo $\hat{\theta}$ se determina $\hat{\mathbf{f}}$ por sustitución directa en (13):

$$\hat{\mathbf{f}} = (\mathbf{I} - \mathbf{T}[\mathbf{T}^\top \mathbf{W}^\top \mathbf{W} \mathbf{T}]^{-1} \mathbf{T}^\top \mathbf{W}^\top \mathbf{W}) \mathbf{S} \mathbf{g}. \quad (16)$$

Resumiendo, en esta Sección se han discutido seis métodos no iterativos de restauración: `mindft`, `mindct`, `minL3`, `minLoG`, `minio2`, `pinv` y `pinv2` y dos maneras distintas de resolver el problema de optimización, la primera usando multiplicadores de Lagrange, la segunda mediante una deducción directa. Las soluciones explicadas se presentan en la Tabla 1.

4 Experimentos y resultados

En esta Sección se presentan los experimentos realizados y los resultados obtenidos. La metodología a emplear es utilizando una imagen original y un proceso de degradación conocido. La imagen original será de $L \times M$ elementos y se almacenará en la matriz \mathbf{F} . La imagen degradada tendrá $L \times N$ elementos y se almacenará en la matriz \mathbf{G} . La restauración de \mathbf{F} se denominará $\hat{\mathbf{F}}$ y será del mismo tamaño que \mathbf{F} .

Para evaluar el desempeño de la restauración se utilizará el valor $ISNR$ (*improvement in signal-to-noise-ratio*) definido como [3]:

$$ISNR = 20 \log_{10} \left(\frac{\| \mathbf{F}_N - \mathbf{G} \|}{\| \mathbf{F} - \hat{\mathbf{F}} \|} \right) = 20 \log_{10} \left(\frac{\sum_{i=1}^L \sum_{j=1}^N (F_{ij} - G_{ij})^2}{\sum_{i=1}^L \sum_{j=1}^M (F_{ij} - \hat{F}_{ij})^2} \right) \quad (17)$$

En este criterio es necesario utilizar \mathbf{F}_N , definido como las primeras N columnas de \mathbf{F} , para poder compararlo con \mathbf{G} que tiene sólo N columnas. Se observa que el $ISNR$ es grande cuando la restauración es buena, y tiende a infinito cuando \mathbf{F} y $\hat{\mathbf{F}}$ son iguales. Este criterio sólo puede ser aplicado a problemas de restauración en los que la imagen degradada ha sido simulada, ya que es necesario conocer la imagen original \mathbf{F} .

Para evaluar el costo computacional se medirá el tiempo de ejecución en un computador Intel(R) Pentium(R) 4 CPU 1500 MHz 256 KB RAM. Todos los programas han sido ejecutados en ambiente MATLAB Versión 6.1 (R12.1).

En la primera parte de esta Sección se compararán los tres métodos iterativos de la Sección 2 versus dos métodos directos representativos de la Sección 3. Debido al alto costo computacional que representan estas soluciones no iterativas, las pruebas se realizaron sólo en la imagen *Cameraman* de 128×128 mostrada en la Fig. 1. En estos experimentos se probó con degradación lineal uniforme, es decir con $\mathbf{h} = [h_1 \dots h_n]^T$ donde $h_i = 1/n$ para $n = 10, 20, 30$ y 40 píxeles. En estos experimentos los tres métodos iterativos obtuvieron desempeños muy paracidos. Los resultados del desempeño se pueden apreciar en la Fig. 4. Se observa que los valores $ISNR$ entre los métodos iterativos y los directos son similares. Cabe destacar que para la vista humana no hay diferencia entre las distintas restauraciones. En lo que sí hay una diferencia significativa es en el tiempo de ejecución. A manera de ejemplo, para restaurar la imagen degradada con $n = 20$ los tres métodos iterativos necesitaron: 38.16, 47.40 y 51.30 minutos, mientras que los otros métodos obtuvieron la imagen restaurada en menos de 0.3 segundos³.

