# Inpainting and Zooming using Sparse Representations

Enrique Marcelo Albornoz
*Fac. de Ingeniería y Cs. Hídricas, emalbornoz@fich.unl.edu.ar*

## I. Introducción

El inpainting es una técnica de restauración, puede verse como una interpolación o una desoclusión de píxeles. La clave está en rescatar información relevante de los píxeles observados y utilizar esto para inferir robustamente los datos perdidos. Suponga una imagen completa $\mathbf{x}$ definida en un dominio finito $\Omega$ (el plano), y su versión degradada $\mathbf{y}$ (aunque no completamente observada). La imagen observada (incompleta) $\mathbf{y}_{obs}$ es el resultado de aplicar el operador de pérdida $\mathcal{M}$ sobre $\mathbf{y}$: $\mathcal{M} : \mathbf{y} \mapsto \mathbf{y}_{obs} = \mathcal{M}[\mathbf{y}] = \mathcal{M}[\mathbf{x} \odot \varepsilon]$, donde $\odot$ indica una composición y $\varepsilon$ es ruido. Ejemplo de $\mathcal{M}$ son las máscaras binarias. Con inpainting se recupera $\mathbf{x}$ a partir de $\mathbf{y}_{obs}$, y esto es un problema inverso mal condicionado. Desde un punto de vista estadístico, los procesos de inpainting/interpolación de imágenes pueden verse como un problema de estimación de datos perdidos. En este trabajo, se propone la idea de usar un algoritmo de maximización de la esperanza (*EM*) en un marco Bayesiano para atacar este problema. Está basado en una versión penalizada de la máxima verosimilitud, formulada usando representaciones ralas lineales ($\mathbf{x} = \mathbf{\Phi}\alpha$). Se impone una penalización de la distribución a priori, que promueve la dispersión, sobre los coeficientes reconstruidos. El marco del algoritmo *EM* da los principios para formalizar la idea de que los datos perdidos pueden ser recuperados basándose en representaciones ralas. Se introduce un algortimo iterativo eficiente y se derivan sus propiedades de convergencia teóricas. Éste supera a sus competidores en cuanto permite un alto grado de fexibilidad para recuperar componentes de diversas estrucutras en la imagen. Se sugieren formas de ajuste automáticas para el parámetro de regularización.

### *Representaciones ralas y elección del diccionario*

Se supone una imagen $x$ de $\sqrt{n} \times \sqrt{n}$ que se puede escribir como la superposición de (unas pocas) funciones elementales $\phi_\gamma(s)$ parametrizadas por $\gamma \in \Gamma$ tal que:

$$x(s) = \sum_{\gamma \in \Gamma} \alpha_\gamma \phi_\gamma(s), \qquad \# \Gamma = L \tag{1}$$

Usualmente en casos prácticos, el diccionario ($\Phi$) se construye con la unión de una o varias (suficientemente incoherentes) transformaciones. En general cada una de éstas es una *base ortogonal* o un *marco ajustado*. Cuando las transformaciones son unidas en un sólo diccionario, serán elegidas de manera tal que cada una sirva para representar de forma rala un sector de la imagen que se está analizando, esta es la idea central aquí (concepto de *diversidad morfológica*). Es deseable que las transformaciones combinadas satisfagan, que cuando una transformación representa de forma rala una parte de la imagen, brinde representaciones no ralas en otro tipo de contenido (relacionado con el concepto de *incoherencia mutua*).

### *Estimación de PMLE*

Ignorando el mecanismo de pérdida de datos, se considera un vector observado $\mathbf{y}$ $n$-dimensional completo. Adoptando el marco de modelos lineales generalizados (GML), se asume que $\mathbf{y}$ tiene una densidad en la familia de las exponenciales clásicas, y toma la forma:

$$f_{\mathbf{Y}}(\mathbf{y}, \eta; \sigma) = h(\mathbf{y}, \sigma) e^{\left[(\eta^T \mathbf{y} - A(\eta))/g(\sigma)\right]} \tag{2}$$

para algunas funciones específicas $A(.)$, $g(.)$ y $h(.)$, $\eta$ es el parámetro canónico y $\sigma$ es el parámetro de escala.

A partir de resultados sobre familia exponenciales, la media de $\mathbf{Y}$ está relacionada con $\eta$ por $\mathsf{E}[\mathbf{Y}_i] = a'(\eta_i)$, porque típicamente $A(\eta) = \sum_{i=1}^n a(\eta_i)$. Existe también una transformación uno-a-uno llamada función de enlace (*link*), la que relaciona $\mathsf{E}[\mathbf{Y}_i]$ con $\Phi\alpha$. El caso más importante se corresponde con la función de enlace canónica $(a')^{-1}(\cdot)$, o $\eta = \Phi\alpha$, lo que implica $(\Phi\alpha)_i = (a')^{-1}(\mathsf{E}[\mathbf{Y}_i])$. En este caso $\alpha$ puede verse como parámetros de una sub-familia canónica de la familia exponencial original con suficiencia estadística $\Phi^T \mathbf{Y}$.

