

RECONSTRUCCIÓN DE CAMPOS DE LUZ A PARTIR DE PROYECCIONES 2D BASADO EN MUESTREO COMPRESIVO

LIGHT FIELD RECONSTRUCTION FROM 2D PROJECTIONS BY COMPRESSIVE SENSING

Kareth León¹, Laura V. Galvis², Henry Arguello Fuentes³

¹ Escuela de Ingeniería de Sistemas, Universidad Industrial de Santander, Bucaramanga, Colombia.
E-mail: kareth.leon@correo.uis.edu.co

² Electrical and Computer Engineering Department, University of Delaware, Newark, DE, USA

³ Escuela de Ingeniería de Sistemas, Universidad Industrial de Santander, Bucaramanga, Colombia.

Recibido: 3 Diciembre

Aceptado: 15 Diciembre

*Correspondencia del autor. Escuela de Ingeniería de Sistemas, Universidad Industrial de Santander, Bucaramanga, Colombia.

E-mail: kareth.leon@correo.uis.edu.co

RESUMEN

La fotografía de campos de luz ha ganado popularidad recientemente debido a sus capacidades para el procesamiento de imágenes. La captura de campos de luz es una técnica que almacena información espacial y angular de una señal óptica proveniente de un objeto tridimensional en una representación de cuatro dimensiones (4D). La adquisición de campos luz en alta resolución sigue siendo un desafío, y aunque existen cámaras comerciales de campo de luz, estos dispositivos sacrifican resolución espacial para ofrecer diferentes vistas angulares de la misma escena. Un nuevo enfoque para obtener alta resolución en campos luz es el uso de la técnica de muestreo compresivo (CS) desarrollada recientemente. El elemento esencial en CS es el proceso de codificación, que puede ser realizado a través de máscaras litográficas. En este trabajo, se presenta un modelo de adquisición de campos de luz usando muestreo compresivo por patrones de codificación. Esta investigación busca reconstruir un campo de luz de alta resolución a partir de una sola proyección 2D. En el modelo, la información redundante de las diferentes vistas angulares del campo de luz es aprovechada por la modulación a través de los patrones de codificación, y luego se soluciona un problema mal condicionado a través de técnicas de codificación dispersa. Con el objetivo de utilizar las técnicas de codificación dispersa, se asume un diccionario donde las imágenes naturales son altamente compresibles. Inicialmente, se realiza el entrenamiento de un diccionario usando 5000 muestras de campos de luz de entrenamiento con el objetivo de encontrar la representación del campo de luz más dispersa y así, reconstruir el campo de luz 4D. Para la reconstrucción, se obtiene un campo de luz con 5×5 vistas angulares a partir de una sola proyección codificada 2D. La proyección codificada 2D es dividida en parches superpuestos, luego cada parche 2D es reconstruido y fusionado usando un filtro de mediana para obtener finalmente el campo de luz 4D. Los resultados indican el alto PSNR (pico de señal a ruido) de las imágenes 4D reconstruidas.

Palabras claves: Fotografía computacional, campo de luz, muestreo compresivo.

ABSTRACT

Light field photography has gained popularity recently due to its imaging capabilities. The light field is a technique that stores the spatial and angular information of an optical signal from a three-dimensional object in a four-dimensional (4D) representation. The acquisition of high-resolution light fields is still a challenge, although there exist commercial light field cameras, those devices trade spatial resolution by offering different views of the same scene. A novel approach to obtain high-resolution light fields is by using the recently developed technique of compressive sensing (CS). The essential element in CS is the codification process, which can be realized by lithographic masks. In this work, a compressive sensing light field camera model based on codification by mask patterns is presented. This research tries to recover a high-resolution light field from a single 2D projection. In the model, the redundant information in the different views of a light field is exploited by the modulation through mask patterns and then an ill-posed underdetermined problem is solved by sparse coding techniques. In order to use the sparse techniques, it is assumed a dictionary where the natural images are compressible enough. Simulations of a dictionary-learning are first performed using a complete dictionary with 5000 light field samples to find the sparsest representation of the light field in order to reconstruct the 4D light field. For the sparse reconstruction, a light field with a 5x5 views is obtained from a single coded 2D projection. The 2D projection is divided into overlapping patches, and then each 2D patch is reconstructed and merged with a median filter to finally obtain the 4D light field. The results indicate the high PSNR (Peak Signal to Noise Ratio) of the reconstructed 4D images.