Dado que los métodos iterativos son demasiado lentos y tienen un desempeño similar a otros métodos directos, el segundo experimento se llevo a cabo sólo con los métodos directos de la Sección 3. Debido a que estos métodos son más rápidos se pudieron realizar las pruebas en más imágenes simulando por una parte un ruido Gaussiano de media cero y desviación estándar s variable, y por otra parte la discretización de la señal degradada usando sólo 8 bits. Esta última simulación quiere decir que el valor de gris de un píxel de la señal degradada, que es calculado como un promedio de n píxeles de la imagen original, es redondeado al entero más próximo. Esto, porque en la realidad las imágenes digitales de 8 bits por ejemplo sólo almacenan valores enteros de intensidad entre 0 y 255.

Las imágenes originales empleadas se muestran en la Fig. 5. Son todas de 256×256 píxeles. En la figura se aprecia también el efecto de degradaciones con $\mathbf{h} = [h_1 \dots h_n]^T$ donde $h_i = 1/n$ para $n = 3, 10, 20, 40, 50$ y 64 píxeles⁴. Métodos para estimar n en este caso se encuentran en [7, 9].

La Figura 6a muestra los resultados obtenidos en el desempeño de 10 métodos de restauración para la situación en la que no hay ruido y no hay discretización. En estos resultados se promedia el valor de $ISNR$

³Ambos métodos fueron programados usando la mayor cantidad de funciones compiladas de MATLAB. Para los métodos iterativos se utilizó el comando `fminsearch`, mientras que para los métodos no iterativos se emplearon las funciones de manipulación de matrices que ofrece MATLAB. Por estas razones, es posible realizar comparaciones de los tiempos de ejecución.

⁴Si bien es cierto que todos los experimentos aquí presentados se hicieron para degradaciones de tipo movimiento lineal uniforme, los métodos explicados en este artículo pueden ser utilizados para otro valores de h_i no constantes. Experimentos realizados con nuestro programa de pruebas muestran que la restauración es posible si \mathbf{h} representa un filtro pasa bajos. Esto es una gran ventaja frente a otros métodos que están restringidos a h_i constante (ver por ejemplo [9]).

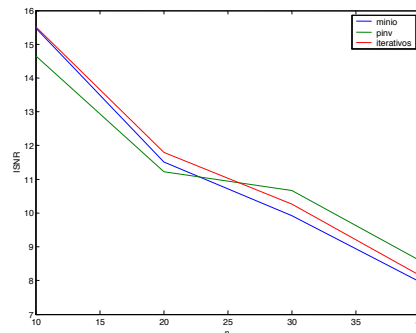


Figura 4: Comparación del desempeño de la restauración entre los métodos iterativos y dos métodos directos (*minio* y *pinv*).