Se busca una estimación regularizada de $\alpha$ desde $\mathbf{y}$ usando el estimador de máxima verosimilitud penalizado (PMLE):

$$\hat{\alpha} = \arg\max_{\alpha} \left[ ll(\mathbf{y}|\alpha) + \log p_\alpha(\alpha) \right] \tag{3}$$

también llamado estimador de probabilidad a posteriori máxima (MAP). El primer término de la ecuación da cuenta de la fidelidad en la observación y $p_\alpha(\cdot)$ introduce la distribución a priori de los coeficientes $\alpha$.

Aquí se restringe el análisis al caso de ruido aditivo blanco Gaussiano (AWGN): $P(\mathbf{Y}|\alpha) \sim \mathcal{N}(\mathbf{\Phi}\alpha, \sigma^2 \mathbf{I})$. De esta forma, el problema de estimación **MAP/PMLE** puede expresarse en términos de los coeficientes de descomposición $\alpha$, lo que sería, para AWGN con varianza $\sigma^2$:

$$\hat{\alpha} = \mathrm{i}\min_{\alpha} \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{\Phi}\alpha\|_2^2 + \lambda \Psi(\alpha) \tag{4}$$

donde $\lambda > 0$ es el parámetro de regularización, $\Psi : \mathbb{R}^L \to \mathbb{R}^+$ es una función de penalización (PF) que promueve la reconstrucción con baja complejidad tomando ventaja de la dispersión (sparsity).

### *Algoritmo EM*

Se vuelve sobre al caso de imágenes con datos perdidos. Se considera $\mathbf{y} = (\mathbf{y}_o, \mathbf{y}_{miss})$, donde $\mathbf{y}_{miss} = \{\mathbf{y}_i\}_{i \in I_m}$ refiere a los datos perdidos, y $\mathbf{y}_o = \{\mathbf{y}_i\}_{i \in I_o}$. Las observaciones incompletas no contienen toda la información para aplicar los métodos estándar para resolver (6) y obtener el PMLE de $\theta = (\alpha^T, \sigma^2)^T \in \Theta \subset \mathbb{R}^L \times \mathbb{R}^{+*}$. Sin embargo, el algoritmo *EM* puede aplicarse para reconstruir los datos perdidos de forma iterativa y luego resolver (6) para la nueva estimación.

## II. Análisis crítico ”general”

### A. *La estructura general del artículo: resumen, contenidos, conclusiones.*

El resumen presenta de manera correcta el trabajo.

Los contenidos del artículo son adecuados para seguir la dinámica del método. Está bastante detallado, se presentan las pruebas de convergencia del algoritmo para cada uno de los tipos de diccionario planteados y para complementarlo se incluye un apéndice con las pruebas de las proposiciones planteadas en los desarrollos de las pruebas de convergencia.

Las conclusiones son adecuadas e interesantes y dan pie a investigaciones futuras, aunque los ejemplos analizados si bien son interesantes parecen ser muy pocos y particulares.

### B. *Cómo los autores realizaron el análisis del estado del arte: estructura general, secuencia, “completitud” y crítica al trabajo original.*

Los autores realizan una reseña del avance del interés en aplicaciones de desoclusión de imágenes que luego daría lugar al inpainting de imágenes. Se comentan las diversas técnicas que se han utilizado y se destacan resultados satisfactorios y limitaciones. Luego, plantean trabajos relacionados, cómo este trabajo se inserta allí y cuales son los avances que éste presenta. Se cita el surgimiento de una nueva línea de interés en el inpainting cuando se unifican las aplicaciones en la industria del cine, películas y restauración de arte.

Algunos autores proponen usar métodos variacionales basados en ecuaciones diferenciales parciales no lineales (PDE) para inpainting de imágenes o partes de imágenes no texturadas (o parte natural de la imagen). Otros proponen resolver el problema de las texturas completando las partes faltantes con información de un contorno, sintetizando una imagen completa y coherente. Para ésto, las partes visibles sirven de entrenamiento para inferir luego las partes desconocidas, se ha propuesto usar Markov aquí. Hay trabajos que combinan las técnicas, primero reconstruyen las partes naturales y luego las texturas, otros que reconstruyen las partes natural y mientras van llenando las texturas progresivamente. Recientemente se introdujo un algoritmo capaz de reconstruir texturas y partes naturales simultáneamente, extendiendo la técnica de análisis de componentes morfológicos (MCA) y diseñado para separar la imagen según los distintos componentes semánticos que presenta.