Keywords: Computational photography, Light field, Compressive sensing

I. INTRODUCCION

Un campo de luz se define como el flujo de rayos que viajan en cada posición y cada dirección del espacio, esto conlleva a tener dimensiones espaciales y angulares en un arreglo de 4 dimensiones (4D) (1). Este concepto se fundamenta en la parametrización de la luz considerada sobre el espacio de rayos 5D o función plenoptic (2), se consideran tres dimensiones para cada posición espacial y dos dimensiones para cada dirección del flujo. La restricción de la función plenoptic permite la reducción de 5D a 4D, o campo de luz, haciendo pasar el rayo de luz por dos planos paralelos, eliminando así, una dimensión (2, 3).

En la literatura se encuentran métodos tradicionales para la captura y construcción de campos de luz. Una primera aproximación, usa un arreglo de lentes para captar una escena, generando una red de muestras angulares ordenadas similar a una fotografía integral (4). Emplear matrices de microlentes ubicados entre el lente principal y el sensor en el sistema hizo posible la captura del campo de luz a partir de una sola adquisición. De esta forma, cada microlente, permite capturar información angular. Estas aproximaciones demandan recursos computacionales considerables, debido al procesamiento necesario para la generación y almacenamiento de las imágenes captadas. Actualmente, las técnicas de captura de campos de luz están

avanzando hacia aperturas codificadas que modulan los rayos de luz (5). Estas permiten o bloquean el paso de luz proporcionando una captura reducida de la escena en la que se descartan valores redundantes que no aportan mayor información.

Este trabajo se basa en dichas aperturas codificadas. Para la captura y procesamiento óptimo de los campos de luz se tienen en cuenta tres componentes claves: los átomos de campo de luz, que son los bloques fundamentales de construcción de campos de luz natural; un diseño óptico optimizado, para la captura de proyecciones de campo de luz en 2 dimensiones (2D); y un método de dispersión robusto, para la reconstrucción de campos de luz 4D (5). Se plantea simular la obtención de proyecciones de campos de luz en 2D y a partir de éstas, reconstruir el respectivo campo de luz en 4D y evaluar la calidad de su reconstrucción, bajo la premisa de que es posible su representación en algún diccionario.

II. Reconstrucción de Campos de Luz

Una imagen capturada en un sensor o cámara es la proyección de un campo de luz que contiene tanto información espacial como angular. La inclusión de una máscara de codificación a cierta distancia d del sensor, modula ópticamente la proyección del campo de luz. En notación discreta, esta proyección del campo

de luz puede expresarse como

$$\mathbf{i} = \Phi \mathbf{l}, \quad \Phi = [\Phi_1 \ \Phi_2 \dots \Phi_{v_2}], \quad (1)$$

donde $\mathbf{i} \in \mathbb{R}^m$ y $\mathbf{l} \in \mathbb{R}^{n \times d}$ representan la imagen capturada y el campo de luz respectivamente en forma vectorizada, donde $m \ll n$. El campo de luz está compuesto por un conjunto de vistas angulares de $v \times v$ que se encuentran apiladas en \mathbf{l} . Las submatrices $\Phi_j \in \mathbb{R}^{m \times m}$ son matrices dispersas que contienen en su diagonal códigos aleatorios de 1s y 0s, donde los valores en 0 bloquean la luz.

El problema inverso de reconstruir un campo de luz de sus respectivas proyecciones codificadas requiere invertir el sistema lineal de ecuaciones en (1). Debido a que el problema es un problema subdeterminado, pues se cuenta con más incógnitas (rayos de luz) que ecuaciones o medidas conocidas (píxeles del sensor), se aplican técnicas de codificación dispersa. Para hacerlo, se asume que los campos de luz son dispersos en alguna base o diccionario $\mathbf{D} \in \mathbb{R}^{n \times d}$, tal que

$$\mathbf{i} = \Phi \mathbf{l} = \Phi \mathbf{D} \boldsymbol{\alpha}, \quad (2)$$

donde $\boldsymbol{\alpha} \in \mathbb{R}^d$ contiene la mayoría de sus coeficientes en 0.

Se busca, entonces, reconstruir el conjunto de coeficientes dispersos $\boldsymbol{\alpha}$ que formen una representación exacta del campo de luz en algún diccionario \mathbf{D} . La solución a (2) puede plantearse como la respuesta al problema,

$$\begin{aligned} & \text{minimizar}_{\{\boldsymbol{\alpha}\}} \|\boldsymbol{\alpha}\|_1 \\ & \text{sujeto a } \|\mathbf{i} - \Phi \mathbf{D} \boldsymbol{\alpha}\|_2 \leq \epsilon, \end{aligned} \quad (3)$$

que es conocido en la literatura entre los métodos de relajación convexa como basis pursuit denoise (BPDN), en el cual el parámetro ϵ representa el nivel de ruido capturado por el sensor del sistema. Comúnmente, este problema se formula usando la función Lagrangiana

$$\text{minimizar}_{\{\boldsymbol{\alpha}\}} \|\mathbf{i} - \Phi \mathbf{D} \boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_1. \quad (4)$$

Este tipo de problemas tienen parte en la teoría de compresión dispersa (Compressive Sensing en Inglés), y a partir de algoritmos robustos es posible encontrar una aproximación de la solución (5).