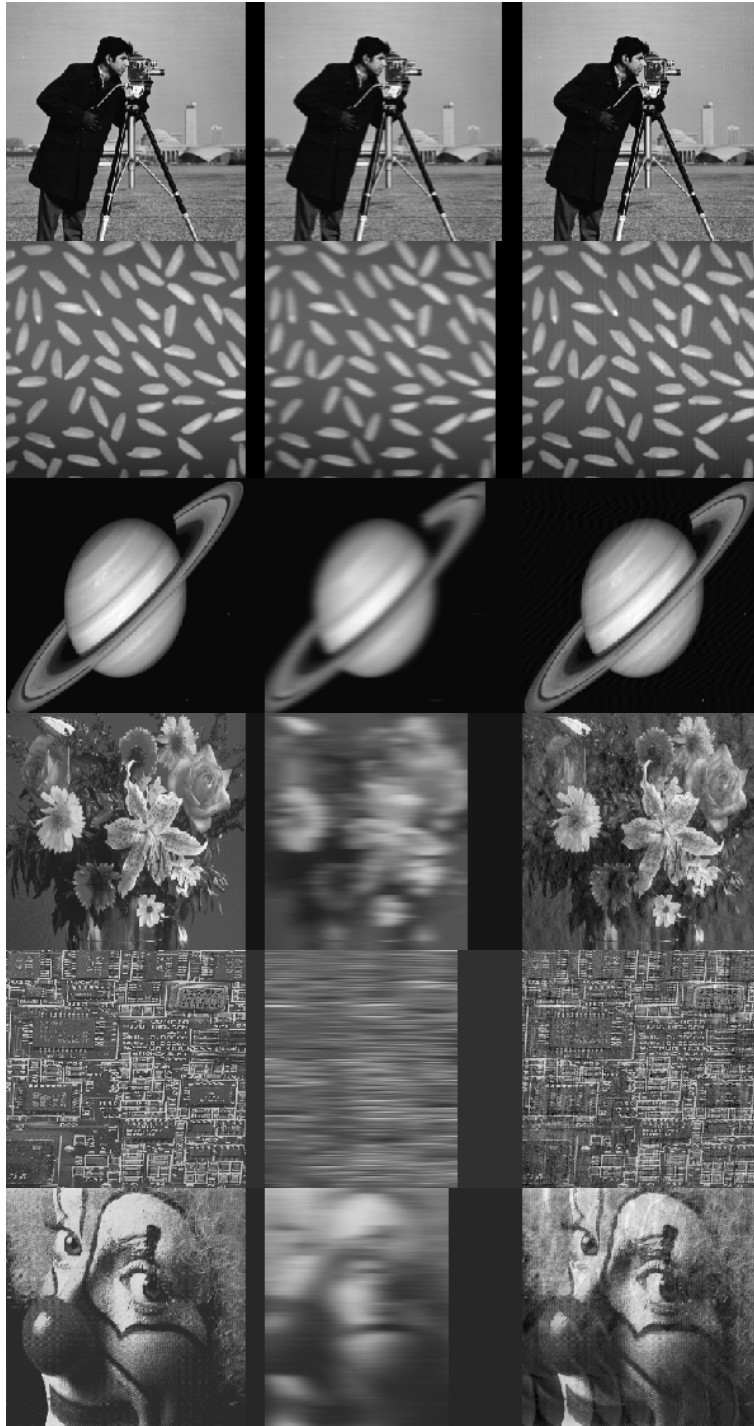


Figura 5: Restauración en 6 imágenes usando el método de la pseudo-inversa para $n = 3, 10, 20, 40, 50$ y 64 . Las tres columnas representan las imágenes originales, degradadas y restauradas.

obtenido en las seis imágenes. Los métodos, resumidos en la Tabla 1 son: `minio`, `minio2`, `minio-dir`, `pinv`, `pinv2`, `pinv-dir`, `minL3`, `minLoG`, `mindft/dct1` y `mindft/dct2`. El sufijo `-dir` significa que en vez de usar la ecuación (9) para la minimización se utilizó la ecuación propuesta en (16). Por otra parte, como `mindft` y `mindct` obtuvieron el mismo resultado se graficaron juntos. Sin embargo la diferencia entre `mindft/dct1` y `mindft/dct2` consiste en que la ventana de Gauss utilizada en la ponderación (ver Sección 3) para el primer caso fue con $\sigma_F = M/4$ y para el segundo fue $\sigma_F = M/2$. Se observa que el peor desempeño fue obtenido por `minL3` seguido por `minLoG`. Asimismo, el primer lugar fue obtenido por `pinv` y `pinv-dir`, obteniendo el mismo desempeño, seguido por `minio-dir`.

La Figura 7 muestra los desempeños para 6 niveles de ruido Gaussiano ($s = 0, 0.1, 0.25, 0.5, 1$ y 3) y con