Existen varios trabajos que plantean resolver el problema de inpainting (imágenes con texturas y partes naturales) a partir de métodos basados en representaciones ralas. Se hace referencia a trabajos que tratan el problema de inpainting, donde se plantea como una tarea de limpieza de ruido (*denoising*) en dos etapas: Se estima la imagen asumiendo un soporte conocido y se estima el soporte progresivamente. Todo el análisis se desarrolla en base a un soporte conocido, lo cual no es cierto en datos reales. Aquí se establece un problema lineal inverso mal condicionado para encarar el problema del inpainting, lo que se resuelve en un marco de bases bayesianas, y para el cual el mecanismo *EM* surge como un algoritmo iterativo debido a que existen datos perdidos. El marco de trabajo es más general y permite estimar los parámetros del ruido, además se puede extender a otros tipos de ruido y aplicaciones (interpolación, zooming). También se ha realizado un análisis detallado y teóricamente de las bases de la convergencia para caracterizar el tipo de soluciones. Algunos trabajos citados tienen una estructura similar y utilizan como diccionario una unión de bases ortogonales o un simple marco ajustado como ser *framelet*. Por ésto, otra capacidad que hace interesante al método propuesto es que permite utilizar cualquier diccionario, incluyendo la unión de transformaciones que pueden ser marcos o marcos ajustados como ser la transformada *curvelet*. Así, permite usar diccionarios sobre-completos con los cuales se puede representar de forma rala (basada en el contenido morfológico) una gran variedad de imágenes.

Si bien el método da pie a la implementación para diversos tipos de ruidos, sólo se plantea para ruido aditivo blanco gaussiano.

### C. *La bibliografía presentada: principales referencias, omisiones importantes.*

La omisión detectada es la referencia al método PSNR de evaluación de la calidad de la imagen resultante del método.

Las citas más interesantes:

- Sobre las propiedades de la convergencia del algoritmo *EM*: Wu, C. (1983) *On the convergence properties of the EM algorithm*. Ann. of Stat.,**11**, 95-103.
- Sobre la máxima verosimilitud de datos incompletos mediante el algoritmo *EM*: Dempster, A., Laird, N., and Rubin, D. (1977) *Maximum Likelihood from Incomplete Data via the EM Algorithm*. J. Roy. Stat. Soc. B, **39**, 1-38.
- Sobre diversidad morfológica: Starck, J.-L., Elad, M., Donoho, D. (2005) *Image Decomposition via the Combination of Sparse Representations and a Variational Approach*. IEEE Trans. Im. Proc., **14**, 1570-1582.
- Sobre análisis de componentes morfológicos: Starck, J.-L., Elad, M., Donoho, D. (2004) *Redundant Multiscale Transforms and Their Application for Morphological Component Separation*. Adv. Imaging Electron Phys. **v132**, 287-348.
- Sobre búsquedas de bases con limpieza de ruido: Chen, S. S.(1995) *Basis Pursuit*. PhD thesis, Department of Statistics, Stanford University.
- Sobre el trabajo analizado: Fadili, M.J. Starck, J.-L, Murtagh, F. (2007) *Inpainting and Zooming using Sparse Representations*. The Computer Journal, Julio (2007).

## III. DESARROLLO

### A. MLE penalizado con datos perdidos

Recordemos que se trabaja con una imagen completa $\mathbf{x}$ definida en un dominio finito $\Omega$ (el plano), y su versión degradada $\mathbf{y}$ (aunque no completamente observada). La imagen observada (incompleta) $\mathbf{y}_{obs}$ es el resultado de aplicar el operador de pérdida $\mathcal{M}$ sobre $\mathbf{y}$: $\mathcal{M} : \mathbf{y} \mapsto \mathbf{y}_{obs} = \mathcal{M}[\mathbf{y}] = \mathcal{M}[\mathbf{x} \odot \varepsilon]$, donde $\odot$ indica una composición y $\varepsilon$ es ruido. Con inpainting se recupera $\mathbf{x}$ a partir de $\mathbf{y}_{obs}$, y esto es un problema inverso mal condicionado.

### B. Estimación de PMLE

En principio se deja de lado el mecanismo de pérdida de datos y se considera un vector observado $\mathbf{y}$ $n$-dimensional completo (con un simple reordenamiento): $y = \Phi\alpha + \varepsilon$.