El entrenamiento del diccionario $\mathbf{D} \in \mathbb{R}^{n \times d}$ que representa la señal dispersa se plantea como la solución al problema

$$\begin{aligned} & \text{minimizar}_{\{\mathbf{D}, \mathbf{A}\}} \|\mathbf{L} - \mathbf{D} \mathbf{A}\|_F \\ & \text{sujeto a } \mathbf{G}; \|\mathbf{A}_i\|_0 \leq k, \end{aligned} \quad (5)$$

donde $\mathbf{L} \in \mathbb{R}^{n \times d}$ es el conjunto de entrenamiento que contiene q señales o parches que representan la señal y $\mathbf{A} \in \mathbb{R}^{d \times q}$ es un conjunto de k coeficientes dispersos.

El planteamiento del problema permite que este se pueda resolver por diferentes métodos, encontrar el que ofrezca una buena representación y alta calidad de reconstrucción es uno de los objetivos de éste trabajo, el cual se buscará basado en el PSNR promedio de todos los parches que compone cada campo de luz reconstruido. Como parámetro adicional, se tiene en cuenta el diccionario a utilizar. El entrenamiento del diccionario a usar en la representación de los campos de luz, afecta la calidad de reconstrucción y es importante tener presente que las imágenes empleadas para realizar dicho entrenamiento hacen que el diccionario adquiera nuevas características, es decir, que un diccionario depende directamente de las imágenes con las que se entrena.

III. Simulaciones y Resultados

Para las simulaciones, se utilizaron los algoritmos implementados en el paquete SPGL1 tales como BPDN (Basis Pursuit Denoise) y LASSO [6, 7]. Se cuenta con una base de datos de 5 imágenes de 840 "x" 593 píxeles de resolución espacial y cada una de ellas, cuenta con 5 "x" 5 vistas angulares.

A. Proyecciones codificadas 2D a 4D

Las proyecciones codificadas 2D se recrean a partir de parches o pequeñas ventanas de 8 "x" 8 píxeles, las cuales permiten una reconstrucción 4D del campo de luz; cada parche es reconstruido por cada píxel de la imagen del sensor. Para cada imagen, se calcula la calidad de reconstrucción PSNR (Peak signal-to-noise ratio) en decibeles (dB).

B. Entrenamiento de diccionarios

Se realizó el entrenamiento de tres diccionarios diferentes. La parte fundamental de cada diccionario es el conjunto de imágenes sintéticas escogidas para el entrenamiento.

Para cada diccionario, se emplearon cuatro grupos di-

ferentes de imágenes. Luego, mediante estos diccionarios, se realizan las reconstrucciones de tres campos de luz.

C. Reconstrucción a partir de diccionarios entrenados
La Figura 1 ilustra la variación de la calidad en una de las imágenes reconstruida a partir de cada uno de los diccionarios; los valores de PSNR muestran numéricamente el resultado obtenido al utilizar el diccionario 1, 2 y 3 respectivamente. El algoritmo de reconstrucción robusto empleado en la figura 1 es el llamado LASSO.

En las reconstrucciones, se calcula el PSNR de cada ventana o parche; el promedio de los valores del PSNR para cada campo de luz presentado en la figura 1 se muestra en la tabla 1. El diccionario 1 proporciona una calidad de reconstrucción alta respecto a la

obtenida en las reconstrucciones con los diccionarios 2 y 3.

D. Comparación de métodos de reconstrucción

Con el fin de analizar el comportamiento de los diferentes diccionarios y algoritmos de reconstrucción, se realizó un análisis variando estos dos parámetros.

En la tabla 2, se observan los resultados de hacer ésta variación en dos conjuntos de imágenes denominados Imagen 1 e Imagen 2. La Fig. 2 muestra el comportamiento de los resultados de reconstrucción de la imagen 1, utilizando los dos algoritmos, BPDN Y LASSO y, teniendo en cuenta los 3 diccionarios entrenados.

Los valores promedio de PSNR para cada campo de luz reconstruido y los respectivos métodos usados para dicha simulación son mostrados en la tabla 2.