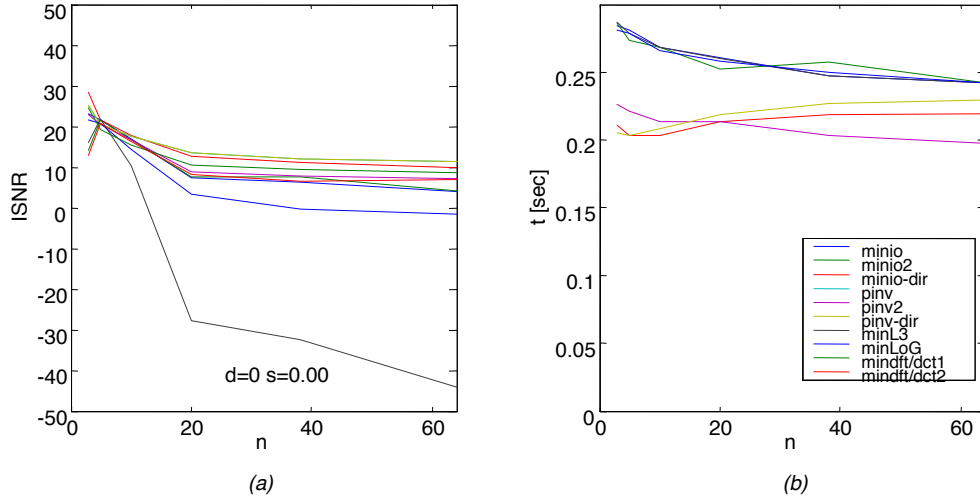


Figura 6: Desempeño y tiempo de ejecución de 10 métodos de restauración para $d = 0$ (sin discretización) y $s = 0$ (sin ruido).

y sin discretizar la imagen degradada ($d = 1$ y $d = 0$ respectivamente). De esta figura sólo podemos apreciar la tendencia. Los resultados por orden de desempeño son similares, los primeros lugares los obtuvieron los métodos **pinv-x** o **minio-x**, y los últimos los métodos **minL3** y **minLoG**. Se aprecia el deterioro del desempeño al incrementar el valor de ruido s o al simular la discretización.

El tiempo de ejecución es independiente del nivel de ruido y discretización, por esta razón sólo se presentan los tiempos para la situación ideal $s = 0$ y $d = 0$ (ver Fig. 6b). El menor tiempo de ejecución fue logrado por el método **pinv2** (0.20 s), seguido por **minio-dir** (0.22 s) y **pinv-dir** (0.23 s). Resulta interesante comparar estos resultados con el método propuesto por MATLAB (**pinv**) en el que tiempo de ejecución es 0.60 s.

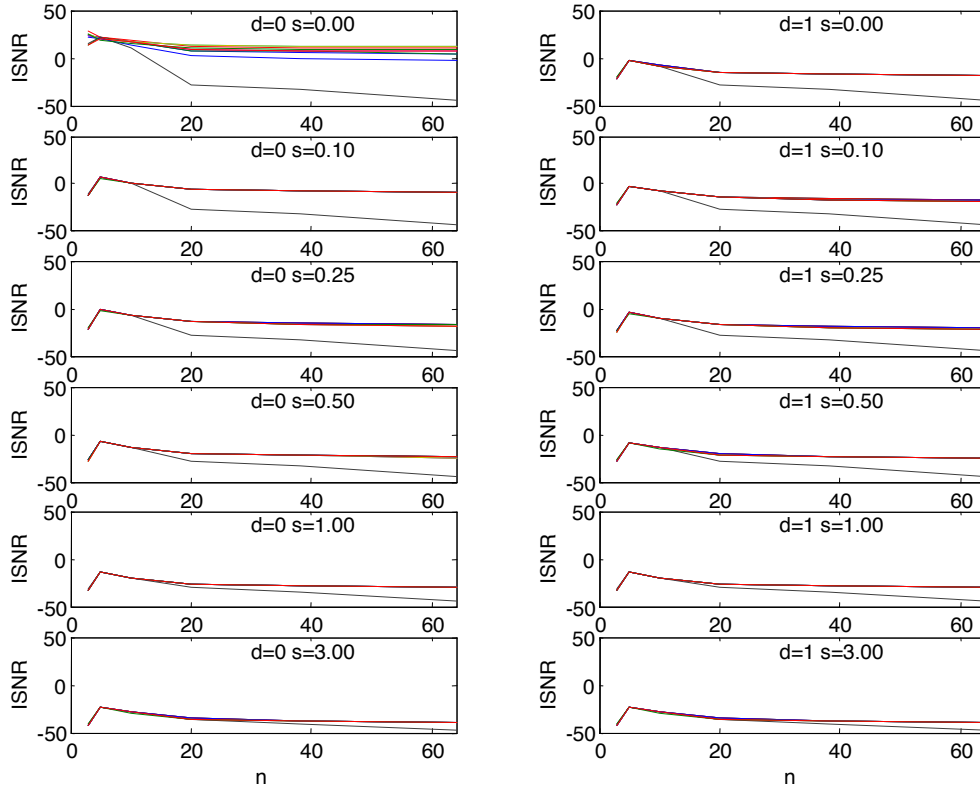


Figura 7: Desempeño de 10 métodos de restauración para ruido (s) y discretización (d) variables.