Se busca una estimación regularizada de $\alpha$ a partir de $\mathbf{y}$ usando un estimador de máxima verosimilitud penalizado (PMLE):

$$\hat{\alpha} = \arg\max_{\alpha} [ll(\mathbf{y}|\alpha) + \log p_\alpha(\alpha)] \tag{5}$$

esto también es llamado estimador de máxima a posteriori (MAP). El primer término de (5) es el aporte del logaritmo de la verosimilitud que promulga la fidelidad en la observación. $p_\alpha(\cdot)$ soporta la distribución a priori impuesta sobre los coeficientes $\alpha$, por ejemplo: Gaussiana, Laplaciana, Gaussiana generalizada, etc.

Aquí se restringe el análisis al caso de ruido Gaussiano blanco aditivo (AWGN) con media $\mu = 0$ y desviación $\sigma$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Podemos expresar a $P(\mathbf{y}|\alpha)$ como la distribución normal:

$$\begin{aligned} P(\mathbf{y}|\alpha) &\sim \mathcal{N}(\mathbf{\Phi}\alpha, \sigma^2\mathbf{I}) \\ P(\mathbf{y}|\alpha) &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} exp\left[-\frac{\|\mathbf{y}-\Phi\alpha\|_2^2}{2\sigma^2}\right] \end{aligned}$$

Así, el problema de estimación **MAP/PMLE** puede expresarse en términos de los coeficientes de descomposición $\alpha$, lo que sería para AWGN con varianza $\sigma^2$, reemplazando en (5):

$$\hat{\alpha} = \arg\max_{\alpha}\left[ll\left(\frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{\left[-\frac{\|\mathbf{y}-\Phi\alpha\|_2^2}{2\sigma^2}\right]}\right) + \lambda\Psi(\alpha)\right]$$

$$\hat{\alpha} = \arg\max_{\alpha}\left[-log\left(\sigma\sqrt{2\pi}\right)^n - \frac{\|\mathbf{y}-\Phi\alpha\|_2^2}{2\sigma^2} + \lambda\Psi(\alpha)\right]$$

$$\hat{\alpha} = \arg\max_{\alpha}\left[-\frac{n}{2}\,log(2\pi\sigma^2) - \frac{\|\mathbf{y}-\Phi\alpha\|_2^2}{2\sigma^2} + \lambda\Psi(\alpha)\right]$$

Esto es equivalente a minimizar:

$$\hat{\alpha} = \mathrm{i}\min_{\alpha}\left[\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{\Phi}\alpha\|_2^2 + \lambda\Psi(\alpha)\right] \tag{6}$$

donde $\lambda > 0$ es el parámetro de regularización, $\Psi : \mathbb{R}^L \to \mathbb{R}^+$ es una función de penalización (PF) que promueve la reconstrucción. Adicionalmente se asume que la distribución a priori asociada a $\Psi(\alpha)$ es separable (que los coeficientes $\alpha$ son independientes e idénticamente distribuidos (**iid**)). Entonces: $\Psi(\alpha) = \sum_{l=1}^{L} \psi(\alpha_l)$

Usualmente la $\psi$ de regularización se plantea como una función en $\mathbb{R}^+$ no negativa, continua, simétrica-par y no decreciente, pero no necesariamente convexa en $\mathbb{R}^+$ y puede ser irregular en el punto cero para producir soluciones de tipo ralas. Una elección típica de $\psi$ es la penalización de norma-$\ell 1$ la que permite una regla de umbralamiento suave bien conocida. El problema planteado en (6) con penalización de norma-$\ell 1$ es una variante de la búsqueda de base para ruido conocido (noise-aware basis pursuit).

### C. Definiciones del algoritmo EM

El algoritmo básicamente permite, de forma iterativa, estimar datos cuando se tienen observaciones de datos incompletos. Ésto implica la existencia de dos espacios de muestras $\mathcal{X}$ y $\mathcal{Y}$ y diversos mapeos $\mathcal{X} \to \mathcal{Y}$. Los datos observados $y$ son una realización de $\mathcal{Y}$. El $x$ en $\mathcal{X}$ no es observado directamente sino a través de $y$. Se asume un mapeo $x \to y(x)$ de $\mathcal{X}$ a $\mathcal{Y}$, entonces $x$ sólo se conoce $\mathcal{X}(y)$ (el subconjunto de $\mathcal{X}$ determinado por la ecuación $y = y(x)$, donde $y$ son los datos observados). Aquí $x$ son los datos completos.