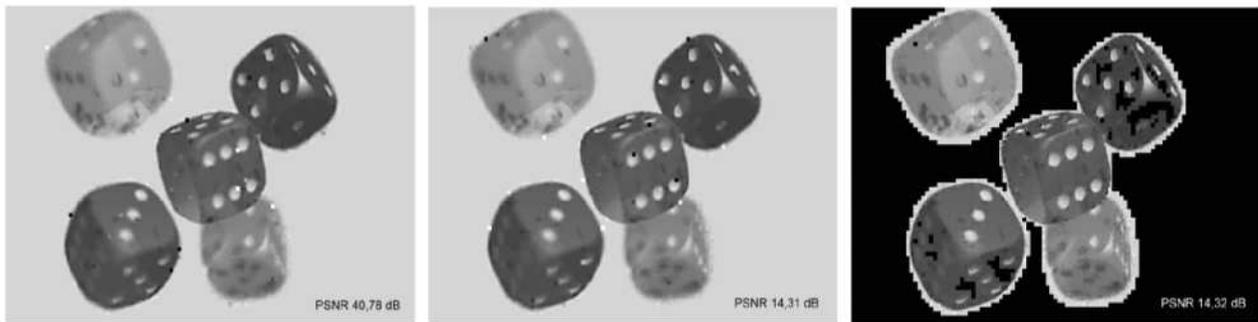


Figura. 1 Evaluación de reconstrucción con diferentes diccionarios; en la imagen de la izquierda se hizo uso del diccionario 1, la imagen del centro fue reconstruida con el diccionario 2 y para la imagen de la derecha se usó el diccionario 3. El algoritmo de reconstrucción empleado fue LASSO.

TABLA I
COMPARACIÓN DEL RENDIMIENTO UTILIZANDO TRES DICCIONARIOS

DICCIONARIO	PSNR prom (Método LASSO)
1	40,781
2	14,318
3	14.325

TABLA II
COMPARACIÓN DEL RENDIMIENTO EN LA RECONSTRUCCIÓN DE DOS IMÁGENES UTILIZANDO TRES DICCIONARIOS Y DOS ALGORITMOS DE RECONSTRUCCIÓN

Imagen a reconstruir	Diccionario	PSNR promedio	
		LASSO	BPDN
Imagen 1	1	35,201	35,711
	2	35,329	35,821
	3	35,237	35,598
Imagen 2	1	40,781	47,294
	2	14,318	47,187
	3	14,325	47,306

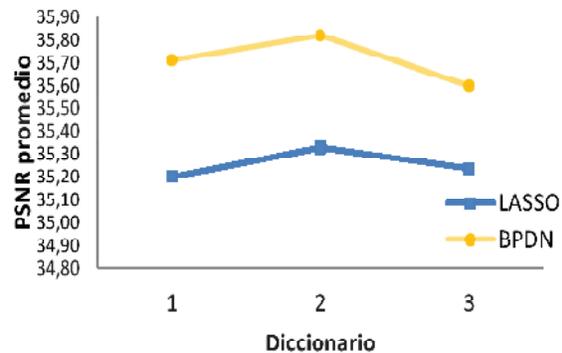


Figura. 2 Comparación gráfica de los resultados de la reconstrucción de la imagen 1. En promedio, se obtiene que la implementación BPDN contribuye en una mejora en la calidad de Reconstrucción.

El algoritmo BPDN proporciona valores altos en los dos campos de luz. Se puede observar en los valores de PSNR de la Imagen 2, una disminución en la calidad considerable, es decir, al utilizar el algoritmo

LASSO, los valores de PSNR obtenidos con los diccionarios 2 y 3 son menores respecto al obtenido con el diccionario 1, contrario a esto, los valores para el algoritmo BPDN no varían significativamente al cambiar el diccionario.

E. Reconstrucción a partir de parches de 9 "x" 9

Adicional a las simulaciones realizadas con parches de 8 "x" 8, se realizaron simulaciones tomando parches de 9 "x" 9 para realizar el entrenamiento de los diccionarios. Con esta simulación se encontraron resultados variables. Para la reconstrucción de la Imagen 1, se obtiene una mejora de la calidad del PSNR promedio, caso contrario sucede para la Imagen 2, en donde se obtiene menor calidad.

En la figura 3, se presenta la reconstrucción con dos tamaños de parches, para el campo de luz de la Imagen 1. Es de notar visualmente que la calidad mejora cuando se toma un parche de mayor tamaño.