5 Conclusiones

En este artículo se han evaluado según desempeño y tiempo de ejecución dos tipos de metodologías basadas en la minimización de rizado para restaurar imágenes degradadas por un movimiento horizontal invariante de posición. La primera metodología consiste en maximizar métricas de correlación de la señal restaurada y la señal degradada. Si bien es cierto que los resultados de desempeño son buenos (y similares a otros métodos) el tiempo requerido en la restauración es tan elevado que su uso no resulta recomendable. La segunda metodología se basa en la minimización de la norma de la ponderación de la señal restaurada con el fin de minimizar su rizado. Se proponen distintas ponderaciones, algunas de ellas derivadas de un planteamiento en el dominio de la frecuencia, y otras en el dominio del espacio. Para este problema se presentan dos tipos de soluciones: la primera, conocida en la comunidad científica de restauración de imágenes como restauración directa regularizada, resuelve el problema de optimización por medio del multiplicador de Lagrange; mientras que la segunda solución, sugerida en este trabajo, logra simplificar el problema de optimización a una minimización dependiente de menos variables y presenta una solución directa sin multiplicadores de Lagrange, basada en una técnica de mínimos cuadrados.

En los experimentos se utilizaron 6 imágenes en blanco y negro de 256×256 píxeles, y se simularon degradaciones con 6 movimientos horizontales ($n = 3, 10, 20, 40, 50$ y 64 píxeles), 6 niveles de ruido Gaussiano ($s = 0, 0.1, 0.25, 0.5, 1$ y 3) y con y sin discretizar la imagen degradada. Esto significa que en total fueron probadas 432 imágenes degradadas. A su vez para cada una de ella fue restaurada usando 10 métodos distintos, los que fueron evaluados según desempeño y tiempo de ejecución. En estas pruebas se obtuvo que los métodos basados en la minimización de la norma de la señal restaurada (métodos `pinv-x`), y aquellos basados en la minimización de la norma de la diferencia entre la señal restaurada y la señal degradada (métodos `minio-x`), fueron los de mejor desempeño. Es difícil determinar cuál de ellos es el mejor ya que los desempeños fueron muy similares. Sin embargo, se observó que la solución propuesta basada en una optimización directa presenta una buena alternativa debido a un tiempo de ejecución bajo, casi un tercio de la pseudo-inversa de Monroe propuesta por MATLAB.

Como trabajo futuro se pretende comparar estas técnicas con otras conocidas como las basadas en la variación total (método de Rudin-Osher-Fatemi) y *wavelet thresholding* [6].

Agradecimientos

Los autores agradecen el apoyo brindado por el Departamento de Ingeniería Informática y el Departamento de Investigaciones Científicas y Tecnológicas de la Universidad de Santiago de Chile (proyecto 06-0119MQ). Asimismo, se agradece los comentarios realizados por los revisores anónimos.

Referencias

- [1] H. Andrews and B. Hunt. *Digital Restoration*. Englewood Cliffs, Prentice-Hall, NJ, 1977.
- [2] M.R. Banham and A.K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *Trans. on IEEE Image Processing*, 1(3):322–336, 1992.
- [3] M.R. Banham and A.K. Katsaggelos. Digital image restoration. *IEEE Signal Processing Magazine*, 14(2):24–40, 1997.
- [4] R. Bates and M. McDonnell. *Restoration and Reconstruction*. The Oxford engineering science series; 16, Oxford University Press, NY, 1986.
- [5] K.R. Castleman. *Digital Image Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1996.
- [6] F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Trans. Image Processing*, 11(12):1450–1456, 2003.
- [7] D. Mery and D. Filbert. A fast non-iterative algorithm for the removal of blur caused by uniform linear motion in X-ray images. In *Proceedings of the 15th World Conference on Non-Destructive Testing (15th-WCNDT)*, Rome, Oct. 15-21 2000.
- [8] J.G. Proakis and D.G. Manolakis. *Digital signal processing*. Prentice Hall, New Jersey, 1996.
- [9] M. Sordhi. Image restoration: The removal of spatially invariant degradations. *Proc. IEEE*, 60(7):842–853, 1972.