Se postula una familia de densidades muestrales $f(x|\Phi)$ dependiendo de los parámetros $\Phi$ y se deriva su familia de densidades muestrales correspondiente $g(y|\Phi)$. La especificación de los datos completos $f(\ldots|\ldots)$ es relacionada a la especificación de datos incompletos $g(\ldots|\ldots)$ por

$$g(y|\Phi) = \int_{\mathcal{X}(y)} f(x|\Phi)dx \tag{7}$$

Con el algoritmo *EM* se quiere encontrar el valor de $\Phi$ que maximize $g(y|\Phi)$ dado los datos observados $y$, haciendo un esencial uso de la familia asociada $f(x|\Phi)$. Existen muchas especificaciones de datos completos generados a partir de una especificación de datos incompletos $g(y)|\Phi$ dada. Ahora suponga que $f(x|\Phi)$ tiene la forma regular de la familia exponencial:

$$f(x|\Phi) = b(x)exp(\phi t(x)^T)/a(\Phi) \tag{8}$$

donde $\Phi$ denota un vector de parámetros de $1 \times r$ y $t(x)$ es un vector de $1 \times r$ con la suficiencia estadística de los datos completos. El término regular significa que $\Phi$ está restringida a un conjunto convexo $\Omega$ $r$-dimensional tal que define una densidad para todo $\Phi$ en $\Omega$.

Suponemos que $\Phi^{(p)}$ es el valor de $\Phi$ después de $p$ ciclos del algoritmo. El próximo ciclo se describe en dos pasos:

**Paso *E*:** Estimar la suficiencia estadística de los datos completos t(x) hallando:

$$t^{(p)} = E(t(x)|y, \Phi^{(p)}) \tag{9}$$

**Paso *M*:** Calcular $\Phi^{(p+1)}$ como solución de las ecuaciones:

$$E(t(x), \Phi) = t^{(p)} \tag{10}$$

Suponiendo que $t^{(p)}$ representa la suficiencia estadística computada de un $x$ observado, entonces (10) define el estimador de la máxima verosimilitud de $\Phi$. Para un $x$ dado, maximizar el $log\, f(x|\Phi) = -log\, a(\Phi) + log\, b(x) + \phi t(x)^T$ es equivalente a maximizar: $-log\, a(\Phi) + \phi t(x)^T$ lo que depende de $x$ sólo a través de $t(x)$. Si (10) puede resolverse para $\Phi$ en $\Omega$, entonces la solución es única debido a la propiedad de convexidad del logaritmo de la verosimilitud para familias exponenciales regulares.

A continuación se plantea una pequeña explicación de como la repetida aplicación de los pasos lleva a obtener un valor $\Phi^*$ de $\Phi$ que maximiza:

$$L(\Phi) = log\, g(y|\Phi) \tag{11}$$

donde $g(y|\Phi)$ es la especificación de datos incompletos (7).

La densidad condicional de $x$ dados $y$ y $\Phi$ es: $k(x|y,\Phi) = f(x|\Phi)/g(y|\Phi)$, entonces:

$$L(\Phi) = log\, f(x|\Phi) - log\, k(x|y,\Phi) \tag{12}$$

Para familias exponenciales tenemos que: $k(x|y,\Phi) = b(x)exp(\Phi t(x)^T)/a(\Phi|y)$, donde:

$$a(\Phi|y) = \int_{\mathcal{X}(y)} b(x)exp(\Phi t(x)^T)dx \tag{13}$$

Luego como $f(x|\Phi)$ y $k(x|y,\Phi)$ representan familias exponenciales con los mismos parámetros naturales $\Phi$ y la misma suficiencia estadística $t(x)$, pero definidos sobre diferentes espacios de muestras $\mathcal{X}$ y $\mathcal{X}(y)$. Se reescribe (12) como:

$$L(\Phi) = -log\, a(\Phi) + log\, a(\Phi|y) \tag{14}$$

donde el paralelo a (13) es:

$$a(\Phi) = \int_{\mathcal{X}} b(x)exp(\Phi t(x)^T)dx \tag{15}$$

Diferenciando (15) y (13) y notando $t(x)$ como $t$, se tiene que

$$\mathbf{D}log\, a(\Phi) = (\partial/\partial\Phi)log\, a(\Phi) = E(t|\Phi) \tag{16}$$

de igual manera

$$\mathbf{D}log\, a(\Phi|y) = E(t|y,\Phi) \tag{17}$$

de esta forma

$$\mathbf{D}L(\Phi) = -E(t|\Phi) + E(t|y,\Phi) \tag{18}$$

Entonces la derivada el logaritmo de la verosimilitud tiene una representación en base a la diferecia de una esperanza incondicional y una esperanza condicional de la suficiencia estadística. La clave para entender el método está en (18), para la cual el algoritmo converge a $\Phi^*$ (esto es en el límite $\Phi^{(p)} = \Phi^{(p+1)} = \Phi^*$), entonces combinando (9) y (10) nos conduce a $E(t|\Phi^*) = E(t|y,\Phi^*)$ o bien $\mathbf{D}L(\Phi) = 0$ en $\Phi = \Phi^*$.