Por otro lado, se calculó el PSNR de la reconstrucción del campo de luz de la Imagen 2 y se encontraron resultados opuestos, la calidad obtenida, utilizando parches de 9 "x" 9 fue menor en comparación con la obtenida a través de parches de 8 "x" 8. La tabla 3, muestra los resultados de esta última simulación, al utilizar los 2 algoritmos de reconstrucción BPDN y LASSO. Estas simulaciones fueron realizadas utilizando el diccionario 1, debido a que este presentó el mejor rendimiento en las simulaciones anteriores.

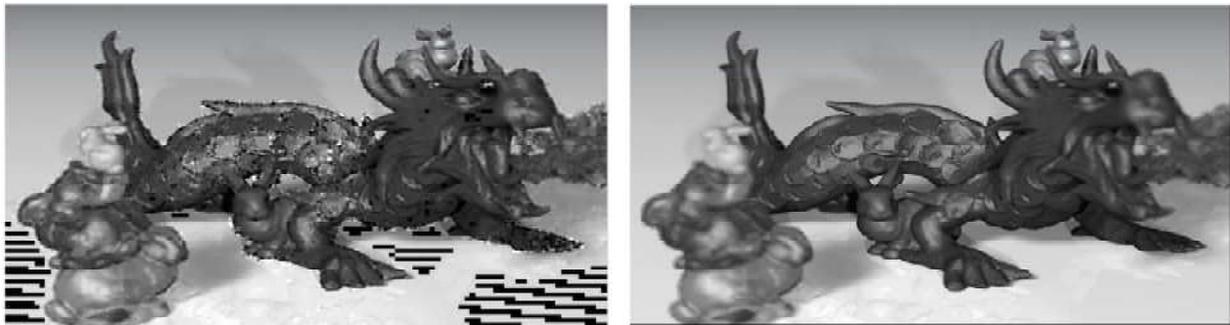


Figura. 3 Comparación de la reconstrucción con un parche de tamaño 8x8 (Izquierda) y un parche de tamaño 9x9 (Derecha). La mejora en el cambio del tamaño del parche o ventana es evidente visualmente.

TABLA III
COMPARACIÓN DE RESULTADOS CAMBIANDO EL TAMAÑO DE LOS PARCHES UTILIZADOS

Imagen	Ventana o parche de 8x8		Ventana o parche de 9x9	
	LASSO	BPDN	LASSO	BPDN
Imagen 1	35,201	35,711	36,273	35,543
Imagen 2	40,781	47,294	36,279	47,481

IV. CONCLUSIONES

Mediante este trabajo se inicia el trabajo en el área del procesamiento de campos de luz. Se realizaron simulaciones con imágenes sintéticas en busca de analizar los diferentes componentes que intervienen en la reconstrucción de campos de luz. Las simulaciones realizadas permitieron analizar el comportamiento de diferentes algoritmos así como la influencia de las imágenes de entrenamiento, el tamaño de los parches utilizados y los diccionarios que se entrenan en las reconstrucciones obtenidas.

BIBLIOGRAFÍA

1. Gortler SJ, Grzeszczuk R, Szeliski R, Cohen MF. The lumigraph. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96 [Internet]. 1996. p. 43–54. Disponible en: <http://portal.acm.org/citation.cfm?doid=237170.237200>
2. Adelson E, Bergen J. The plenoptic function and the elements of early vision. Computational models of visual processing . [Internet]. 1991;3–20. Disponible en: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6277403&abstractAccess=no&userType=inst
3. Levoy M, Hanrahan P. Light field rendering. Proc 23rd Annu Conf Comput Graph Interact Tech - SIGGRAPH '96 [Internet]. 1996;31–42. Disponible en: <http://portal.acm.org/citation.cfm?doid=237170.237199>
4. Lippmann G. Epreuves reversibles, photographies integrales. Comptes Rendus de l'Académie des Sciences [Internet]. 1908;444:446–51. Disponible en: http://jphystap.journaldephysique.org/articles/jphystap/abs/1908/01/jphystap_1908__7__821_0/jphystap_1908__7__821_0.html
5. Marwah K, Wetzstein G, Bando Y, Raskar R. Compressive light field photography using overcomplete dictionaries and optimized projections. ACM Trans Graph [Internet]. 2013;32:1. Disponible en: <http://dl.acm.org/citation.cfm?doid=2461912.2461914>
6. Chen SS, Donoho DL, Saunders MA. Atomic Decomposition by Basis Pursuit. SIAM Journal on Scientific Computing. 1998. p. 33–61.
7. Natarajan BK. Sparse Approximate Solutions to Linear Systems. SIAM Journal on Computing. 1995. p. 227–34.