### D. *Derivación de los pasos E y M*

Volviendo sobre el caso de imágenes con datos perdidos. Se considera $\mathbf{y} = (\mathbf{y}_o, \mathbf{y}_{miss})$, donde $\mathbf{y}_{miss} = \{\mathbf{y}_i\}_{i\in I_m}$ refiere a los datos perdidos, e $\mathbf{y}_o = \{\mathbf{y}_i\}_{i\in I_o}$. Las observaciones incompletas no contienen toda la información para aplicar los métodos estándar para resolver (6) y obtener el PMLE de $\theta = (\alpha^T, \sigma^2)^T \in \Theta \subset \mathbb{R}^L \times \mathbb{R}^{+*}$. Sin embargo, el algoritmo *EM* puede aplicarse para reconstruir los datos perdidos de forma iterativa y luego resolver (6) para la nueva estimación. Las estimaciones se refinan iterativamente hasta la convergencia.

*El paso E*

Aquí se computa la esperanza condicional del logaritmo de la verosimilitud penalizado de los datos completos, dado $\mathbf{y}_o$ y los parámetros actuales $\theta^{(t)} = (\alpha^{(t)^T}, \sigma^{2^{(t)}})^T$:

$$\begin{aligned} Q\left(\theta|\theta^{(t)}\right) &= \mathsf{E}\left[ll(\mathbf{y}|\theta) - \lambda\Psi(\alpha)|\mathbf{y}_o, \theta^{(t)}\right] \\ &= \mathsf{E}\left[ll(\mathbf{y}|\theta)|\mathbf{y}_o, \theta^{(t)}\right] - \lambda\Psi(\alpha) \end{aligned} \tag{19}$$

Para familias exponenciales regulares, el paso *E* se reduce a encontrar los valores esperados de la suficiencia estadística de los datos completos de $\mathbf{y}$ dado por la información observada $\mathbf{y}_o$ y la estimación de $\alpha^{(t)}$ y $\sigma^{2^{(t)}}$. Luego, y como el ruido es blanco Gaussiano y de media cero, el paso *E* se reduce a calcular la esperanza condicional de los valores y la esperanza condicional de los valores al cuadrado de la información perdida, esto sería:

$$\begin{aligned} y_i^{(t)} &= \mathsf{E}\left[y_i|\mathbf{\Phi}, \mathbf{y}_o, \alpha^{(t)}, \sigma^{2^{(t)}}\right] = \\ &= \begin{cases} y_i & \text{para datos observados}, i \in I_o \\ (\Phi\alpha^{(t)})_i & \text{para datos perdidos}, i \in I_m \end{cases} \end{aligned} \tag{20}$$

$$\mathsf{E}\left[y_i^2|\mathbf{\Phi}, \mathbf{y}_o, \alpha^{(t)}, \sigma^{2^{(t)}}\right] = \begin{cases} y_i^2 & i \in I_o \\ (\Phi\alpha^{(t)})_i^2 + \sigma^{2^{(t)}} & i \in I_m \end{cases} \tag{21}$$

*El paso M*

En este paso se maximiza la función de penalización sustituta con las observaciones perdidas por sus estimaciones en el paso *E* en la iteración *t*, esto es:

$$\theta^{(t+1)} = \mathrm{i}\min_{\theta\in\Theta} -Q(\theta|\theta^{(t)}) \tag{22}$$

Entonces el paso *M* actualiza $\sigma^{2^{(t+1)}}$ de acuerdo a:

$$\sigma^{2^{(t+1)}} = \frac{1}{n}\left[\sum_{i\in I_o}(y_i - (\Phi\alpha^{(t)})_i)^2 + (n - n_o)\sigma^{2^{(t)}}\right] \tag{23}$$

donde $n_o = \mathsf{tr}\, M = \#\, I_o$ es el número de píxeles observados. La convergencia (cuando se consideren a todos los píxeles observados) vendría dada por:

$$\sigma^{2^{(t+1)}} = \sigma^{2^{(t)}} = \hat{\sigma}^2 = \frac{1}{n_o}\sum_{i\in I_o}(y_i - (\Phi\hat{\alpha})_i)^2$$

que viene a ser la estimación de máxima verosimilitud de la varianza del ruido dentro de la máscara.

Se actualiza $\alpha^{(t+1)}$ que no sólo depende de la estructura del diccionario $\Phi$, sino también de las propiedades de la PF de $\psi$ (por ejemplo si es suave o no, si es convexa o no convexa). Estos dos parámetros tienen un claro impacto en la manera en que converge el algoritmo de inpainting. En todos los casos la estructura general del algoritmo de inpainting es:

**Algoritmo 1**

**Se requiere:** La imagen observada $\mathbf{y}_o = \mathcal{M}\mathbf{y}$ y una márcara $\mathcal{M}$, $t = 0$, $\alpha^{(0)}$ inicial

1: **Repetir**
2: Paso **E**
3: Actualización de la imagen estimada:

$$\mathbf{y}^{(t)} = \mathbf{y}_{obs} + (\mathbf{I} - \mathcal{M})\mathbf{\Phi}\alpha^{(t)} = \mathbf{\Phi}\alpha^{(t)} + (\mathbf{y}_{obs} - \mathcal{M}\mathbf{\Phi}\alpha^{(t)})$$

4: Paso **M**
5: Resolver la ec. 6 con el $\mathbf{y}^{(t)}$ reconstruido sustituyendo por $\mathbf{y}$, y se obtiene $\alpha^{(t+1)}$.
6: Para actualizar $\sigma^{2^{(t+1)}}$, ver la ec. (22).
7: $t = t + 1$,
8: **Hasta que** se llegue a la convergencia.
9: Recontrucción de $\hat{x}$ con $\hat{\alpha}$.

Si se conoce $\sigma^2$, el paso 6 puede eliminarse del esquema.

## IV. ANÁLISIS CRÍTICO ”DE RESULTADOS”

### A. *Métodos experimentales (si corresponde)*

Se han realizado experimentos con el algoritmo *ECM* en diversas situaciones. Se han analizado problemas de inpainting, interpolación de datos perdidos y super-resolución (zooming). Basicamente, se elije un problema, se forma el diccionario, se fijan los parámetros requeridos por el algortimo y se corre.

Una dificultad que surge por ser requerimiento del algoritmo es la máscara ($\mathcal{M}$), en los ejemplos de las jaulas, éstos se han dibujado a mano. Sin embargo, el método en este tema es muy dependiente de la imagen que se analiza y no siempre se pueden dibujar las máscaras a mano, ésto así planteado le quita lo *automático* al método. En el caso real, se debería tener en cuenta la especificación exacta del instrumento de medición de la imagen para determinar que píxeles se pierden; sí el obstáculo es externo o la pérdida se manifiesta de forma aleatoria el problema requiere demasiada intervención del usuario. No se plantea en el trabajo alguna alternativa automática para la estimación de las máscaras.

Se han realizado evaluaciones subjetivas no formales y objetivas basadas en el valor pico de la relación señal-ruido (PSNR). Sin embargo, no se da lugar en el trabajo a un breve comentario de la medida PSNR, por lo tanto se define a continuación y se explica su forma de uso.

PSNR representa la relación entre la potencia máxima posible de una señal y la potencia del ruido que afecta la fidelidad de su representación. Como muchas señales tienen un rango dinámico muy amplio, usualmente se expresa al PSNR en términos de la escala logarítmica (decibel). Es usada para medir la calidad de la reconstrucción (en este trabajo la calidad de la restauración, interpolación y super-resolución) de una señal (imagen), etc. Es simple definirlo a través del error cuadrático medio (MSE) el cual, para dos imágenes monocromáticas ($I$, $K$) de $m \times n$ donde una es la versión “ruidosa” de la otra, se define como:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} ||I(i,j) - K(i,j)||^2$$

Entonces el PSNR se define como:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

Aquí $MAX_I$ es el valor máximo que puede tomar un pixel en la imagen. Por ejemplo, si se codifica con $8$ *bits* por muestra, este valor es $255$. El valor real no es significativo, pero la comparación entre valores de diversas imágenes da una medida de calidad.

Una técnica importante para exhibir errores que podría incorporarse al análisis es la de construir una imagen del error que muestre los errores del pixel por pixel (crear una imagen tomando la diferencia entre los pixeles reconstruidos y originales). Ésto permitiría evaluar que morfologías dan más errores y así evaluar la inclusión de alguna otra transformación a los diccionarios utilizados.

**Análisis de imágenes en color**: para estos casos se ha elegido una representación de la imagen en un espacio *RGB*, a cada banda se le aplica el proceso de manera independiente y luego son combinadas en la imagen final resultante.

### B. *Cómo se presentan y discuten las tablas y figuras.*

Los resultados se presentan con una breve discusión acerca de cada ejemplo y luego se exhiben las imágenes originales, las *ruidosas* y las restauradas; junto al cálculo de la métrica PSNR para las imágenes ruidosas y las restauradas. Aunque PSNR es una de las medidas más utilizadas para medir la calidad de la reconstrucción/restauración de una imagen, los autores podrían haber planteado algunas medidas perceptivas. Recordando que los valores de PSNR no se correlacionan perfectamente con la calidad visual percibida de la imagen debido al comportamiento no lineal del sistema visual humano.

Se presentan unos pocos ejemplos que si bien son muy buenos, no se discuten falencias del método, inclusive argumentan una falta de análisis exahustivo de mayor cantidad de resultados para realizar conclusiones más precisas en los resultados.

Sólo se compara a la técnica con otro método (el PDE) en ejemplos de inpainting y para el caso particular de remoción de jaulas en imágenes con animales.

No se han utilizado tablas para la comparación con otros métodos que se aplican en estos casos de ejemplo.

### C. *¿Hay trabajos futuros propuestos? ¿Es un trabajo preliminar o está “cerrado”?*

Si bien se han demostrado las propiedades de convergencia del método y se dan ejemplos de su aplicación en inpainting, interpolación y super-resolución, el trabajo da pie a continuar con varias líneas de investigación. A partir de este trabajo, se plantean otros a futuro: la investigación formal del parámetro de regularización $\alpha$ (con métodos de seguir la trayectoria/continuación homotópica) y la extensión de esta técnica a imágenes multivaluadas como ser los datos hiper-espectrales.

## V. FORMULARIO DE REVISIÓN DE ARTICULO

### A. *Evaluación Global*

- **Autores**: M.J. Fadili, J.-L. Starck y F. Murtagh
- **El título**: Inpainting and Zooming using Sparse Representations

Por favor, asigne puntaje en [0 ..10] para los aspectos detallados debajo.

1. Relevancia respecto a la temática central de la revista: 8
2. Originalidad: 7
3. Calidad técnica: 10
4. Claridad y organización del trabajo: 9
5. Título y presentación del resumen: 9
6. Correspondencia con el formato de estilo: 10

*Valoración global:* ***8***

*10: Aceptación Fuerte (Calidad excelente, ideas claras)*
*8: Aceptado (Calidad buena, resultados sólidos e interesantes)*
*6: Aceptación débil (Vota por la aceptación, pero no lo defenderá)*
*4: Débil Rechazo (no le gusta, pero no argumentará contra él)*
*2: Rechazo (argumentará para rechazar este trabajo)*
*0: Fuerte Rechazo (Impropio para esta revista)*

### B. *Comentarios Detallados*

Por favor, describa las contribuciones principales del trabajo, proporcionando detalles que ayudarán a los autores a mejorar la calidad de su artículo.

1. Contribuciones principales: se propone un marco de trabajo más general que los existentes y da pie para el análisis con otro tipo de ruido. Lo más interesante es que combina la restauración de imágenes que presentan partes naturales o geométricas y partes con texturas, además se evalua la utilización de cualquier tipo de transformaciones en los diccionarios.
2. Metodología y resultados: la metodología es adecuada, aunque es notorio que los resultados se han planteado con unos 3 o 4 ejemplos y que no se compara a todos con otros métodos. Sin embargo, en los resultados se aprecian buenos resultados en la métrica (PSNR) y mediante una inspección ocular se ven restauraciones muy significativas.
3. Desarrollo teórico y contenido matemático: el contenido matemático es adecuado en el trabajo y tiene un apéndice que complementa las demostraciones de las proposiciones planteadas en el análisis de la convergencia del método.
4. Formato y organización general: el formato es adecuado, sin embargo la organización general no me resulto agradable. Ésto es porque las imágenes resultantes de los experimentos se encuentran todas al final del trabajo luego del apéndice y de la bibliografía (varias páginas después de su análisis), se dificulta la lectura del análisis de los resultados y la visualización simultánea de las imágenes. Además se incluye una sección de *trabajos relacionados* donde se revisan aspectos de otros trabajos y las ventajas de este trabajo, sin embargo podrían plantearse en una sección de discusión los aspectos resultantes del actual desarrollo y llevar la información de trabajos relacionados a la subsección de estado del arte.
5. Estilo y lenguaje: el estilo es del tipo de publicación en revista y el lenguaje es accesible y de buena redacción.
6. Aspectos positivos: se han tomado varias técnicas existentes y se convinaron de forma adecuada para la realización de estas tareas. Se presenta un algoritmo sencillo, que pemite la incorporación de diccionarios de uniones de transformaciones. Además se presenta la convergencia teórica del método.
7. Aspectos negativos: se ha planteado sólo un método para medir la calidad de los resultados y no se han planteado medidas perceptivas. Los resultados son escasos y no se ha hecho un análisis significativo que permita plantear las mejoras en tablas comparativas con varios métodos. Sólo para una aplicación en particular se compara al método con otro.
8. Comentarios adicionales: