

Metodología de adquisición y reconstrucción de vídeos espectrales mediante el uso de la teoría de muestreo compresivo y redes neuronales convolucionales

Camilo Andrés Calderón Carrillo y Pablo Andrés Gomez Toloza

Trabajo de grado para optar al título de Ingeniero de Sistemas e Ingeniero Electrónico

Director:

Ph.D(c) Miguel Angel Marquez Castellanos

Codirectores:

Ph.D Henry Arguello Fuentes

Ph.D Julian Rodríguez Ferreira

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingeniería de Sistemas e Informática

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2022

Dedicatoria

*A mi padre por su amor incondicional y darme
su apoyo, enseñanzas y sabiduría*

*A mi madre por ser una mujer luchadora, darme
sus consejos y demostrarme su amor cada día*

*A mi hermana por enseñarme sobre disciplina y
resiliencia y ser un ejemplo a seguir*

*A mi hermano gemelo por ser mi compañero y
amigo durante toda mi vida*

*A Jessica por apoyarme en cada paso que doy
y llenarme de su amor cada día*

*A Daniel, Santiago y Miguel por su cariño y
convertirse en mis hermanos*

- Camilo Calderón

Dedicatoria

A mi hermano por mantener la alegría en mi día a día, sin importar los problemas.

A mi padre por su apoyo incondicional, darme sus consejos, enseñarme de la disciplina, sabiduría y capacidad resolutiva para la vida.

A mi madre por enseñarme sobre la resiliencia, el trabajo duro y brindarme su amor incondicionalmente.

A mi novia Daniela, por acompañarme en todas mis decisiones, por brindarme su amor, apoyo y consejos cuando más los necesitaba.

En general a todas las personas, que de una u otra forma se vieron involucrados en mi formación tanto profesional como personal proporcionandome su apoyo.

- Pablo Gómez

Agradecimientos

A nuestro director Miguel Marquez por todo su tiempo, enseñanza y acompañamiento en nuestra formación personal y profesional.

Al profesor Julian Rodríguez, codirector del proyecto por su guía y recomendaciones.

Al profesor Henry Arguello, codirector del proyecto por brindarnos sus consejos y recomendaciones, y por ofrecernos la oportunidad de ser parte del grupo de investigación HDSP.

A los integrantes del grupo HDSP por todo el apoyo y paciencia brindado en nuestra formación académica y profesional.

A nuestros compañeros de carrera que han sido de gran importancia en nuestra vida tanto universitaria como profesional.

Tabla de Contenido

Introducción	19
1 Objetivos	23
2 Imágenes Espectrales Compresivas	24
2.1 Métodos de adquisición de SI	25
2.2 Muestreo compresivo	27
2.3 Códigos de apertura	29
2.4 Espejos deformables segmentados	30
2.5 Sistemas Tradicionales de Adquisición	31
2.5.1 SD-CASSI	31
2.5.2 DD-CASSI	35
2.5.3 CACTI	38
2.6 Algoritmos de filtrado	43
2.6.1 Filtro recursivo	43
2.6.2 Coincidencia de bloques y filtrado 3D	43
2.6.3 Filtrado de medios no locales	44
2.7 Redes Neuronales	44
2.7.1 Redes Neuronales Convolucionales	44

METODOLOGÍA DE ADQUISICIÓN Y RECONSTRUCCIÓN DE VIDEO ESPECTRAL	6
2.7.2 U-Net	45
2.7.3 ResNet	47
2.7.4 FFDNet	48
2.8 Algoritmo de Variación Total	50
3 Sincronización de Elementos Optoelectrónicos	51
3.1 Caracterización	51
3.1.1 Caracterización de Cámara Stingray F-080B	51
3.1.2 Caracterización de Monocromador TLS-300XR	54
3.2 Sincronización de Elementos Optoelectrónicos	59
3.2.1 Sincronización Cámara Stingray F-080B y Equipo de Computo.	59
3.2.2 Sincronización de Equipo de Computo, Microcontrolador Arduino y Servo Motor.	61
3.2.2.1 Servomotor	61
3.2.2.2 Microcontrolador Arduino	63
3.2.2.3 Implementación	64
3.2.3 Sincronización Equipo de Computo, Cámara Stingray F-080B, Microcontrolador Arduino y Servo Motor	66
3.2.4 Sincronización Equipo de Computo y Monocromador TLS-300XR	67
3.2.5 Sincronización Equipo de Computo, Monocromador, Cámara Stingray F-145B, Mi- crocontrolador Arduino y Servo Motor	69
4 Método Propuesto	71

4.1	Metodología de Adquisición	72
4.2	Matriz de sensado	75
4.3	Algoritmo de Reconstrucción PNP-ADMM	76
4.4	Reconstrucción Espectro-Temporal de SV	78
4.5	Captura en el Laboratorio de Vídeos Espectrales	80
5	Resultados	82
5.1	Conjunto de Datos de SV	82
5.2	Métricas de Evaluación	84
5.3	Análisis de Transmitancia	86
5.4	Resultados de Simulación	89
5.4.1	Análisis Espacio-Temporal	89
5.4.2	Análisis Espectro-Temporal	92
5.4.3	Análisis Espectral	94
5.4.4	Análisis para capturas experimentales de SV	102
5.4.5	Análisis Comparativo del método propuesto	109
6	Trabajo Futuro	111
7	Conclusiones	112
8	Apéndice	114
	Referencias Bibliográficas	116

Lista de Figuras

Figura 1	Representación de la división espectral de una imagen en 8 bandas con la respuesta espectral de dos puntos.	24
Figura 2	Comparación de una imagen (a) en escala de grises, (b) en representación RGB en el espectro electromagnético visible y (c) en el infrarrojo cercano. Adaptado de Monno et al. (2018).	25
Figura 3	Métodos de escaneo de puntos (A), de línea (B), de longitudes de onda (C) y única captura (D) para SI. Adaptado de Wang et al. (2017)	27
Figura 4	Codificación de una escena mediante código de apertura binario.	29
Figura 5	Espejo deformable segmentado (DMD)	30
Figura 6	Sistema óptico SD-CASSI adaptado de Marquez et al. (2019)	32
Figura 7	Matriz de Sensado del Sistema óptico SD-CASSI adaptado de Wagadarikar et al. (2008a)	34
Figura 8	Medidas del Sistema óptico SD-CASSI tomado de Wagadarikar et al. (2008a)	35
Figura 9	Sistema óptico DD-CASSI adaptado de Marquez et al. (2019)	36
Figura 10	Matriz de sensado del Sistema óptico DD-CASSI adaptado de Gehm et al. (2007)	37
Figura 11	Medidas del Sistema óptico DD-CASSI, datos adquiridos a partir de dataset KAIST Hwang et al. (2015), las cuales son ingresadas al sistema DD-CASSI como imágenes de 512×512 .	38

Figura 12	Sistema óptico CACTI adaptado de Llull et al. (2013)	39
Figura 13	Matriz de sensado del sistema óptico CACTI, adaptado a partir de Llull et al. (2013).	41
Figura 14	Medidas del Sistema óptico CACTI tomado de and Simulation (2020)	42
Figura 15	Red neuronal convencional conformada por una capa de entrada, una capa oculta y una capa de salida.	45
Figura 16	Arquitectura de la red convolucional U-NET para la segmentación de imágenes, adaptado de Jiao et al. (2020).	46
Figura 17	Arquitectura de la red convolucional residual ResNet para la estabilización del gradiente, adaptado de Talo (2019).	48
Figura 18	Arquitectura de la red neuronal FFDNet, adaptado de Zhang et al. (2018a)	49
Figura 19	Cámara Stingray F-080B.	52
Figura 20	Respuesta Espectral Cámara F-080B, donde se observa que cada curva es la respuesta espectral de la cámara respectiva a su color, y la curva gris es la respuesta espectral de la cámara a una iluminación monocromática.	53
Figura 21	Conectividad Cámara F-080B, en (a) se observa la conexión externa tipo trigger pin-hirose 12 pines, la cual comúnmente se utiliza para conexión de un generador de señales. En (b) aparece la conexión IEEE 1394 b normalmente utilizada para sincronización industrial .	54

Figura 22	Monocromador TLS-300XR que se encuentra en el laboratorio de óptica del grupo HDSP.	55
Figura 23	Resultados de caracterización espectral del monocromador obtenido a través de una conexión entre el monocromador TLS-300XR y el espectrómetro de onda de referencia Flame para espectro visible (VIS)	57
Figura 24	Conexiones que posee Fuente de alimentación del Monocromador, tanto para sincronización como para transmisión de información	58
Figura 25	Montaje de sincronización Computador, y Cámara F-080B, mediante conexión USB	59
Figura 26	Captura de imagen en escala de grises por medio de sincronización Computador y Cámara F-080B	60
Figura 27	Servomotor	61
Figura 28	Funcionamiento de servomotor de tipo 180°, tomado de Area Tecnología (2020)	62
Figura 29	Funcionamiento de servomotor de tipo 360°, adaptado de Control Automático Educación (2020).	63
Figura 30	Placa Arduino Uno/Mega, utilizada para aplicaciones de adquisición, tratamiento y transmisión de señales digitales y análogas.	64
Figura 31	Montaje de sincronización Computador, Arduino y Servomotor, por medio de conexión serial, realizando un giro de 180° por medio de pasos de 10°.	65
Figura 32	Montaje de sincronización Computador, Cámara Arduino y Servomotor, por medio de cable USB y conexión serial, respectivamente.	66

- Figura 33 Resultados de captura de video de 18 fotogramas en escala de grises por medio de sincronización Computador, Arduino y Servomotor. 67
- Figura 34 Montaje de sincronización Computador y Monocromador a través de cable USB. 68
- Figura 35 Montaje de sincronización Computador, Cámara, Monocromador, Arduino y Servomotor para la captura de videos espectrales. 69
- Figura 36 Metodología de adquisición propuesta basada en la combinación de un filtro tuneable, un DMD y un único sensor para la captura de SV comprimidos. 71
- Figura 37 Construcción del código de apertura, codificación de una escena y adquisición de medidas para el enfoque de ventana espacio-temporal con un tamaño de $S = 2$ 74
- Figura 38 Ejemplo de matriz de sensado para un código de apertura de tamaño $N_x = 5$, $N_y = 5$, $N_\lambda = 4$, $N_t = 1$. 76
- Figura 39 Flujo del método propuesto: El algoritmo recibe como entrada el SV, posteriormente, se obtienen las medidas compresas bajo la metodología propuesta y estas se pasan al algoritmo que implementa la FFDNet y la TV en el marco de trabajo del PNP. Luego, se obtiene como salida un vídeo en escala de grises como aproximación temporal que es utilizado para inicializar un PNP tradicional, el cual a partir de las medidas compresas y esta aproximación, reconstruye el vídeo espectral. 79
- Figura 40 Captura de SV con 16 bandas espectrales en el rango de 450-700 nm equiespaciadas y 32 fotogramas, a partir del sistema sincronizado en el laboratorio de óptica. 81

- Figura 41 Red Neuronal HRNet, ingresando vídeos reales RGB con tamaños desde 1280×720 hasta 1920×1080 , finalizando el procesamiento con vídeos espectrales de tamaño $1024 \times 1024 \times 16 \times 32$ 83
- Figura 42 Resultados de diagramas de caja para 10 reconstrucciones de video espectrales usando (a) PSNR y (b) SSIM. En (c) se muestran los códigos de apertura a diferentes valores de transmitancia. 87
- Figura 43 Resultados para un SV de un juego de billar en donde se aprecian los fotogramas originales y la reconstrucción temporal en escala de grises. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente. 90
- Figura 44 Resultados para un SV de un balón de fútbol rebotando en donde se aprecian los fotogramas originales y la reconstrucción temporal en escala de grises. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente. 91
- Figura 45 Resultados espectro-temporales SV de un juego de billar, donde se muestran los fotogramas originales y la reconstrucción temporal en representación RGB. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente. 93
- Figura 46 Resultados espectro-temporales para un SV de un balón de fútbol rebotando en donde se muestran los fotogramas originales y la reconstrucción temporal en representación RGB. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente. 95

- Figura 47 Resultados espectrales del SV con la escena de un juego de billar en la que se observan 6 bandas espectrales para la escena original y la reconstrucción. Se incluyen las métricas de PSNR y SSIM con las longitudes de onda de cada banda correspondiente. 96
- Figura 48 Respuesta espectral de 3 regiones de interés de un fotograma espectral para la escena de un SV de un juego de billar. Se muestra la firma espectral original con su reconstrucción y la métrica SAM. 98
- Figura 49 Resultados espectrales del SV con la escena de un balón de fútbol en la que se observan 6 bandas espectrales para la escena original y la reconstrucción. Se incluyen las métricas de PSNR y SSIM con las longitudes de onda de cada banda correspondiente. 99
- Figura 50 Respuesta espectral de 3 regiones de interés de un fotograma espectral para la escena de un SV de un balón de fútbol. Se muestra la firma espectral original con su reconstrucción y la métrica SAM. 101
- Figura 51 Resultados para un SV capturado experimentalmente de un Lego girando en donde se aprecian los fotogramas originales y la reconstrucción temporal en escala de grises. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente. 103
- Figura 52 Resultados espectro-temporales para un SV capturado experimentalmente de un Lego girando en donde se muestran los fotogramas originales y la reconstrucción temporal en representación RGB. Se incluyen las métricas de SSIM y SAM con los números de cada fotograma correspondiente. 105

Figura 53 Resultados espectrales del SV con la escena de un lego girando en la que se observan 6 bandas espectrales para la escena original y la reconstrucción. Se incluyen las métricas de PSNR y SSIM con las longitudes de onda de cada banda correspondiente. 106

Figura 54 Respuesta espectral de 3 regiones de interés de un fotograma espectral para la escena de un SV de un Lego girando. Se muestra la firma espectral original con su reconstrucción y la métrica SAM. 108

Lista de Tablas

Tabla 1	Comparación de los resultados obtenidos para la reconstrucción de 10 SV usando el método tradicional y la metodología propuesta. Se usan los promedios alcanzados para las métricas de PSNR, SSIM y SAM y se muestra la ganancia que se obtiene en cada una de las métricas al usar el enfoque propuesto.	110
---------	---	-----

Lista de Apéndices

	pág.
Apéndice A Función de Sincronización de elementos optoelectronicos.	114
Apéndice B Función de Sincronización de Servomotor al Sistema.	115

Resumen

Título: Metodología de adquisición y reconstrucción de videos espectrales mediante el uso de la teoría de muestreo compresivo y redes neuronales convolucionales *

Autores: Camilo Andrés Calderón Carrillo ** y Pablo Andres Gomez Toloza ***

Palabras Clave: Muestreo compresivo, Video espectral, Aprendizaje profundo, Sistema óptico.

Descripción: La adquisición de video espectral a partir de la teoría de muestreo compresivo (CSV, de sus siglas en inglés *compressive spectral video*) se ha convertido en un tema de investigación de gran interés en la comunidad de visión por computadora, debido a que estos sistemas permiten comprimir la información espacial, espectral y temporal en un conjunto de imágenes 2D para posteriormente recuperarla usando algoritmos de reconstrucción. Sin embargo, los enfoques de CSV se han limitado a comprimir únicamente la dimensión espectral de cada cuadro, es decir, no se desarrolla compresión temporal. Los enfoques alternativos son los sistemas de brazo doble basados en un generador de imágenes espectrales y temporales compresivas que fusiona las mediciones adyacentes para recuperar el vídeo espectral. Por lo tanto, es deseable el desarrollo de un generador de imágenes CSV que permita recuperar un vídeo espectral a partir de una única medición comprimida. Este trabajo propone una metodología de adquisición y compresión de video espectral mediante el uso de un elemento de codificación espacial binario tipo segmentado y un filtro tuneable. A través del sistema óptico propuesto, tanto la dimensión espectral como la temporal se pueden modular durante un tiempo de integración utilizando solo dos elementos ópticos.

* Trabajo de grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de sistemas e Informática. Director: Ph.D(c) Miguel Angel Marquez Castellanos.

*** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones. Director: Ph.D(c) Miguel Angel Marquez Castellanos.

Abstract

Title: Spectral video acquisition and reconstruction methodology using compressive sensing theory and convolutional neural networks. *

Authors: Camilo Andrés Calderón Carrillo ** and Pablo Andres Gomez Toloza ***

Keywords: Compressive sensing, Spectral video, Deep learning, Optical system.

Description: Spectral video acquisition from compressive sensing theory (CSV) has become a high-interest research topic in the computer vision community due to the fact that these systems allow the spatial, spectral, and temporal information to be compressed into a set of 2D images to subsequently recover the spectral video using reconstruction algorithms. However, CSV approaches have been only limited to compress the spectral dimension for each frame, i.e., temporal compression is not performed. Alternative approaches are dual-arm systems that rely on a compressive temporal and spectral imager that fuse adjacent measurements to retrieve the spectral video. Therefore, the development of a CSV imager that allows a spectral video to be retrieved from a single compressed measurement is highly desired. This work proposes a spectral video acquisition and compression methodology through the use of a segmented-type binary spatial coding element and a tunable filter. Through the proposed optical system, both the spectral and temporal dimensions can be modulated during an integration time using only two optical elements.

* Degree Work

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de sistemas e Informática. Director: Ph.D(c) Miguel Angel Marquez Castellanos.

*** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones. Director: Ph.D(c) Miguel Angel Marquez Castellanos.

Introducción

Los vídeos espectrales (SV, de sus siglas en inglés spectral video) son representados como función de cuatro dimensiones $f(x, y, \lambda, t)$ donde x e y representan una dimensión espacial, λ la dimensión espectral y t la temporal. Los SV se pueden representar como un tensor de cuarto orden donde las bandas representan la información espectral en un rango arbitrario del espectro electromagnético y para un intervalo de tiempo. Los SV contienen información que permite determinar distintas características sobre los elementos que componen una escena así como sus cambios espectrales a lo largo del tiempo, motivo por el cual se han convertido en una herramienta de gran interés en aplicaciones comerciales y de investigación, tales como, detección de objetivos Chang (2016), análisis de tejidos Leitner et al. (2013), detección de materiales Liu et al. (2019a), tomografía óptica Nassif et al. (2004), monitoreo clínico Scabbia et al. (2020), entre otras aplicaciones. Sin embargo, dadas sus características multidimensionales, la adquisición de SV requiere de altas capacidades de almacenamiento, debido a esto, tanto la transmisión como el almacenamiento de los datos adquiridos podrían resultar complejos y costosos computacionalmente a medida que aumenta su resolución espacial, espectral y temporal, lo que implica la necesidad de equipos de cómputo potentes, detectores sensibles y un mayor ancho de banda debido a que los cubos generados son tensores de datos de gran tamaño y de una alta dimensionalidad. Por lo tanto, resulta útil reducir el tamaño de la información para disminuir los costos de cómputo y tratamiento de los vídeos capturados.

Para resolver los problemas relacionados a la adquisición de SV los investigadores han abordado el estudio de la teoría del muestreo compresivo (CS, de sus siglas en inglés *Compressive Sensing*) Donoho (2006), para el desarrollo de metodologías de adquisición que permitan capturar la información espectral y temporal a partir de pocas medidas. Con base en la teoría de CS se han propuesto sistemas ópticos que permiten proyectar la información espacial-temporal o la información espacial-espectral a una medida bidimensional sin tener pérdidas significativas de información. Este proceso de compresión de la información se logra tradicionalmente mediante el uso de moduladores de luz, tales como, los espejos deformables segmentados (DMD, de sus siglas en inglés *digital micromirror device*), pantallas de cristal líquido (LCOS, de sus siglas en inglés *Liquid crystal on silicon*), espejos deformables continuos (DM, de sus siglas en inglés *Deformable Mirrors*), entre otros. Entre estos elementos de modulación el más popular y mayormente usado en los enfoques de CS espectral o temporal han sido los DMD Wu et al. (2011b), debido a su alta resolución espacial, bajo costo, y sus altas velocidades de modulación. Además del uso de DMD, los códigos de apertura de color también son utilizados como moduladores de luz puesto que en ellos cada píxel actúa como un filtro óptico y la codificación de la información depende de la respuesta espectral de los mismos.

Basado en los DMD se han propuesto múltiples sistemas ópticos para el muestreo compresivo de la información temporal y espectral entre los que se encuentran las técnicas de estimación numérica para imágenes espectrales con diseño de apertura codificada (CASSI, de sus siglas en inglés *coded aperture snapshot spectral imager*) Wagadarikar et al. (2008b). Entre esta técnica de muestreo se encuentra el método de diseño de dispersor dual para imágenes espectrales ins-

instantáneas de apertura codificada (DD-CASSI, de sus siglas en inglés *dual disperser design for coded aperture snapshot spectral imaging*) Gehm et al. (2007), que consta de un doble dispersor para la obtención de las imágenes espectrales y el método de diseño de dispersor único para imágenes espectrales instantáneas de apertura codificada (SD-CASSI, de sus siglas en inglés *single disperser design for coded aperture snapshot spectral imaging*) Wagadarikar et al. (2008a), el cual se diferencia del anterior dado que se tiene un único dispersor para la misma tarea. Por otro lado, para el muestreo temporal se tienen técnicas como la de imágenes temporales compresivas de apertura codificada (CACTI, de sus siglas en inglés *coded aperture compressive temporal imaging*) Llull et al. (2013), el cual se trata de una traslación de la apertura codificada durante la exposición, por lo tanto cada plano temporal del vídeo se modula por una versión del código, realizando una modulación por píxel sin utilizar un sensor con ancho de banda mayor. Por otro lado una variación reciente del CACTI es la imagen espectro-temporal de compresión de apertura codificada (CACSTI, de sus siglas en inglés *coded aperture compressive spectral-temporal imaging*) Tsai et al. (2015), que emplea la traslación mecánica de una apertura codificada y una dispersión espectral para capturar la escena dinámica multispectral en un detector monocromático.

Los enfoques del estado del arte para compresión de SV mencionados anteriormente realizan esta tarea mediante dos cámaras diferentes en donde una se encarga de comprimir la dimensión espectral y la otra la dimensión temporal. Debido a esto, este trabajo propone una metodología de adquisición y reconstrucción de vídeo espectral que permita realizar la compresión de la información multidimensional a través de un sistema óptico compuesto principalmente por un elemento

de codificación binario (DMD) y un filtro tuneable con los que se puede realizar una modulación espectro-temporal. Posteriormente, el tensor 4D puede estimarse mediante el uso de algoritmos de reconstrucción, eliminación de ruido y redes neuronales manteniendo la calidad de la reconstrucción del vídeo y reduciendo de forma efectiva el ancho de banda necesario para transmitir esta información.

1. Objetivos

Objetivo General

- Diseñar una metodología de adquisición comprimida y reconstrucción de vídeos espectrales a partir del uso de elementos moduladores de luz tipo espejo segmentados deformables junto con el uso de redes neuronales profundas.

Objetivos específicos

- Implementar una red neuronal convolucional para la estimación de la información espacio-temporal a partir de medidas comprimidas.
- Implementar un sistema optoelectrónico que permita la adquisición de vídeos espectrales a partir de un protocolo de muestreo tipo barrido espectral y fotograma a fotograma.
- Implementar una red neuronal convolucional del estado del arte para crear imágenes espectrales a partir de imágenes RGB.
- Verificar las reconstrucciones realizadas por el sistema neuronal mediante medidas reales del laboratorio.

2. Imágenes Espectrales Compresivas

Las imágenes espectrales (SI, de sus siglas en inglés spectral image) corresponden a la información espacio-espectral de una escena que es capturada en un rango arbitrario del espectro electromagnético, llamados bandas espectrales. A diferencia de las cámaras e imágenes digitales tradicionales en escala de grises, las SI pueden incluir el rango de luz visible, infrarrojo y luz ultravioleta, lo que permite obtener información que no es visible por el ojo humano, a través de los valores de intensidad en las longitudes de onda discretas capturadas.

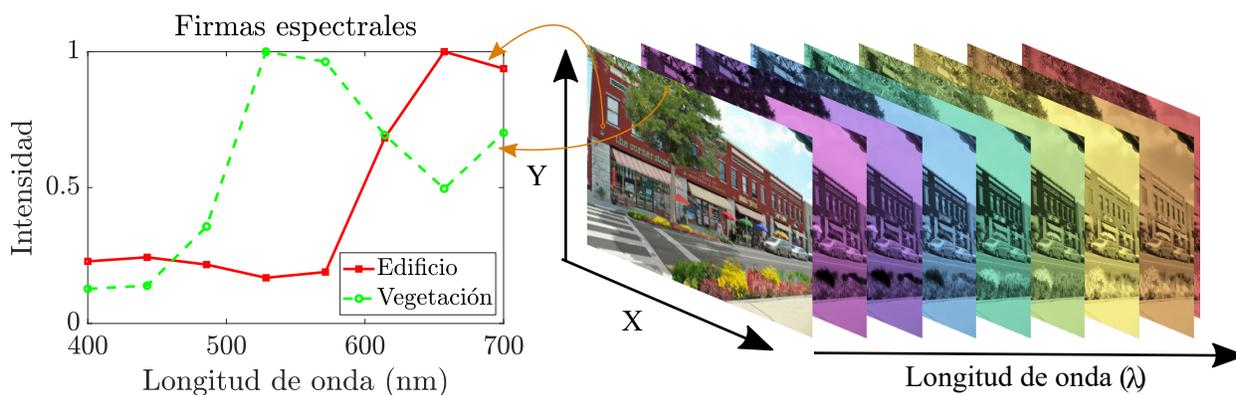


Figura. 1. Representación de la división espectral de una imagen en 8 bandas con la respuesta espectral de dos puntos.

La figura 1 muestra un ejemplo de imagen espectral, aquí, la primera imagen corresponde a la representación RGB de la escena, seguida de su división en 8 bandas espectrales. Las SI capturadas se almacenan como cubos de datos $f(x, y, \lambda)$ que contienen la información espacial en x e y , y la información espectral, con múltiples bandas discretizadas, en λ .

Esto hace posible identificar distintos tipos de materiales y objetos que se encuentren en la escena a través de sus firmas espectrales, lo que significa que entre más información espectral

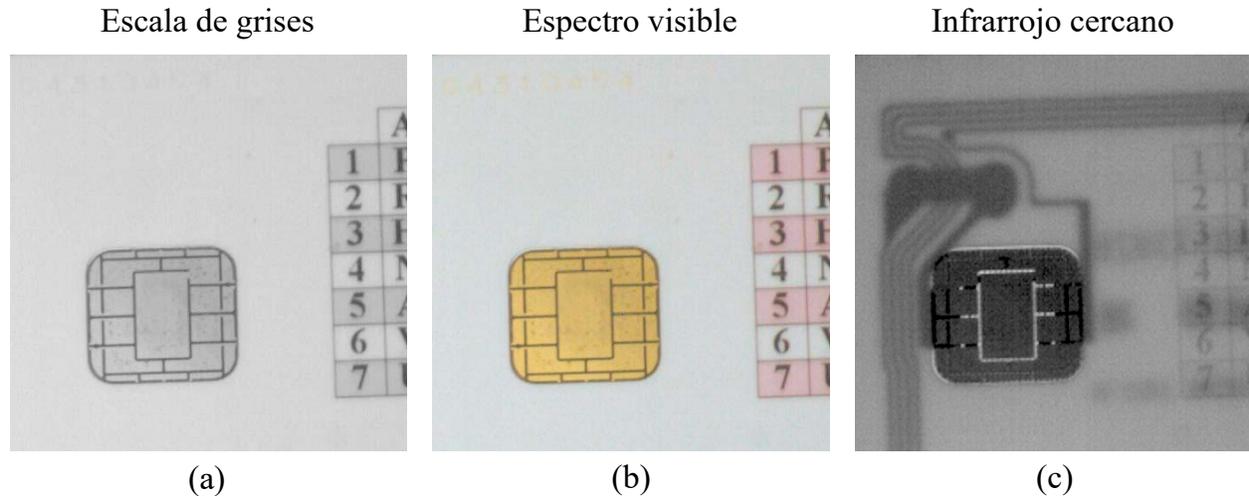


Figura. 2. Comparación de una imagen (a) en escala de grises, (b) en representación RGB en el espectro electromagnético visible y (c) en el infrarrojo cercano. Adaptado de Monno et al. (2018).

se capture, más detalles pueden obtenerse. La figura 2 muestra una comparación entre distintas representaciones de una imagen, en (a) se observa la foto de una tarjeta inteligente en escala de grises donde muy pocos detalles puede observarse, como su forma y características macroscópicas, a su vez, en (b) se aprecia la tarjeta a color que no aporta más que los colores de los números o el chip, finalmente, en (c) se muestra la imagen en el rango del infrarrojo cercano, en la que por ejemplo es posible ver los circuitos integrados en la tarjeta, que no son claros a simple vista. Sin embargo, capturar imágenes a diferentes longitudes de onda implica un mayor costo computacional en su captura, tiempo de adquisición, tratamiento y almacenamiento. A continuación se introducen 4 métodos tradicionales de muestreo espectral por barrido.

2.1. Métodos de adquisición de SI

Las SI suelen obtenerse a través de sensores que capturan la señal y la discretizan en un número determinado de bandas, según las características de este y la tarea de aplicación específica,

para finalmente ser capturadas por el detector. La figura 3 muestra los distintos sistemas ópticos para adquirir estas imágenes a través de la combinación de elementos dispersivos como los prismas, detectores 2D y filtros de colores.

Específicamente en estos sistemas se encuentran las basadas en escaneo y las de área extensa ("Wide-Field"). En el primer grupo se realiza un muestreo a través de un rayo láser enfocado, ya sea en forma de punto o línea. Para el escaneo de puntos Vane et al. (1993) mostrado en la figura 3A, se utiliza un láser en forma de punto para capturar las longitudes de onda de las dimensiones espaciales x e y de la escena. Por otro lado, el escaneo de líneas Gupta and Hartley (1997) mostrado en la figura 3B, realiza un barrido de la escena utilizando un láser en forma de línea, para moverse a lo largo de una dimensión x o y y capturar las longitudes de ondas de múltiples puntos. Los métodos anteriormente mencionados logran obtener una alta resolución espectral por cada punto o línea de píxeles, permitiendo una demultiplexación y cuantificación precisa de las muestras de la imagen, pero en contra posición teniendo una baja velocidad de captura.

Por las dificultades anteriormente expuestas para los métodos de escaneo, otro tipo de sistemas fueron desarrollados para la captura de SI conocidos como de campo amplio, los cuales utilizan una iluminación de la región y detecta la luz emitida desde un grupo de píxeles dentro de un área usando una matriz 2D en el detector, como se observa en la figura 3C y 3D. En el primer caso de escaneo de longitudes de onda Gat (2000), se realiza una adquisición de imágenes bidimensionales capturando secuencialmente una banda λ completa de la escena, usando filtros de color. Por otro lado el método de única captura Arce et al. (2013) que usa múltiples prismas para capturar la información tanto espacial como espectral en un sólo disparo. En comparación

con los métodos de escaneo, estos métodos de campo amplio eliminan la necesidad de un escaneo realizado mecánicamente, adquiriendo imágenes de una gran región con alta resolución espacial en una única captura, aumentando su velocidad de captura.

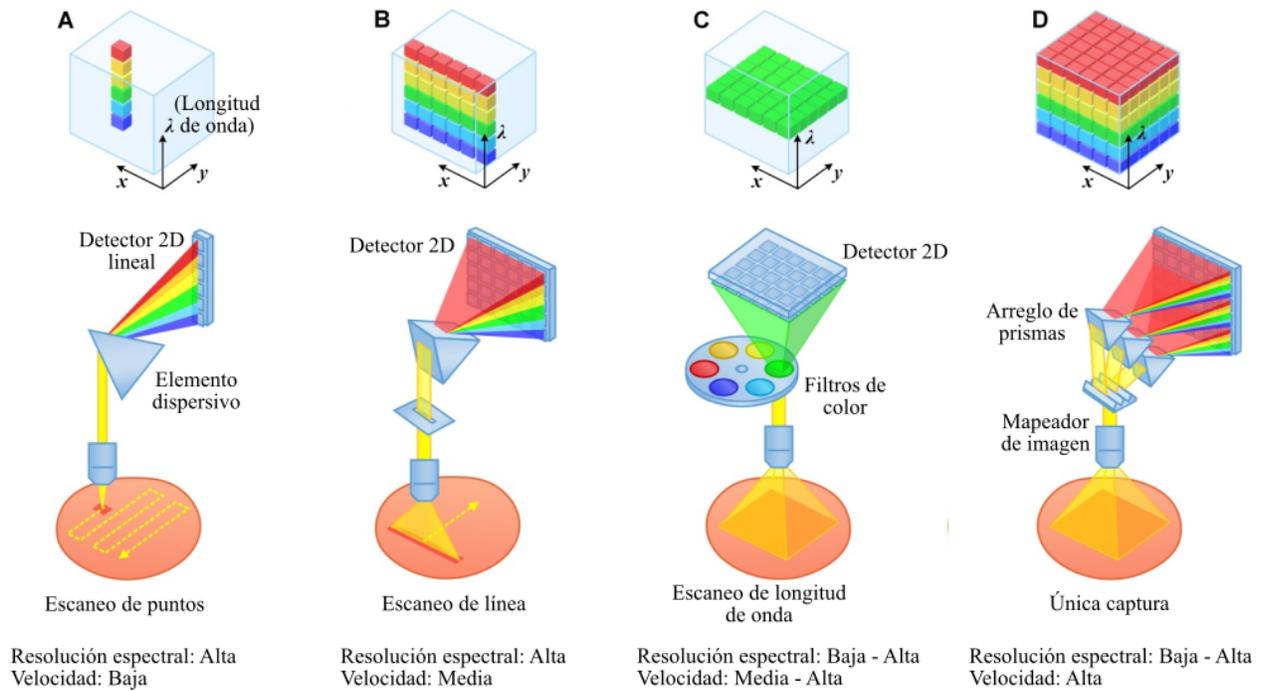


Figura. 3. Métodos de escaneo de puntos (A), de línea (B), de longitudes de onda (C) y única captura (D) para SI. Adaptado de Wang et al. (2017)

2.2. Muestreo compresivo

A pesar de su amplio uso, la captura de SI presenta una carencia de información temporal que permiten obtener características importantes para analizar el comportamiento espectral de una escena a lo largo del tiempo. No obstante, incluir la dimensión temporal presentaría una mayor complejidad para su captura, almacenamiento y tratamiento al tratarse de un SV representado por una matriz 4D, lo que implica la necesidad de nuevas técnicas y métodos para su aplicación. En el desarrollo de metodologías de adquisición y reconstrucción de SV, el muestreo compresivo

ha formado un papel fundamental para el tratamiento de estos datos de alta resolución puesto que presenta una evolución al teorema de muestreo de Shannon-Nyquist. Específicamente, el CS plantea un método en el que una señal adquirida puede ser comprimida al momento de su detección, muestreando la señal a una velocidad muy por debajo de la tasa de Nyquist, en este campo destacan sistemas ópticos como el SD-CASSI Wagadarikar et al. (2008a), DD-CASSI Gehm et al. (2007), CACTI Lull et al. (2013) y CACSTI Tsai et al. (2015). En especial, el CS aplicado a las imágenes espectrales crean un campo denominado imágenes espectrales compresivas (CSI, de sus siglas en inglés *Compressive Spectral Imaging*) el cual establece que se puede estimar la información de una imagen espectral a partir de un conjunto de medidas más pequeñas bajo la condición de que esta imagen sea escasa, es decir, que tenga en su mayoría elementos de valor 0, en alguna base de representación.

Un SV puede denotarse como $\mathbf{F} \in \mathbb{R}^{N \times M \times L \times T}$ representando un tensor discreto de cuatro dimensiones que contiene la información espacial, espectral y temporal donde M y N corresponden a las dimensiones espaciales, L a la dimensión espectral y T corresponde a la dimensión temporal y, sea $\mathbf{f} \in \mathbb{R}^n$ la forma vectorial de \mathbf{F} con $n = NMLT$, teniendo en cuenta la teoría del CS y el CSI se puede establecer la ecuación:

$$\mathbf{f} = \Psi\theta, \quad (1)$$

donde $\theta \in \mathbb{R}^{n \times 1}$ es la representación escasa de las medidas \mathbf{f} en la base de representación $\Psi \in \mathbb{R}^{n \times n}$. De esta forma \mathbf{f} podrá estimarse a partir de la combinación lineal de S vectores de Ψ con

$S \ll MNL T$ y con lo cual la imagen se puede obtener a partir de una menor cantidad de medidas.

2.3. Códigos de apertura

Los códigos de apertura son un elemento considerablemente usado en la captura de imágenes compresivas, estos representan una máscara con patrones inscritos en los que, a través de la opacidad del material usado, se bloquea o no la radiación de una onda electromagnética, de esta forma, se produce una codificación de la imagen.

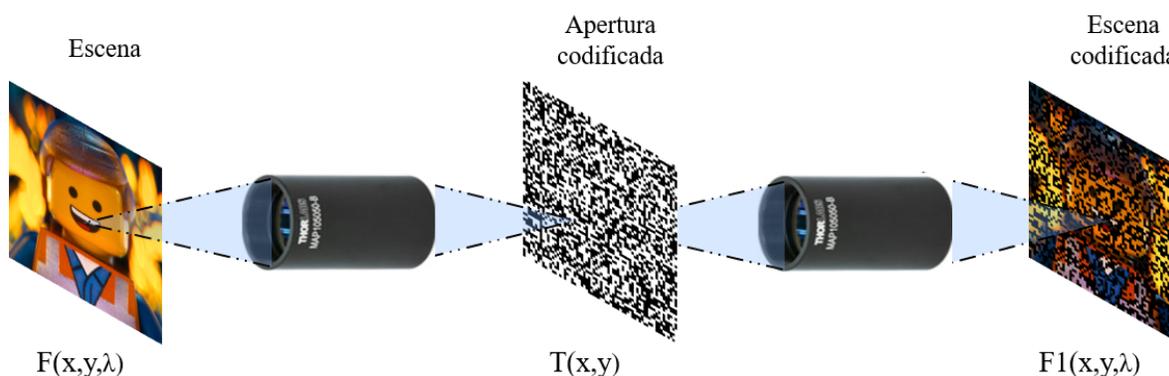


Figura. 4. Codificación de una escena mediante código de apertura binario.

La figura 4 muestra un ejemplo del efecto de una apertura codificada sobre una escena, aquí, la luz es bloqueada por las partes más oscuras de la máscara. El uso de estos elementos tiene gran relevancia en aplicaciones como la tomografía computarizada, en la que los rayos X inciden sobre el cuerpo de una persona, produciendo que esta reciba una gran cantidad de energía ionizante que resulta nociva para su salud. Por esto, mediante una máscara codificada, se puede reducir la cantidad de energía que se irradia y si se conoce el patrón usado, es posible recuperar matemáticamente la fuente original con una alta fiabilidad.

2.4. Espejos deformables segmentados

Para la adquisición de datos de una escena los investigadores han desarrollado diversas técnicas por medio del uso de moduladores de luz, uno de más utilizados son los espejos deformables segmentados (DMD) los cuales constan de una matriz de cientos de miles de microespejos de aluminio, de aproximadamente 16 (μm) de ancho con forma rectangular. Cada espejo de estos representa un píxel de la imagen que se desea sensar.

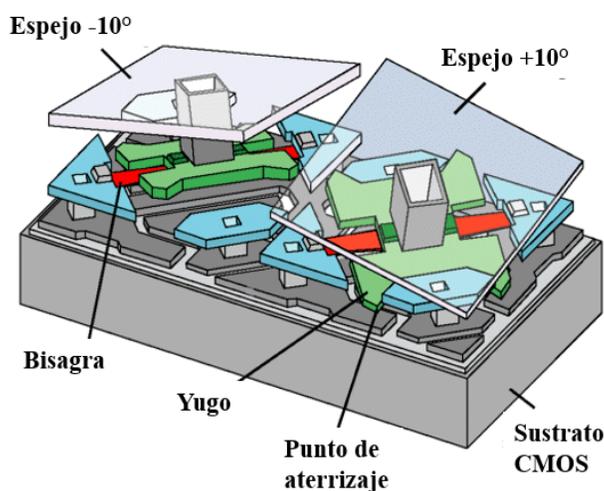


Figura. 5. Espejo deformable segmentado (DMD)

El funcionamiento de este dispositivo opto electrónico, se lleva a cabo debido a que cada microespejo tiene la posibilidad de girar $\pm 10 - 12^\circ$, como lo muestra la figura 5. Donde dependiendo de esta posición la luz procedente de la imagen se refleja o atraviesa el lente, es decir, si el espejo se encuentra en posición encendido, este se levanta y luz procedente de la imagen se refleja y permite que este píxel sea sensado. Por otro lado, si la posición es apagada, el microespejo dirige el haz hacia un disipador de calor, lo cual hace que ese píxel en específico no sea sensado y que

por lo tanto la información de este píxel no se tenga en cuenta.

Dado la relación entre el número de muestras tomadas por el DMD, y la teoría de sensado compresivo, este dispositivo ha sido utilizado ampliamente por los investigadores tanto para modulación temporal como espectral Rueda et al. (2015); Wu et al. (2011a,c), ya que entregan una alta resolución espacial y velocidad de modulación a un bajo costo.

2.5. Sistemas Tradicionales de Adquisición

Diversos sistemas ópticos basados en micro espejos deformables DMD, han sido propuestos para la adquisición comprimida de información espectral-temporal. Entre estos podemos destacar los basados en la técnica de estimación de imágenes espectrales de apertura codificada CASSI, el cual garantiza que toda la información sea detectada, de este método se destacan principalmente dos sistemas ópticos, el SD-CASSI y el DD-CASSI. Adicionalmente, para el dominio temporal sobresale la arquitectura óptica de vídeo CACTI, desarrollada para la adquisición comprimida de datos. Estos sistemas ópticos serán estudiados a continuación.

2.5.1. SD-CASSI. El sistema dispersor único para imágenes espectrales de apertura codificada SD-CASSI se muestra en la figura 6, el cual en primer lugar consta de la escena espacio-espectral que se desea capturar. Posteriormente, se usa un código de apertura que permite el paso de un porcentaje de la información de cada firma espectral, según el valor de transmitancia que se fije. Luego, a través de un prisma es posible modular la escena codificada la cual según el coeficiente de refracción del elemento dispersivo y de la longitud de onda presenta una dispersión. Finalmente, la luz resultante pasa por un sensor que integra todos los datos obteniendo las medidas compresas. Este sistema entrega una codificación espacial con baja complejidad respecto a su

variante de sistema óptico DD-CASSI, obteniendo valores de PSNR aceptables Cao et al. (2016).

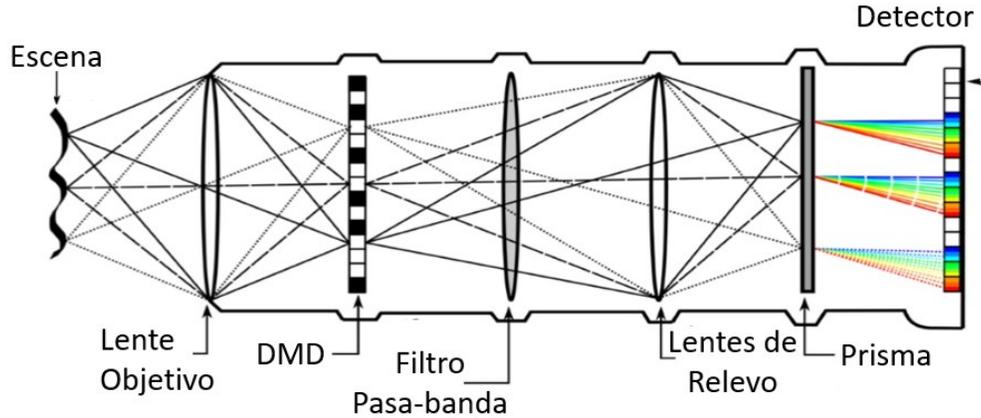


Figura. 6. Sistema óptico SD-CASSI adaptado de Marquez et al. (2019)

En este sistema óptico el arreglo de detectores es insensible a la longitud de onda, por lo que mide la intensidad de la luz que incide sobre estos en lugar de la densidad espectral. De esta forma, la medida continua adquirida se puede representar como

$$g(x,y) = \int f_0(x + \alpha(\lambda - \lambda_c), y; \lambda) T(x + \alpha(\lambda - \lambda_c), y) d\lambda, \quad (2)$$

donde $f_0(x, y; \lambda)$ representa la fuente espectral que entra al detector, α denota la dispersión lineal del elemento dispersivo en una longitud de onda central λ_c , x e y denotan las dimensiones espaciales y λ la dimensión espectral. Además, $T(x, y)$ es el código de apertura que puede ser representado como una matriz espacial de agujeros pequeños (pinhole de su nombre en inglés), cada uno de los cuales tiene una longitud lateral de Δq y donde $t_{n', m'}$ que representa un punto abierto o cerrado en

la posición (n', m') en la matriz de agujeros

$$T(x, y) = \sum_{m', n'} t_{n', m'} \text{rect} \left(\frac{x}{\Delta q} - m', \frac{y}{\Delta q} - n' \right). \quad (3)$$

Con esta representación del código de apertura, teniendo en cuenta la presencia de ruido w y debido a que la matriz de detectores está espacialmente pixelada con un tamaño de píxel Δ , las medidas $g(x, y)$ se muestrean en ambas dimensiones del detector, por lo que pueden representarse como

$$g_{mn} = \sum_{m', n'} t_{n', m'} \int \int \int \text{rect} \left(\frac{x + \alpha(\lambda - \lambda_c)}{\Delta q} - m', \frac{y}{\Delta q} - n' \right) \times \text{rect} \left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n \right) \times f_0(x + \alpha(\lambda - \lambda_c), y; \lambda) dx dy d\lambda + w_{nm}. \quad (4)$$

Así, denotando la densidad espectral de la fuente $f_0(x, y; \lambda)$ en su forma discreta f_{ijk} y el código de apertura $T(x, y)$ como t_{ij} , las mediciones en forma matricial para una posición (n, m) se pueden escribir como

$$g_{nm} = \sum_k f_{(m+k)nk} t_{(m+k)n} + w_{nm} = (Hf)_{nm} + w_{nm}, \quad (5)$$

donde H es un operador lineal conocido como matriz de sensado y formado por el código de apertura, el cual representa el modelo del sistema mostrado en la figura 7. Asimismo, en dicha figura se aprecia debido a la ubicación de los elementos ópticos y el funcionamiento de la codificación la matriz de sensado posee una variación específica de traslación entre dos bandas consecutivas,

dicha variación es exactamente de N y aumenta según el número de banda en el que se encuentre capturando.

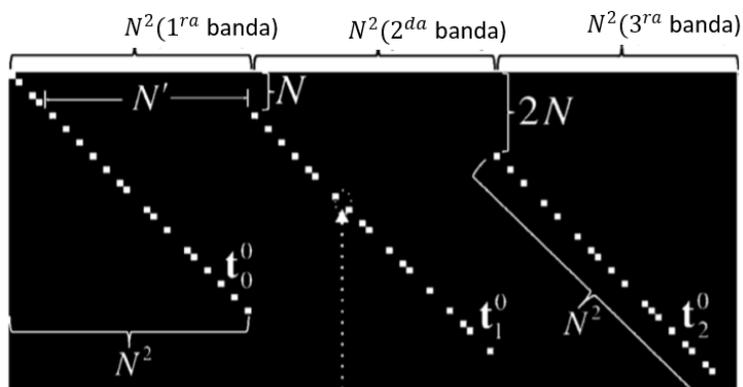


Figura. 7. Matriz de Sensado del Sistema óptico SD-CASSI adaptado de Wagadarikar et al. (2008a)

La figura 8 muestra una comparación entre la escena original y la medición adquirida a partir del sistema SD-CASSI de la figura 6, donde se observa en (a) un par de bolas de ping pong las cuales fueron iluminadas con un láser verde de 543 nm y con una fuente de luz blanca con un filtro de 560 nm y en (b) la medida adquirida. Nótese que debido al funcionamiento del sistema se produce un solapamiento espacio-espectral de las imágenes que fueron moduladas por el código de apertura. Asimismo, es necesario resaltar que el cubo de datos resultante es de $128 \times 128 \times 28$ con un rango espectral de 540 a 640 nm.

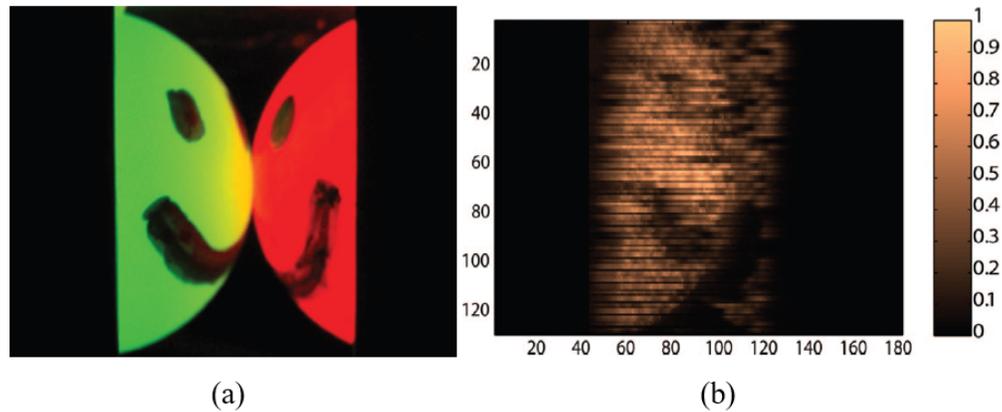


Figura. 8. Medidas del Sistema óptico SD-CASSI tomado de Wagadarikar et al. (2008a)

2.5.2. DD-CASSI. Una variación del sistema CASSI es el diseño de doble dispersor para la obtención de imágenes espectrales de apertura codificada DD-CASSI. En la implementación de este tipo de sistema inicialmente se tiene la escena deseada para la captura, seguida de la dispersión de la luz por medio de un prisma, la cual es modulada posteriormente a través de una apertura codificada como se observa en la figura 9. Adicionalmente, una vez se realiza la modulación ocurre una nueva dispersión con un prisma de las mismas características que difracta la luz. Esta es una diferencia notable en comparación con el sistema anterior SD-CASSI donde se tenía una modulación horizontal, en este caso se obtiene una modulación diagonal, que llega finalmente al sensor el cual se encarga de integrar la información. Este sistema realiza tanto una codificación espacial como espectral, obteniendo valores mayores de PSNR con respecto al SD-CASSI Cao et al. (2016), sin embargo, consta de un mayor número de elementos ópticos.

Al igual que para el SD-CASSI, en este sistema, la matriz de detectores presenta una característica de insensibilidad a la longitud de onda, por lo que la medida adquirida no es la densidad

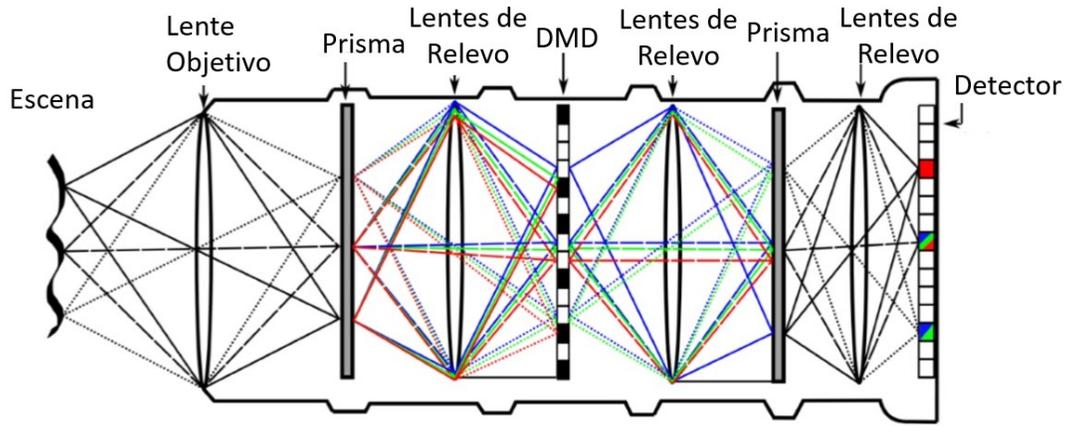


Figura. 9. Sistema óptico DD-CASSI adaptado de Marquez et al. (2019)

espectral sino la intensidad definida en su forma continua como

$$g(x, y) = \int H(x, y; \lambda) S_0(x, y; \lambda) \quad (6)$$

donde $S_0(x, y; \lambda)$ es la densidad espectral que ingresa al sistema y $H(x, y; \lambda) = T(x - \alpha(\lambda - \lambda_c), y)$ corresponde a la función de filtro de densidad espectral. En esta última, α y λ_c son variables de la propagación óptica de imágenes en un elemento dispersivo, correspondientes a la dispersión lineal y la longitud de onda central, respectivamente.

Además, se debe considerar que la matriz de detectores está espacialmente pixelada, si tomamos el tamaño de densidad de píxel como Δ , las medidas en el plano del detector se pueden definir como

$$g_{nm} = \int \int \int \text{rect} \left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n \right) H(x, y; \lambda) S_0(x, y; \lambda) dx dy d\lambda. \quad (7)$$

También, es necesario considerar una función de transmisión del código de apertura $T(x,y)$ en la que este se encuentre pixelado con un tamaño igual al los pixeles del detector

$$T(x,y) = \sum_{m',n'} T_{m',n'} \text{rect} \left(\frac{x}{\Delta} - m', \frac{y}{\Delta} - n' \right) \quad (8)$$

Según esta definición y la representación filtro de densidad espectral $H(x,y;\lambda)$, podemos realizar una representación más completa de las medidas definiendo la ecuación (7) como

$$g_{nm} = \sum_{m',n'} \int \int \int \text{rect} \left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n \right) \text{rect} \left(\frac{x - \alpha(\lambda - \lambda_c)}{\Delta} - m', \frac{y}{\Delta} - n' \right) \times T_{n',m'} S_0(x,y;\lambda) dx dy d\lambda. \quad (9)$$

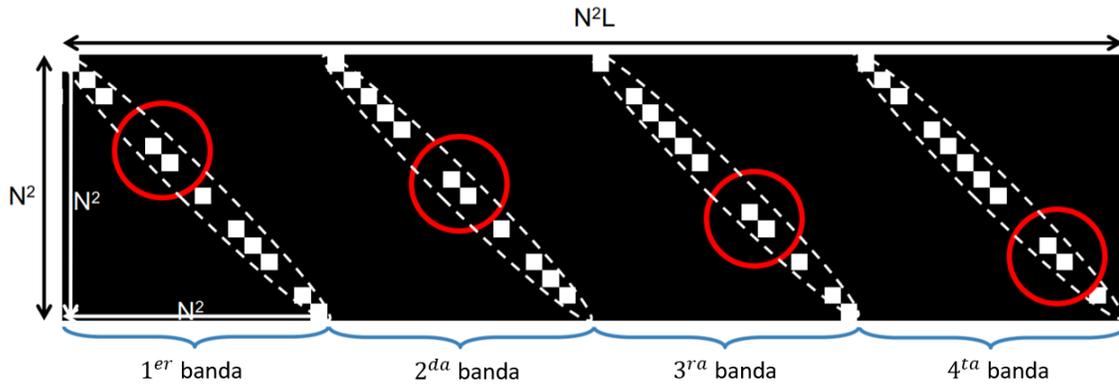


Figura. 10. Matriz de sensado del Sistema óptico DD-CASSI adaptado de Gehm et al. (2007)

En la figura 10, se observa la matriz de sensado procedente a dicho sistema óptico, donde se aprecia una característica remarcable en sus variaciones entre banda y banda de la escena, y se concluye que ocurre una traslación de los valores en las diagonales de la matriz, esto en el

contexto de que el sistema óptico realiza un proceso de dispersión doble debido a los prismas implementados.

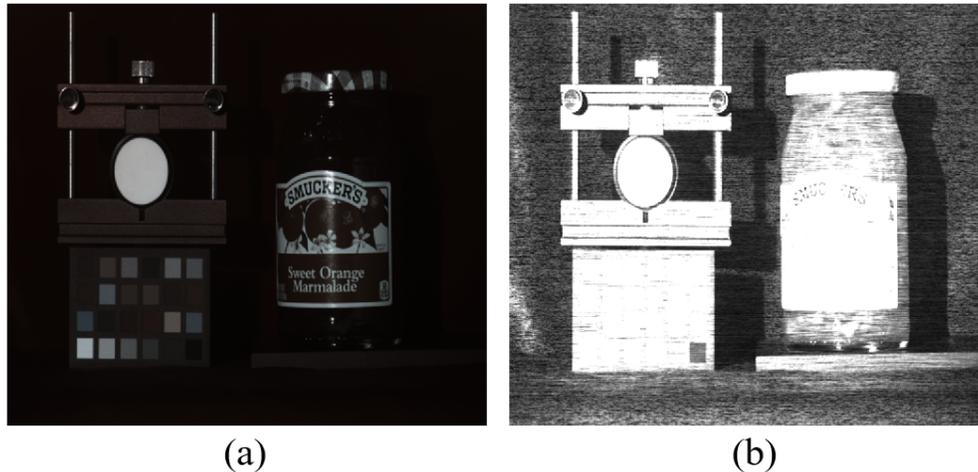


Figura. 11. Medidas del Sistema óptico DD-CASSI, datos adquiridos a partir de dataset KAIST Hwang et al. (2015), las cuales son ingresadas al sistema DD-CASSI como imágenes de 512×512 .

La imagen detectada se expone en la figura 11, donde a partir de la configuración del sistema óptico anteriormente expuesto, figura 9, en a) Se observa una escena con un fondo negro y capturada en el espectro y en b) Se aprecia la captura con una transmitancia del 50% , donde se observa la difuminación existente debido al movimiento que se explico anteriormente en la figura 10.

2.5.3. CACTI. En el dominio temporal se han desarrollado diversas técnicas para la adquisición comprimida de la información, una de estas técnicas se le conoce como la imagen temporal codificada de apertura compresiva CACTI. En este sistema se introduce un esquema de traslación de apertura codificada para la adquisición compresiva de baja potencia de información espacio-temporal, en la que esta codificación se implementa por medio de un código de apertura

ubicado en un plano de imagen intermedio.

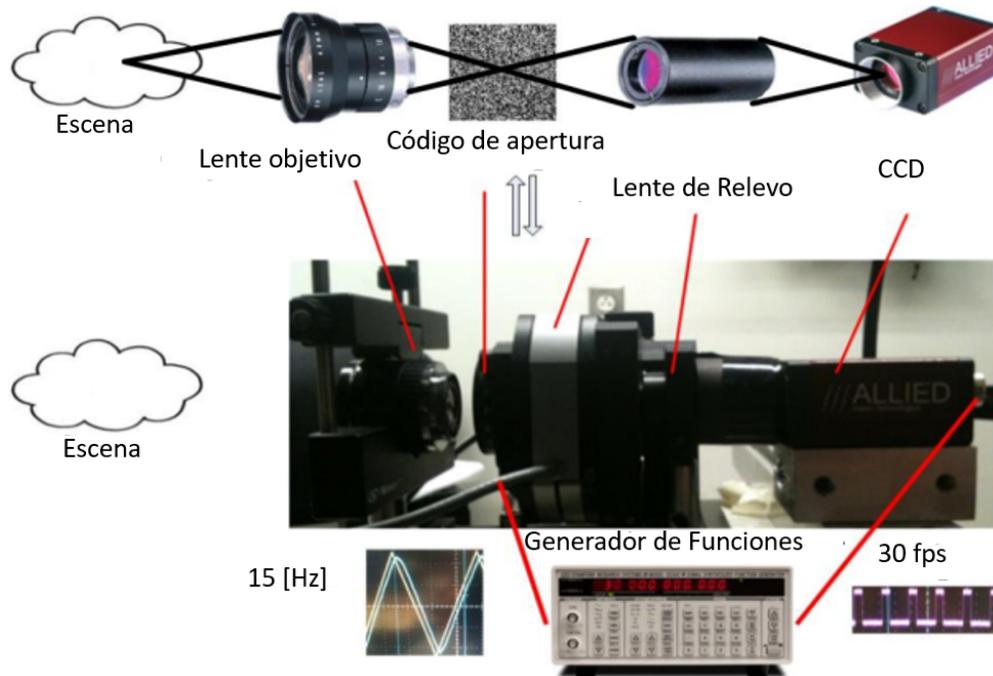


Figura. 12. Sistema óptico CACTI adaptado de Llull et al. (2013)

A diferencia del sistema óptico CASSI en donde cada plano del cubo de datos espectral se encuentra modulado por un código de apertura desplazado, el sistema CACTI realiza una traslación del código de apertura durante la exposición, lo que quiere decir que cada plano temporal es modulado por una versión desplazada del código de apertura, llegando así a una modulación para cada píxel sin utilizar el ancho de banda adicional del sensor, tal y como se observa en la figura 12.

Este sistema de adquisición se puede definir como un patrón único de vóxeles de objetos espacio-temporales $f(x, y, t) \in \mathbb{R}^3$ con una función de transmisión que se desplaza en el tiempo. Si se considera solamente una dimensión espacial en el modelo, mapeando $((x, y) \rightarrow x)$ la medida capturada $g(x', t')$ consiste en muestras discretas de la transformación continua

$$g(x', t') = \int_1^{N_F} \int_1^N f(x, t) T(x - s(t)) \text{rect} \left(\frac{x - x'}{\Delta_x} \right) \text{rect} \left(\frac{t - t'}{\Delta_t} \right) dx dt, \quad (10)$$

donde $T(x - s(t))$ la función de transmisión del código de apertura en la que $s(t)$ describe la posición espacial de este código durante la ventana de integración de la cámara. Además, Δ_x es el tamaño de los píxeles del detector, $\text{rect} \left(\frac{x}{\Delta_x} \right)$ es la función de muestreo de píxeles y Δ_t es el tiempo de integración.

Si se considera un área cuadrada de detección de N píxeles, la forma discreta del cubo espacio temporal es $\mathbf{f} \in \mathbb{R}^{\sqrt{N} \times \sqrt{N} \times N_F}$. Además, teniendo en cuenta que el código de apertura variante en el tiempo $\mathbf{T} \in \mathbb{R}^{\sqrt{N} \times \sqrt{N} \times N_F}$ codifica de forma única cada uno de los canales temporales N_F de \mathbf{f} antes de integrarlos en la medida del detector $\mathbf{g} \in \mathbb{R}^{\sqrt{N} \times \sqrt{N}}$ durante Δ_t , las medidas en los índices espaciales (i, j) y un índice temporal k se pueden representar en su forma discreta como

$$g_{i,j} = \sum_{k=1}^{N_F} T_{i,j,k} f_{i,j,k} + n_{i,j}, \quad (11)$$

donde $n_{i,j}$ representa el ruido de la imagen en el (i, j) ésimo píxel. Por último, tomando la forma vectorial de la fuente $\mathbf{f} \in \mathbb{R}^{N N_F \times 1}$, la medida $\mathbf{g} \in \mathbb{R}^{N \times 1}$ y el ruido $\mathbf{n} \in \mathbb{R}^{N \times 1}$ se puede representar la transformación lineal del sistema como

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \quad (12)$$

donde $\mathbf{H} \in \mathbb{R}^{N \times N N_F}$ es la matriz discreta del sistema que representa la función de muestreo de píxe-

les y la función de transmisión variable en el tiempo, representada en la figura 13. Constatando que el funcionamiento del sistema óptico anteriormente mencionado se refleja en la matriz de sensado, debido a que entre fotograma y fotograma existe un cambio en los valores de las diagonales de esta, todo esto debido a la etapa del piezoeléctrico.

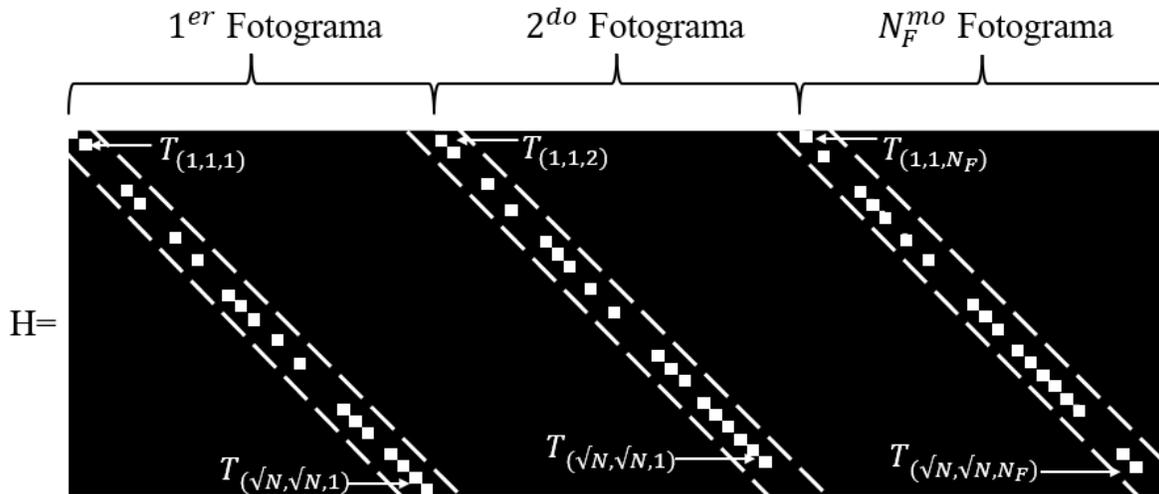


Figura. 13. Matriz de sensado del sistema óptico CACTI, adaptado a partir de Llull et al. (2013).

En la figura 14, se observa en (a) un fotograma de la escena real a capturar, para a continuación en (b) apreciar el mismo fotograma del vídeo capturado a 30 fps con el sistema óptico anteriormente expuesto para obtener resultados experimentales de superresolución temporal, tomado de Llull et al. (2013). Donde se aprecia que se obtiene un aislamiento de los canales temporales de la escena, llegando a una reconstrucción de diversos fotogramas de vídeo de alta velocidad a partir de una única medición, mostrada en (b).

Debido a lo anteriormente presentado, los sistemas tradicionales de adquisición poseen diferencias marcadas, específicamente en el sistema SD-CASSI realiza una codificación fija a cada



Figura. 14. Medidas del Sistema óptico CACTI tomado de and Simulation (2020)

firma espectral y una modulación por medio de un prisma, obteniendo de esta manera codificación únicamente espacial de la escena. Por otro lado, en el sistema CACTI realiza una traslación del código de apertura durante la exposición para de esta manera adquirir cada plano temporal desplazado por el código de apertura, efectuando una codificación temporal de la escena. Adicionalmente, en el sistema DD-CASSI por medio de una dispersión inicial a través de un prisma, una codificación y por último nuevamente una dispersión se genera una modulación diagonal creando una codificación tanto espectral como espacial.

Por esta razón, es remarcable la importancia de desarrollar un método de adquisición imágenes CSV que permita realizar la compresión de la dimensión espectral y temporal en una sola medición y de manera óptima. Este trabajo propone un sistema óptico novedoso compuesto por un filtro tuneable y un DMD bajo un enfoque de codificación de ventana espacio-temporal que permitirá modular ambas dimensiones y adquirirlas mediante un único sensor.

2.6. Algoritmos de filtrado

2.6.1. Filtro recursivo. El filtro recursivo (RF, de sus siglas en inglés *recursive filter*) es un algoritmo de eliminación de ruido utilizado ampliamente en el tratamiento de imágenes. Este filtro aplica convoluciones en la señal utilizando tanto los datos de entrada como los valores previamente calculados de la salida, lo que le da su carácter recursivo y le permite una respuesta eficaz al evitar una convolución más larga. Esta característica hace que el RF se caracterice como un filtro de respuesta al impulso infinita (IIR, de sus siglas en inglés *infinite impulse response*), el cual se puede definir en un n -ésimo punto de la señal como

$$y[n] = b_0x[n] + b_1x[n-1] + \dots + b_Mx[n-M] - a_1y[n-1] - \dots - a_Ny[n-N], \quad (13)$$

donde a y b se conocen como coeficientes de filtro. Específicamente, los valores de b se llaman coeficientes de prealimentación y los valores a se llaman coeficientes de realimentación. De esta manera, cuando los coeficientes de realimentación son no nulos, el filtro se considera recursivo.

2.6.2. Coincidencia de bloques y filtrado 3D. El algoritmo de coincidencia de bloques y filtrado 3D (BM3D, de sus siglas en inglés *block-matching and 3D filtering*) es un tipo de filtrado utilizado para la reducción de ruido en imágenes Dabov et al. (2006). Este método se basa en que una imagen tiene una representación localmente escasa en un dominio de transformación, para esto, agrupa parches de dos dimensiones de la imagen en función de su similitud para obtener un bloque 3D que es posteriormente transformado linealmente, luego, los coeficientes del bloque son reducidos aprovechando la redundancia entre los parches, finalmente, se aplica una

transformación inversa y los parches son reagrupados en sus posiciones originales. Esto permite reducir el ruido presente en la imagen y aumentar los detalles de la misma.

2.6.3. Filtrado de medios no locales. El filtrado de medios no locales (NLM, de sus siglas en inglés *non-local means*) es un método ampliamente usado en el procesamiento de imágenes para la eliminación de ruido Buades et al. (2011). Este algoritmo ajusta cada valor de píxel con un promedio de todos los píxeles de la imagen, ponderada por que tan similares son al píxel objetivo. Esto resulta en una imagen más clara, con bordes reducidos, con una menor pérdida de su estructura y detalles.

2.7. Redes Neuronales

En la actualidad ha tenido un gran auge los sistemas conexionistas, mejor conocidos como redes neuronales artificiales, estos sistemas son un modelo computacional que consiste en un grupo de neuronas las cuales por medio de interconexiones transmiten datos y producen algunos valores de salida, tal y como se observa en la figura 15. Cada neurona se encuentra enlazada con otra por medio de unos enlaces que multiplican el valor de salida por un peso específico y tienen como funcionalidad aumentar o anular el estado de activación de las siguientes neuronas. Estos modelos han tenido gran popularidad en la última década debido a que aprenden y se forman por ellos mismos, sin tener que programarlos de forma explícita, llegando a superar los resultados obtenidos por medio del enfoque tradicional.

2.7.1. Redes Neuronales Convolucionales. Además del modelo de red neuronal convencional, se ha observado un tipo de arquitectura específico de red neuronal artificial que ofrece resultados visiblemente mejores, donde tiene su principal característica es que consiste en

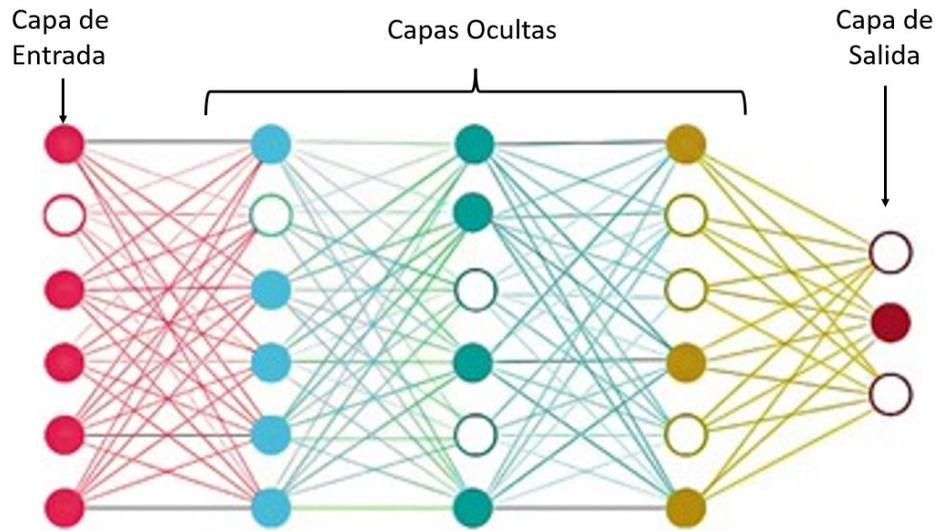


Figura. 15. Red neuronal convencional conformada por una capa de entrada, una capa oculta y una capa de salida.

múltiples capas de filtros convolucionales de una o más dimensiones, y después de cada capa, normalmente se añade una función para realizar un mapeo causal no lineal. Generalmente, este tipo de redes tienen como aplicación la clasificación y resolución de problemas de optimización, donde para esto, se tiene una fase de extracción de características, las cuales la realizan un grupo de neuronas convolucionales y posteriormente una reducción por medio de un muestreo.

Debido a la arquitectura de este tipo de redes neuronales, son principalmente utilizadas para aprender a clasificar todo tipo de datos como imágenes, señales, vídeos, entre otros, sin embargo, dieron origen a otro tipo de redes neuronales utilizadas para segmentación de imagen.

2.7.2. U-Net. Uno de los primeros algoritmos de segmentación semántica que utiliza redes convolucionales completas es la arquitectura conocida como U-NET Ronneberger et al. (2015). La cual se construye en una estructura simétrica en forma de U en el que están las direcciones de compresión y expansión.

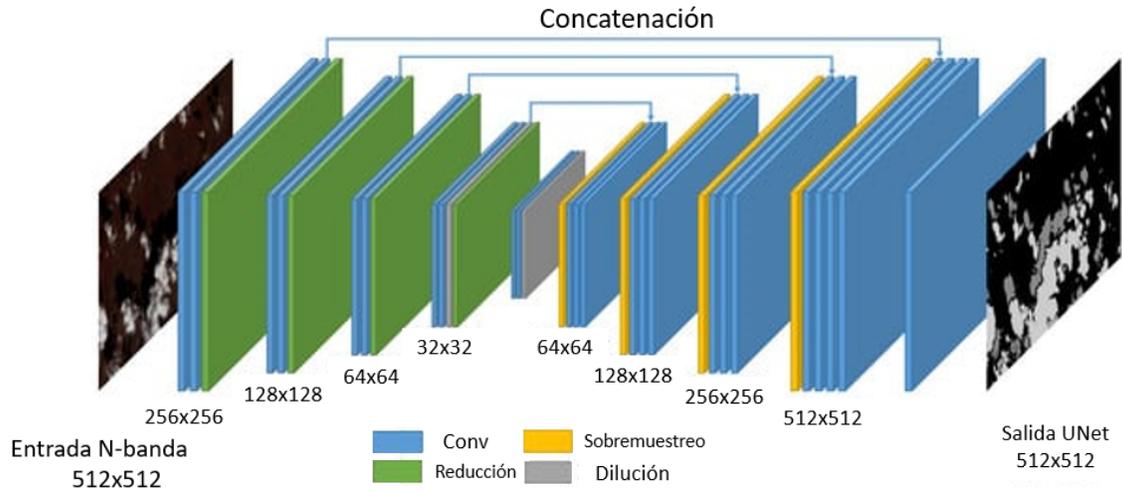


Figura. 16. Arquitectura de la red convolucional U-NET para la segmentación de imágenes, adaptado de Jiao et al. (2020).

El funcionamiento de esta red se subdivide en tres fases principales, la fase de compresión, la fase de puente y la fase de expansión. La primera fase convierte la imagen de entrada en una representación compacta por medio de un aprendizaje jerárquico, aplicando bloques de convolución, agrupación máxima y submuestreo. De esta manera, comprime la información capturando el contexto de la imagen de entrada para poder realizar la segmentación y extrayendo múltiples representaciones de características de esta. Esta fase consta de 4 bloques de contracción donde cada uno de ellos bloque toma la entrada y aplica dos capas de convolución de 3x3 seguidas de una capa de agrupación máxima de 2x2. La segunda fase se encuentra en la capa más inferior y tiene como papel principal realizar el enlace entre las rutas de compresión y expansión siendo la ruta media ellas, esta fase utiliza dos capas de convolución de 3x3. La última fase discrimina en cada píxel utilizando las características aprendidas y aplicando operadores de muestreo, concatenación y convolución. Esta fase consta de 4 bloques de expansión, en esta, cada bloque pasa la entrada a

dos capas convolucionales de 3x3 seguidas de una capa de muestreo de 2x2. El propósito de la fase de expansión es permitir una localización precisa de la información contextual de la fase de compresión, lo que asegurara que las características que se aprenden al contraer la imagen se utilizan para reconstruirla. Finalmente, el mapeo resultante pasa a través de otra capa de convolución de 3x3 con el mismo número de mapas de características que el número de capas de imagen deseadas. El proceso descrito anteriormente se observa en la figura 16 donde en la primera fase se comprime la información, y posteriormente se reconstruye en la fase final. Por lo anteriormente expuesto esta arquitectura de red neuronal ha sido utilizada en múltiples tareas de procesamiento de imágenes como extracción de caminos Zhang et al. (2018b) usando imágenes de teledetección, detección de grietas en hormigón Liu et al. (2019b) e imágenes falsificadas Bi et al. (2019), seguimiento de objetivos en óptica adaptativa predictiva Chen et al. (2021), reconstrucción de campos de partículas 3D usando hologramas bidimensionales Wu et al. (2021) y recuperación de datos sísmicos escasamente muestreados Huang and Nowack (2020), entre otros.

2.7.3. ResNet. Las redes residuales o ResNet, son un tipo de redes neuronales que introducen saltos de conexión en sus arquitecturas. Estas redes surgen teniendo en cuenta los resultados obtenidos por las redes tradicionales en los que se observó que su precisión y velocidad de convergencia se veían afectadas a medida que el número de capas aumentaba. Esto es contrario a lo que se esperaba teóricamente, sin embargo ocurre dado que el error de entrenamiento aumenta, debido al sobre ajuste. Por lo tanto, el problema se convierte a uno de optimización o degradación del gradiente.



Figura. 17. Arquitectura de la red convolucional residual ResNet para la estabilización del gradiente, adaptado de Talo (2019).

Por dicha razón se implementó una nueva arquitectura conocida como red residual (ResNet, de sus siglas en inglés *residual network*) He et al. (2015), la cual, como se observa en la figura 17, se basa en una estructura de saltos entre dos o más capas que contienen una normalización y en la que no existe una linealidad entre ellas, para así disminuir la retropropagación del gradiente a las capas anteriores y evitar el desvanecimiento del mismo.

El funcionamiento de esta arquitectura se basa en un bloque conocido como bloque residual, este bloque tiene una ruta en la que su entrada se agrega directamente a su salida, conocida como conexión de salto. De esta manera, en el peor de los casos las capas más profundas no aprenden nada y el rendimiento y precisión no son afectados por el gran número de capas de la red, debido a la estabilización que se crea por las conexiones de salto. Además, es necesario resaltar que, en la mayoría de las pruebas realizadas las capas más profundas de esta arquitectura aprenden en menor medida, aumentando el rendimiento de la red, pero estableciendo el gradiente y disminuyendo el problema de desvanecimiento.

2.7.4. FFDNet. La red neuronal convolucional de eliminación de ruido rápida y flexible (FFDNet, de sus siglas en inglés *fast and flexible denoising convolutional neural network*)

Zhang et al. (2018a) fue propuesta como un método efectivo y eficiente para la eliminación del ruido en procesamiento de imágenes a diferentes niveles incluso cuando este es espacialmente variable. Este algoritmo de red neuronal, a diferencia de otros eliminadores de ruido del estado del arte Dabov et al. (2007); Zhang et al. (2017), agrega un mapa de nivel de ruido sintonizable que es modelado como entrada de la red y hace que los parámetros del modelo sean flexibles a diferentes niveles de ruido. Además, el modelo trabaja con subimágenes reducidas espacialmente, logrando una compensación entre la velocidad del entrenamiento y prueba y el rendimiento de la eliminación de ruido. Para esto, se utiliza un operador de reducción espacial reversible que ajusta una imagen de entrada de tamaño $N_x \times N_y \times N_\lambda$ en cuatro subimágenes submuestreadas de tamaño $\frac{N_x}{2} \times \frac{N_y}{2} \times 4N_\lambda$ para que el eliminador de ruido sea eficiente sin sacrificar el rendimiento y logre un equilibrio entre la reducción de ruido y la preservación de los detalles.

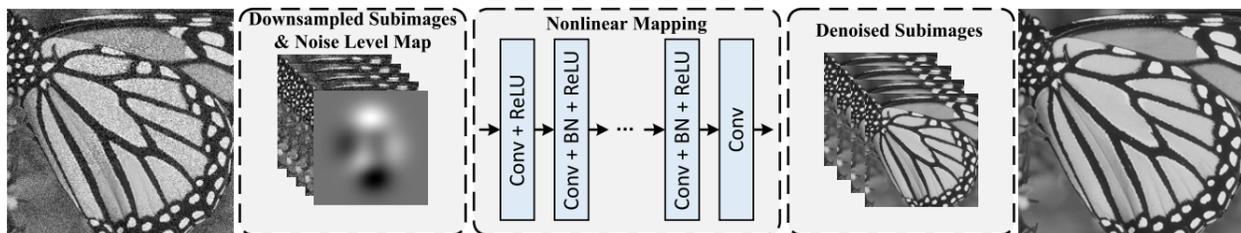


Figura. 18. Arquitectura de la red neuronal FFDNet, adaptado de Zhang et al. (2018a)

La figura 18 muestra la arquitectura de la FFDNet, la primera capa transforma la imagen ruidosa de entrada en las cuatro subimágenes mediante el operador de reducción espacial. Posteriormente, se concatena el nivel de ruido a estos datos para formar un tensor de tamaño $\frac{N_x}{2} \times \frac{N_y}{2} \times (4N_\lambda + 1)$. La red consta de 15 capas convolucionales con un tamaño de filtro 3×3 compuestas por operaciones convolucionales, capas de rectificación lineal (ReLU) y normalización

por lotes, excepto por la primera capa que no incluye la normalización por lotes y la última capa que solo realiza la convolución. Finalmente, luego de la capa de convolución final, se aplica un operador inverso del operador de reducción de resolución aplicado en la etapa de entrada, con el objetivo de aumentar la escala de la imagen y así obtener la imagen limpia estimada de tamaño $N_x \times N_y \times N_\lambda$, reconstruyéndola con las cuatro subimágenes sin ruido. Es importante resaltar que la red neuronal establece 64 mapas de características en sus capas y además realiza un relleno de ceros (zero-padding de su nombre en inglés) para mantener el tamaño de estos mapas de características después de cada convolución.

2.8. Algoritmo de Variación Total

El algoritmo de variación total (TV, de sus siglas en inglés *total variation*) Chambolle et al. (2010) es un filtro de eliminación de ruido que se basa en la idea que las señales con detalles espurios excesivos tienen una variación total alta. Esta premisa principal puede desarrollarse en que al minimizar o reducir la variación total de la señal, se eliminan los detalles no deseados, de esta manera se convierte en un filtro de conservación de bordes notablemente eficaz, conservando los bordes y suavizando el ruido en regiones planas Strong and Chan (2003).

3. Sincronización de Elementos Optoelectrónicos

Para implementar el sistema optoelectrónico propuesto en el segundo objetivo específico para la adquisición de SV por medio de un protocolo de muestreo tipo barrido espectral y fotograma a fotograma, se desarrollaron un grupo de actividades para la caracterización y sincronización de dichos dispositivos y posteriormente cumplir a cabalidad con la captura de los vídeos espectrales en laboratorio.

3.1. Caracterización

Primero que todo, para un correcto funcionamiento e implementación del sistema optoelectrónico para la adquisición de SV, se realizó una caracterización de los dispositivos que se van a utilizar y de esta manera obtener vídeos óptimos, disminuyendo la probabilidad de errores debido a una mala calibración o conexión que podría existir entre estos.

3.1.1. Caracterización de Cámara Stingray F-080B. La cámara Stingray F-080B, cuenta con diversas utilidades, entre las cuales se encuentra la conexión por medio de USB o un trigger externo, asimismo, su funcionamiento es en modalidad de 16 bits, y cuando se encuentra a máxima resolución la velocidad de captura tiene una captura de 16 imágenes por segundo. Este tipo de cámara es útil en aplicaciones de automatización industrial y de laboratorio mediante visión artificial. Igualmente, la integración de diversas funciones de preproceso en la cámara permiten obtener una excelente calidad de imagen, por tanto, se disminuye la necesidad de efectuar procesos por parte de la CPU o GPU aumentando considerablemente la velocidad del sistema, siendo de esta manera un dispositivo óptimo para la captura automatizada de SV por el método de adquisición de

fotograma por fotograma.



Figura. 19. Cámara Stingray F-080B.

Este dispositivo optoelectrónico posee múltiples características específicas, tales como:

- Velocidad de obturación: 49 [μ s] a 67 [s].
- ADC: 14 bit.
- Tamaño máximo de imágenes: 1032 x 776 [píxel].
- Tasa de fotogramas: 60 [fps].
- Requerimientos de Potencia: 8 a 36 VDC por medio de cable IEEE o 12-pin Hirose.
- Numero F= f/1.4.
- Entrada de Cámara 1: Voltaje (Alto)= 3-24 [V] Voltaje (Bajo)= 0-1.5[V]
- Entrada de Cámara 2: Voltaje (Alto)= 3-24 [V] Voltaje (Bajo)= 0-1.5[V]

Adicionalmente, una de las características más relevantes que se tuvieron en cuenta para la elección del dispositivo electrónico encargado de capturar la escena en múltiples longitudes de onda, es su respuesta al espectro, en cuyo caso tal como se observa en la figura 20, cuenta con una sensibilidad de amplio espectro en diferentes bandas como RGB, asimismo posee un extenso espectro con respecto a la respuesta monocromática, infiriendo de esta manera que para capturar un SV en un barrido espectral desde los 400 a 700 nm es idónea.

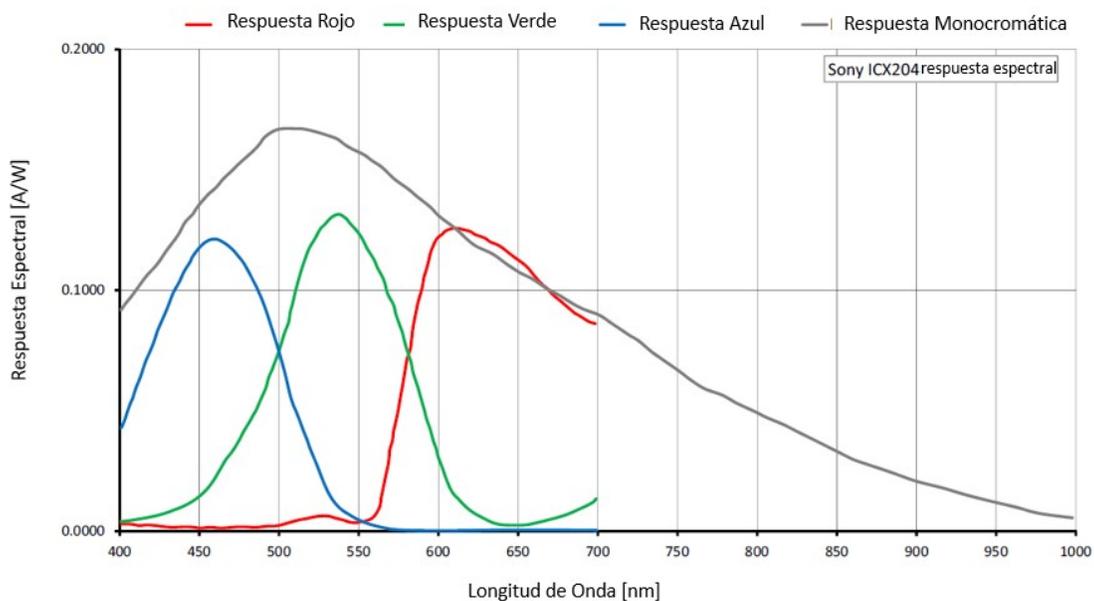


Figura. 20. Respuesta Espectral Cámara F-080B, donde se observa que cada curva es la respuesta espectral de la cámara respectiva a su color, y la curva gris es la respuesta espectral de la cámara a una iluminación monocromática.

Por otro lado, en el ámbito de conectividad cuenta con un trigger externo y un par de puertos USB los cuales son útiles para la sincronización con diversos elementos, figura 21, tanto ópticos como electrónicos, reflejando de esta manera su idoneidad para la captura de SV por medio de la sincronización de múltiples dispositivos.

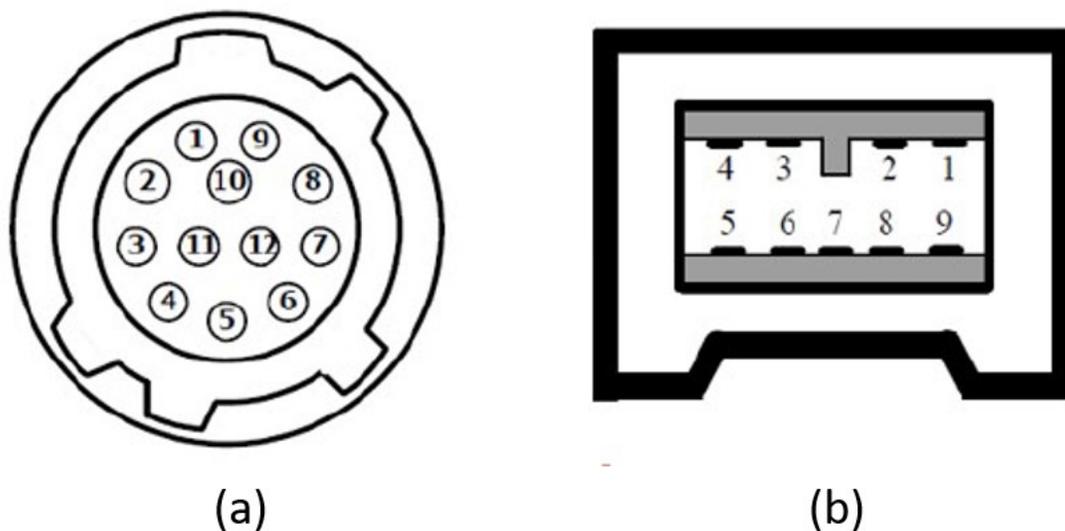


Figura. 21. Conectividad Cámara F-080B, en (a) se observa la conexión externa tipo trigger pin-hirose 12 pines, la cual comúnmente se utiliza para conexión de un generador de señales. En (b) aparece la conexión IEEE 1394 b normalmente utilizada para sincronización industrial .

Por lo anterior expuesto, se observa que la cámara Stingray F-080B es un dispositivo óptimo para realizar la captura de SV de forma automatizada y sincronizada dispositivos optoelectrónicos, dado sus características, más específicamente la particularidad de las múltiples opciones integradas para un control por medio de una conexión externa como el puerto USB o trigger.

3.1.2. Caracterización de Monocromador TLS-300XR. El monocromador TLS-300XR es un dispositivo capaz de dispersar la radiación que refleja una escena, la cual es transmitida en un grupo de bandas específicas de longitudes de onda, asimismo, la luz monocromática se encuentra en el rango de 350-1800 nm con una resolución de 0.7 nm.



Figura. 22. Monocromador TLS-300XR que se encuentra en el laboratorio de óptica del grupo HDSP.

Las principales aplicaciones en las que se utiliza este dispositivo es principalmente la dispersión de longitudes de ondas específicas sobre una escena, para que de esta manera el sensor reciba de manera secuencial una imagen en una longitud de onda particular, la cual finalmente por medio de un procesamiento de datos se visualice una imagen espectral condensada en un rango de bandas determinado con alta precisión, y por consiguiente realizar un análisis espectral de la escena, obteniendo información valiosa que por medio de una captura en el espectro visible es desestimada.

Un aspecto importante en la caracterización del dispositivo es la respuesta al espectro, donde se asegura el ancho de banda efectivo del mismo, y cual es el rango de longitudes de onda en

los cuales se podrá adquirir una imagen espectral. Para comprobar dichas características se realizó un sistema monocromador-espectrómetro realizando un barrido desde 450 a 700 nm con separaciones de 25nm, obteniendo los resultados que se observan en la figura 23, concluyendo de esta manera que el dispositivo posee una respuesta espectral óptima para la captura de SV, asimismo asegurando que no exista una superposición de bandas espectrales, todo esto debido a la configuración del mismo por medio del banco de filtros y el slit, disponiendo de un ancho de banda por slit de aproximadamente 12 nm, tal y como se observa en la figura 23. Adicionalmente se tiene un máximo de intensidad en 545 nm con un valor de 44518 lm, y por otra parte un valor mínimo en la banda espectral de 700 nm, obteniendo un valor de 13318 lm, infiriendo de esta forma que en los rangos espectrales centrales posee mayor intensidad de iluminación y a medida que se extiende a los extremos de infrarrojo y ultravioleta dicha intensidad disminuye.

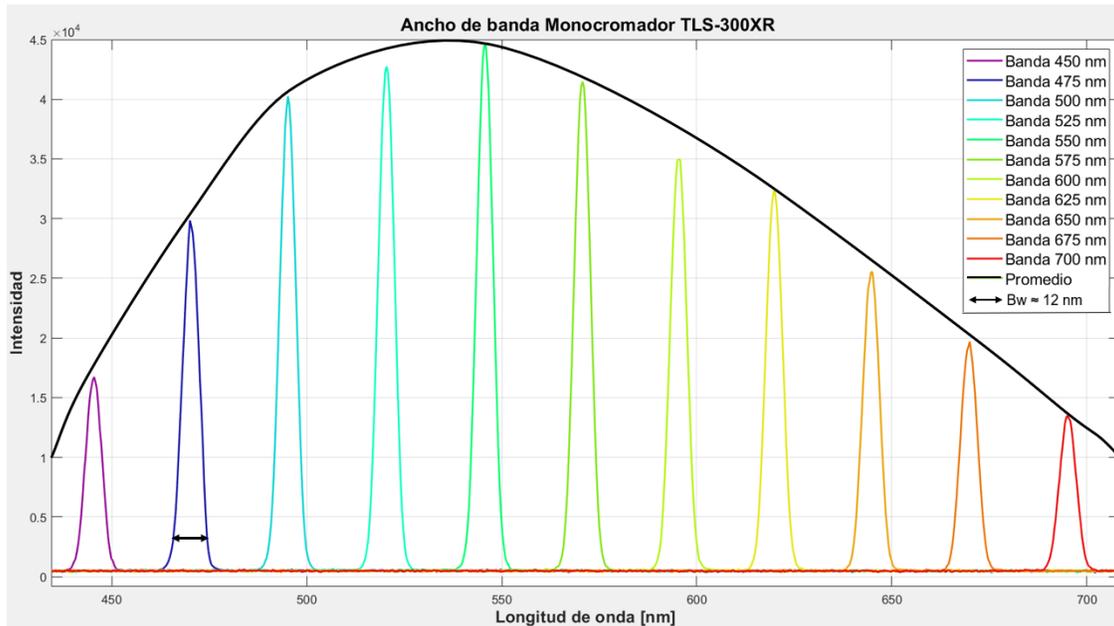


Figura. 23. Resultados de caracterización espectral del monocromador obtenido a través de una conexión entre el monocromador TLS-300XR y el espectrómetro de onda de referencia Flame para espectro visible (VIS)

Asimismo, en lo que concierne a este documento, es fundamental investigar la configuración necesaria para realizar una sincronización de este dispositivo con otros, y en este caso posee tres opciones para sincronización, una de ellas es por medio de una conexión de trigger, la otra es por medio de una comunicación serial por un puerto RS232 y por último, la más utilizada es una conexión USB a un equipo de cómputo.

En el primero de estos, es necesario resaltar que la conexión de Trigger externo nos permite controlar y sincronizar el sistema con el obturador del dispositivo, por medio de una entrada de tipo BNC tal y como se observa en la Figura 24 (F), asimismo el obturador posee otros modos de funcionamiento, tales como el modo manual y temporizado, los cuales se pueden adaptar a la sincronización vía conexión Trigger externo.

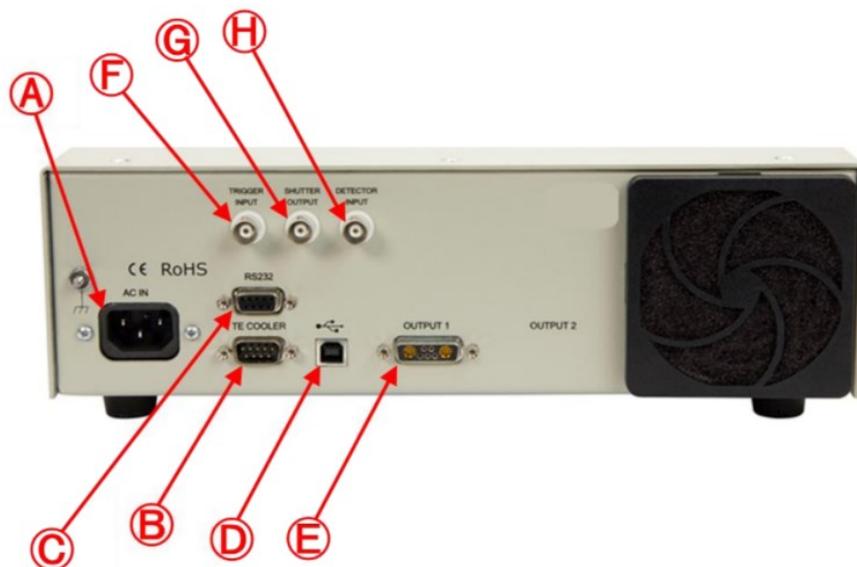


Figura. 24. Conexiones que posee Fuente de alimentación del Monocromador, tanto para sincronización como para transmisión de información

Por otro lado, para controlar las bandas que se desean capturar a través de un dispositivo externo, el monocromador posee la opción de sincronizarlo por medio de una conexión RS232 tal y como se observa en la Figura 24 (C), y en conjunto con un microcontrolador como el Arduino Uno o Raspberry Pi, se realiza una comunicación serial de este dispositivo y de esta manera controlar las bandas que se desean capturar de la escena.

Por último, la manera mas común de conexión es por medio del puerto USB, el cual normalmente se conecta a un equipo de computo, y por medio del software Matlab u otro, se programa el funcionamiento deseado, tanto para el estado de obturación, como para los tipos de bandas con los que se desea iluminar la escena. Por lo anteriormente expuesto, los métodos de conexión óptimos para la aplicación que concierne a este documento es a través de la conexión tanto USB o RS232, la cual será definida dependiendo del tipo de dispositivo que se utilizara en el sistema, ya sea un

microcontrolador o un equipo de cómputo.

3.2. Sincronización de Elementos Optoelectrónicos

Prosiguiendo con la captura de SV en el laboratorio de óptica por medio de un protocolo de muestreo tipo barrido espectral y fotograma a fotograma, es necesario realizar una sincronización de diversos dispositivos ópticos los cuales fueron mencionados anteriormente, en consecuencia se llevó a cabo una sincronización progresiva de estos elementos para culminar con una captura automatizada y óptima de SV.

3.2.1. Sincronización Cámara Stingray F-080B y Equipo de Computo.. El primer paso para la implementación de un sistema optoelectrónico totalmente sincronizado para la captura de SV es sincronizar el computador con la cámara Stingray F-080B para la adquisición de los fotogramas que conformaran el vídeo espectral.



Figura. 25. Montaje de sincronización Computador, y Cámara F-080B, mediante conexion USB

El funcionamiento de la sincronización de estos dos dispositivos se realiza por medio del puerto USB que posee la cámara, tal y como se observa en la figura 25, donde por medio de una programación realizada en el laboratorio en el software Matlab, Anexo 1, se acoplan y se adquiere la escena, tal y como se observa en la Figura 26.



Figura. 26. Captura de imagen en escala de grises por medio de sincronización Computador y Cámara F-080B

Se realiza una captura en escala de grises por medio de la cámara F-080B sin utilizar una iluminación específica proporcionada por un monocromador, tal y como se observa en la Figura 26, concluyendo de esta manera que debido a un óptimo posicionamiento de la cámara con respecto a la escena, se establece un enfoque adecuado, teniendo en cuenta para esto los conocimientos de óptica en el trazado de rayos y las características de la cámara como la distancia focal y las

aberraciones ópticas dado que se tiene un lente objetivo de 8 mm y $f/1.4$.

3.2.2. Sincronización de Equipo de Computo, Microcontrolador Arduino y Servo Motor. Para realizar la captura de vídeos espectrales de forma automatizada, se implementó un montaje y sincronización de diversos elementos, como la tarjeta de control Arduino y un servomotor de 180°.

3.2.2.1. Servomotor. Un servo motor es un servomecanismo de bucle cerrado que utiliza realimentación de posición que permite un control preciso de posición angular y velocidad, por medio de una señal de control tanto analógica como digital.



Figura. 27. Servomotor

En el mercado existen diversos tipos de motores, pero se clasifican puntualmente en dos:

Servomotor de 180°: Este tipo de servomotor ofrece la opción de controlar la posición donde se desea ubicar el motor con una sensibilidad de 1° con un máximo de rotación de 180°, tal y como se observa en la Figura 28, con una limitación de control de velocidad de rotación. Normalmente este tipo de servomotores se utilizan en aplicaciones que requieren precisión en el ángulo de giro sin importar la velocidad de respuesta.

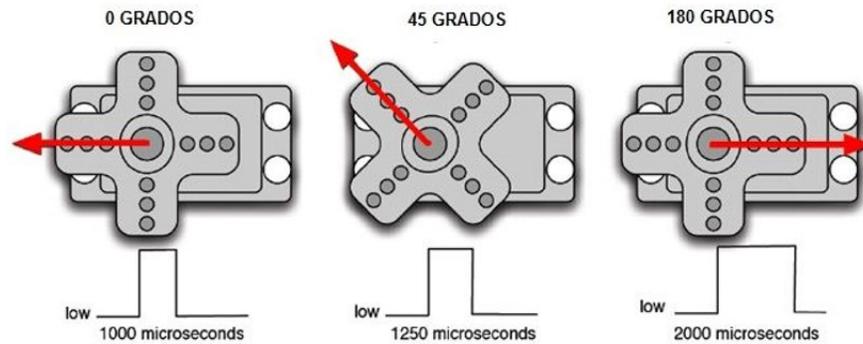


Figura. 28. Funcionamiento de servomotor de tipo 180°, tomado de Area Tecnología (2020)

Tal y como se observa en la Figura 28, dependiendo del ancho del pulso será la posición que tome el servomotor. También es necesario resaltar que todo esto depende de la tarjeta de control que se emplee, la cual normalmente es la Arduino Uno, o la raspberry.

Servomotor de 360°: Por otro lado, este tipo de servomotores poseen la característica de poder realizar una rotación de 360°, esto debido a una modificación realizada a nivel mecánico, pero a cambio de un mayor dinamismo de rotación, se pierde control en el ángulo de rotación, haciendo de esta manera que en lugar de tener en cuenta el ángulo de posición, se controla la velocidad de giro.

Tal y como se observa en la Figura 29, el funcionamiento básico de este tipo de servomotores es a partir del ancho del pulso, tal cual como los servomotores de 180°, pero controlando la velocidad de giro, lo que hace que este tipo de servomotores tengan aplicaciones en donde la precisión del ángulo sea poco importante en comparación con la velocidad necesaria para realizar un giro completo.

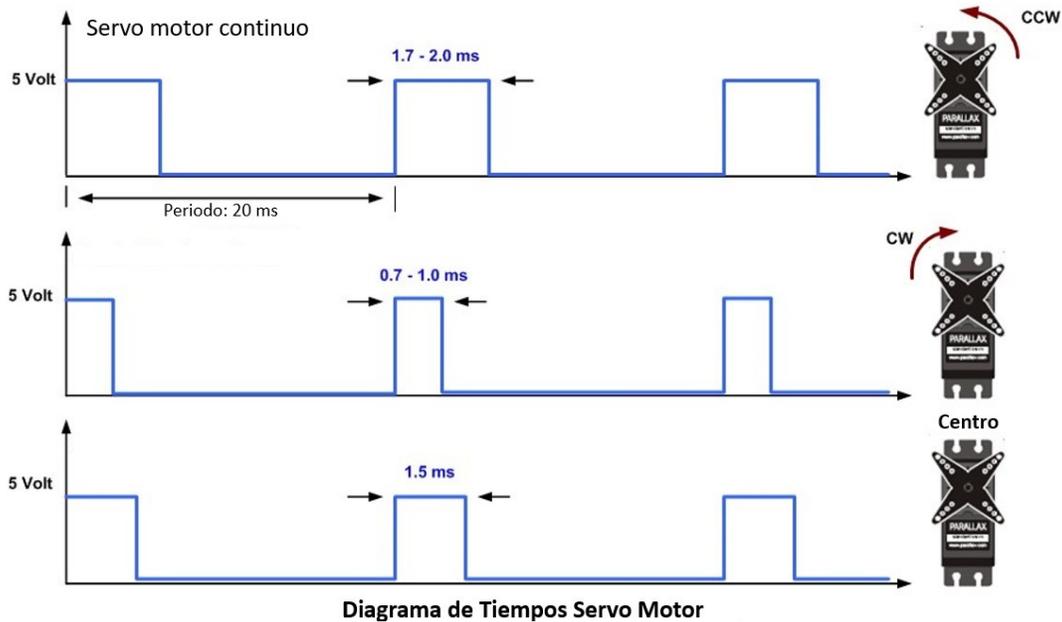


Figura. 29. Funcionamiento de servomotor de tipo 360°, adaptado de Control Automático Educación (2020).

Por lo tanto, para la aplicación de uno de los objetivos específicos de este trabajo de grado, el cual es realizar la captura de manera controlada y específica de un SV automatizado por medio de una escena con dinámica de rotación se utilizó un servomotor de 180°, debido a su precisión en el ámbito de ángulo de rotación y de esta manera asegurar la repetibilidad de la escena.

3.2.2.2. Microcontrolador Arduino. Asimismo, para completar la sincronización del sistema que realiza la dinámica de la escena, se utilizó la placa Arduino uno, la cual está basada en un microcontrolador ATMEL, donde por medio de programación básica en el entorno de Arduino IDE se pueden grabar instrucciones para su funcionamiento en diversos ámbitos.



Figura. 30. Placa Arduino Uno/Mega, utilizada para aplicaciones de adquisición, tratamiento y transmisión de señales digitales y analógicas.

La razón principal del porque se tomó la decisión de utilizar esta tarjeta, es debido a su amplia gama de aplicaciones, y el gran número de puertos que posee, los cuales se pueden utilizar en diferentes implementaciones al mismo tiempo, asimismo, debido a su fácil programación y su velocidad de transmisión de las instrucciones a los dispositivos conectados. Igualmente, de sus características remarcables se encuentran las siguientes:

- Velocidad de Reloj: 16 [MHz].
- Voltaje de trabajo: 5 [V].
- Rango de voltaje de entrada: 7.5-12 [V].
- Posee 14 Pines digitales y 6 pines analógicos.
- Rango de velocidad de comunicación serial: 300-115200 [baudios]

3.2.2.3. Implementación. Debido a los dispositivos seleccionados, se prosigue a realizar el montaje de la sincronización del computador con la placa Arduino y el servomotor, tal y como se observa en a Figura 25.

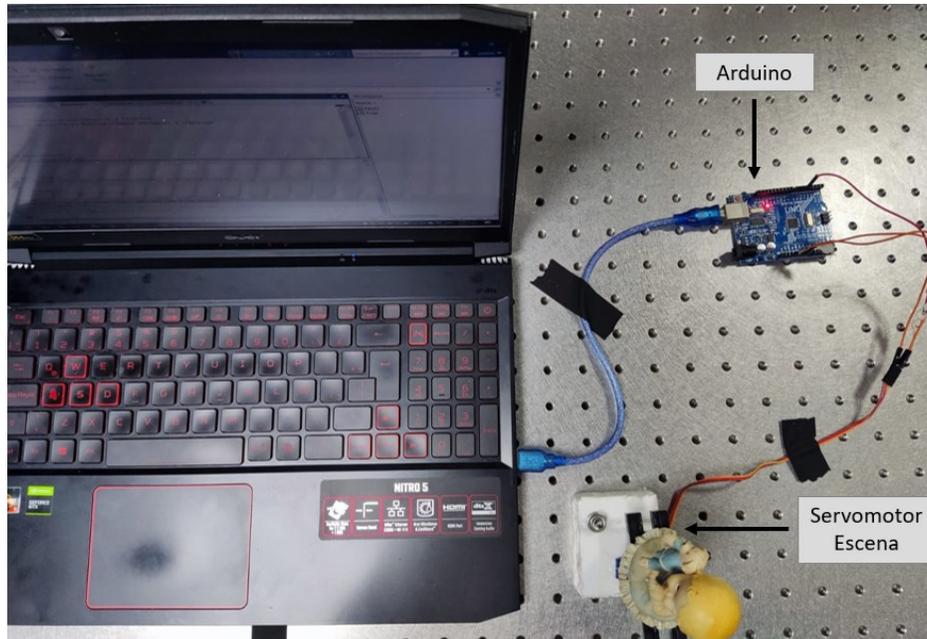


Figura. 31. Montaje de sincronización Computador, Arduino y Servomotor, por medio de conexión serial, realizando un giro de 180° por medio de pasos de 10° .

El funcionamiento de esta sincronización se inicia por medio de una programación realizada en el software Matlab, el cual realiza una rotación específica, donde en este caso es un ángulo de 10° cada 0.5 [s], tal y como se observa en el Anexo 1, llegando al máximo de ángulo en 18 pasos esto debido a la elección del tipo de servomotor de 180° , teniendo también la posibilidad de realizar una programación para que la escena regrese a su posición inicial en el mismo intervalo de tiempo, y de esta manera realizar el movimiento deseado para la elaboración de los fotogramas de un vídeo por medio de la técnica de fotograma a fotograma y asegurándose de que sea un proceso repetible para diferentes escenas. Concluyendo de esta manera que la sincronización realizada para la automatización de la dinámica de la escena se encuentre comprobada y funcional, y en consecuencia continuar con la siguiente fase. Es importante precisar que por medio de una conexión

extra a partir de los mismos pines y un servomotor adicional se obtendría una dinámica en el fondo de la escena mejorando la dinámica de la escena en un grado de libertad.

3.2.3. Sincronización Equipo de Computo, Cámara Stingray F-080B, Microcontrolador Arduino y Servo Motor. El siguiente paso para la captura de SV es la sincronización de los elementos anteriormente mencionados con la cámara stingray F-080B y de esta manera capturar un vídeo en el espectro visible de forma automatizada.

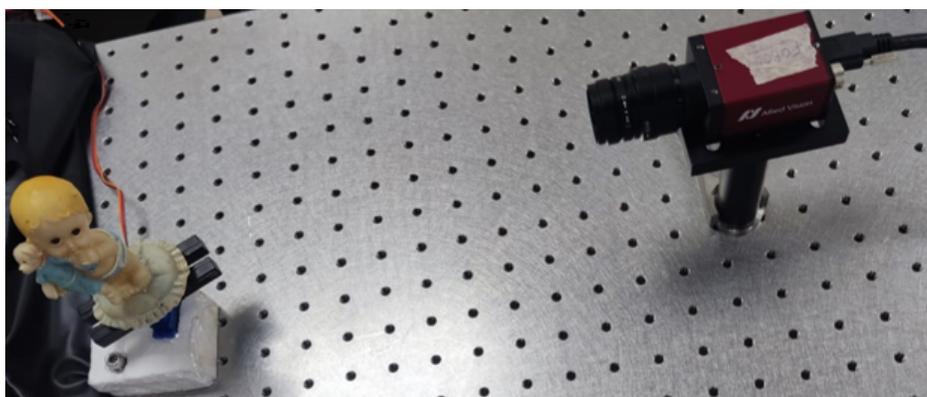


Figura. 32. Montaje de sincronización Computador, Cámara Arduino y Servomotor, por medio de cable USB y conexión serial, respectivamente.

El funcionamiento del sistema optoelectrónico planteado y que se observa en la figura 32, es el siguiente: Por medio de una programación en la cual se combinaron los códigos mencionados en las secciones 3.2.1,3.2.2 se realiza la sincronización, obteniendo de esta manera un sistema en el cual primero se activa el servomotor para realizar la rotación específica de la escena, y después de un retardo predeterminado se activa la cámara por un tiempo de integración específico para capturar un fotogramas, después de esto para cerrar el lazo del sistema nuevamente se activa el servomotor y así proceder a capturar el siguiente fotograma del vídeo, adquiriendo de forma óptima,

automatizada y repetible un vídeo en el espectro visible por medio de la técnica de fotograma a fotograma, tal y como se aprecia en la figura 33.

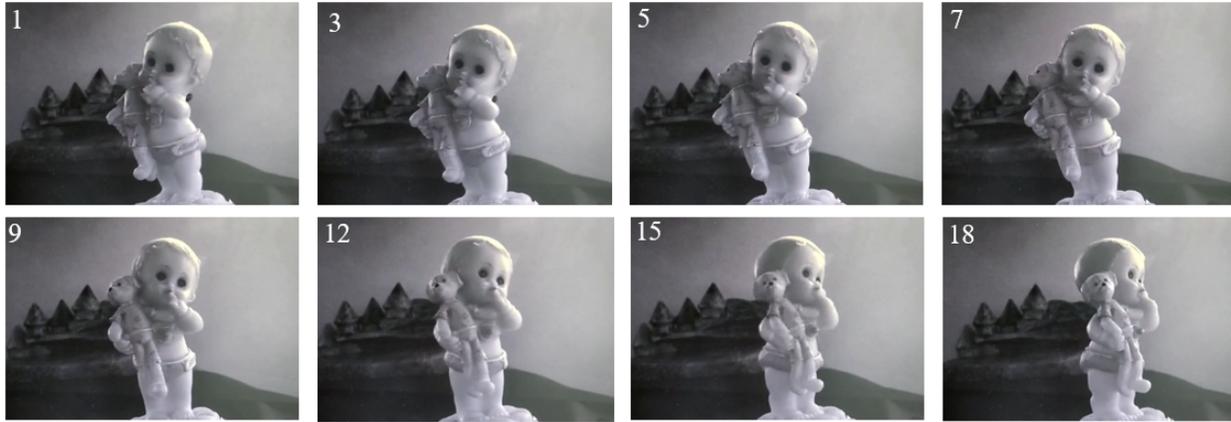


Figura. 33. Resultados de captura de video de 18 fotogramas en escala de grises por medio de sincronización Computador, Arduino y Servomotor.

Adquiriendo una escena en escala de grises de 18 fotogramas, donde se aprecia el movimiento de la escena a través de la sincronización del servomotor el cual tiene un paso de 10° por fotograma, asimismo, se observa que el posicionamiento de la cámara es el óptimo, sin importar si la dinámica de la escena varia dado que tanto la escena como el sensor se encuentran fijados a través de un soporte.

3.2.4. Sincronización Equipo de Computo y Monocromador TLS-300XR. Posteriormente a la sincronización de los elementos para la captura de vídeos en el espectro visible, es necesario sincronizar un dispositivo que se encargue de la dispersión de las bandas para de esta manera generar finalmente vídeos espectrales, en cuyo caso se utilizará el monocromador TLS-300XR

Tal y como se explicó en la caracterización de este dispositivo, la sincronización se puede

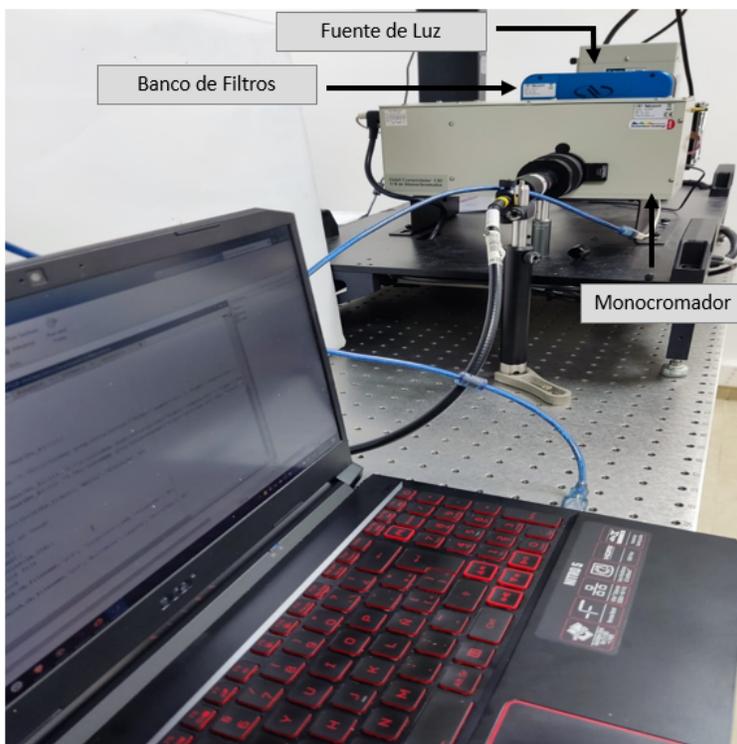


Figura. 34. Montaje de sincronización Computador y Monocromador a través de cable USB.

llevar a cabo por diversos modos, en cuyo caso se utilizó la conexión por puerto USB para realizar una programación a nivel de software, figura 34 y modificar de esta manera la longitud de onda con la que se desea iluminar la escena y descartar la sincronización de la apertura del obturador.

Por dicha razón, el funcionamiento de este sistema es el siguiente. A través de un código creado en el software Matlab se realiza la comunicación con el monocromador para de esta manera iluminar una escena en una longitud de onda específica y adquirir la información de la escena en esa banda espectral, asimismo, por medio de una configuración predeterminada se realiza un barrido en un rango efectivo de 400-700 nm, para de esta forma al final condensar esta información y obtener una imagen espectral.

3.2.5. Sincronización Equipo de Computo, Monocromador, Cámara Stingray

F-145B, Microcontrolador Arduino y Servo Motor. Finalmente, para cumplir con el objetivo de realizar una captura de SV automatizada, se realiza una sincronización general de todos los dispositivos optoelectrónicos anteriormente mencionados, tal y como se observa en la figura 35, donde en dicho montaje se posicionan tanto la cámara como la salida del monocromador a una distancia especifica, para de esta manera asegurar que tanto el enfoque de la escena en la cámara sea el óptimo, como que la iluminación sea la correcta para realizar el barrido de las bandas espectrales deseadas.

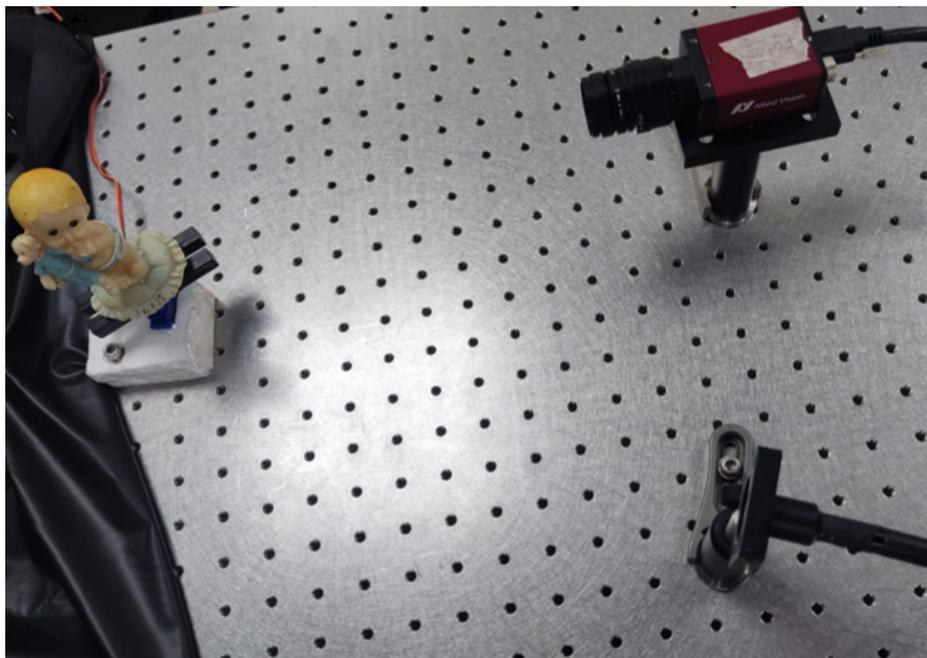


Figura. 35. Montaje de sincronización Computador, Cámara, Monocromador, Arduino y Servomotor para la captura de videos espectrales.

Tal y como se ha llevado el proceso durante la seccion la sincronización de los dispositivos anteriores, en esta etapa final se conectan todos los dispositivos al equipo de cómputo por medio

de sus puertos USB y/o puerto serial, después de esto por medio de la combinación de todos los códigos realizados, se genera un código final, Anexo 1. La iluminación ofrecida por el monocromador fue configurada en pasos de 3 nm, entre un rango espectral de 400-700 nm, igualmente la cámara utilizada posee un lente objetivo de 8 [mm] EFL y f/1.4, con una dinámica entre fotograma y fotograma de 10° y un tiempo de captura de 0.5 [s]. Concluyendo de esta manera con un sistema totalmente automatizado y sincronizado para la captura de vídeos espectrales por medio de la metodología de fotograma a fotograma.

4. Método Propuesto

En esta sección se presentará el método propuesto basado en una metodología de adquisición y reconstrucción de SV compresivo. Para esto, se plantea el sistema óptico mostrado en la figura 36 compuesto principalmente la combinación de un DMD y un filtro tuneable, cuya principal función es generar códigos de apertura de color variantes en el tiempo que permitan comprimir la información espectro-temporal del SV. A través del uso de estos dos elementos ópticos y un enfoque de ventana espacio temporal la dimensión espectral y temporal pueden modularse en un tiempo de integración para lograr capturar las medidas escasas mediante un único sensor conformado por una cámara CCD.

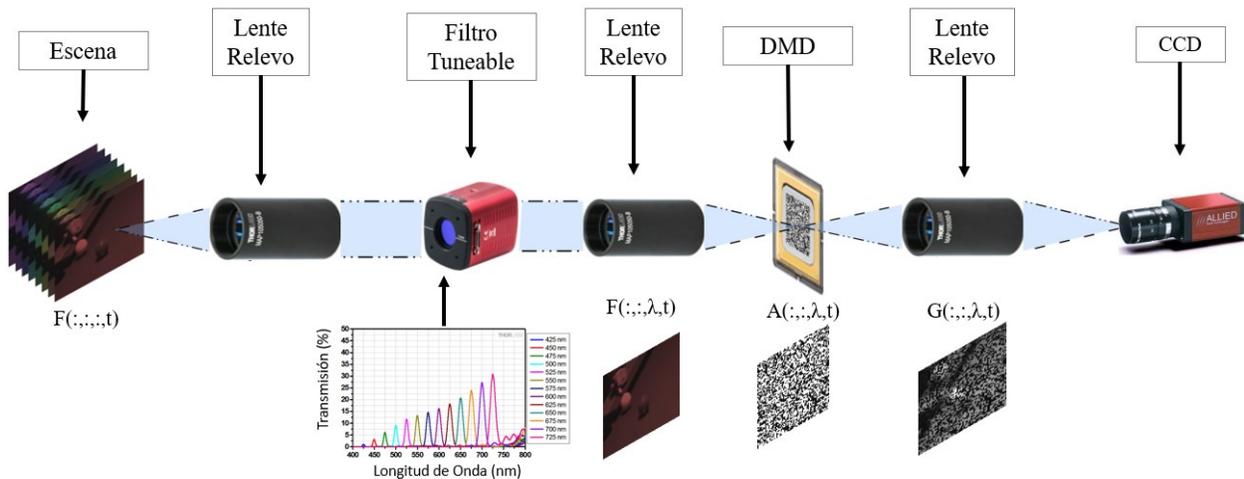


Figura. 36. Metodología de adquisición propuesta basada en la combinación de un filtro tuneable, un DMD y un único sensor para la captura de SV comprimidos.

Para la reconstrucción del SV capturado, en primer lugar, se implementó un método de estimación de imágenes mediante una red de eliminación de ruido del estado del arte y el algoritmo

de variación total, que obtiene una aproximación a escala de grises del vídeo espectral. Posteriormente, el algoritmo Plug and Play es alimentado con esta aproximación como una inicialización para obtener una reconstrucción de mejor calidad del vídeo espectral. Finalmente, el algoritmo se probará con capturas reales de vídeos espectrales obtenidos en un laboratorio mediante la sincronización de los elementos optoelectrónicos descritos en el capítulo 3.

4.1. Metodología de Adquisición

Este trabajo propone una metodología de adquisición y reconstrucción de SV en la que se logra comprimir la dimensión espectral y temporal con un mismo sensor. Para esto, se presenta un enfoque de ventana espacio-temporal en la que el SV se submuestra en sus dimensiones espaciales, de esta manera, la información espacio-espectral de múltiples disparos se contiene en cada uno de los elementos de la matriz propuesta, aumentando la compresión temporal. Además, se propone un novedoso sistema óptico compuesto principalmente por un filtro tuneable y un DMD, cuya función principal es generar aperturas codificadas de color en donde tanto la dimensión espectral como la temporal se pueden modular durante un tiempo de integración utilizando solo estos dos elementos ópticos.

La adquisición de las medidas del sistema en forma continua puede expresarse como

$$\mathbf{G}(x, y) = \int_T \int_\Lambda \mathbf{F}(x, y, \lambda, t) \mathbf{A}(x, y, \lambda, t) d\lambda dt, \quad (14)$$

donde \mathbf{F} es la fuente espectro-temporal continua, \mathbf{A} el código de apertura y \mathbf{G} las medidas adquiridas, además, x e y indexan las coordenadas espaciales, λ las espectrales y t la dimensión

temporal. La salida del sistema se obtiene integrando sobre la sensibilidad del rango espectral del detector Λ y durante un tiempo de integración T para los fotogramas.

Adicionalmente, se presenta el modelo matemático de detección discreta capaz de representar datos espaciales, espectrales y temporales, proyectando la información multidimensional en una medición bidimensional, definido como

$$\mathbf{G}_{i,j} = \sum_{t=0}^{N_t-1} \sum_{\lambda=0}^{N_\lambda-1} \mathbf{F}_{i,j,\ell,t} \cdot \mathbf{A}_{i,j,\ell,t}, \quad (15)$$

donde $i = \{0, \dots, N_x - 1\}$, $j = \{0, \dots, N_y - 1\}$, $\ell = \{0, \dots, N_\lambda - 1\}$, y $t = \{0, \dots, N_t - 1\}$ dónde N_x y N_y corresponden a la resolución espacial, N_λ es la resolución espectral, N_t es el número de fotogramas. Además, $\mathbf{G} \in \mathbb{R}^{N_x \times N_y}$ son las medidas compresas, $\mathbf{F} \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_t}$ es el video espectral, y $\mathbf{A} \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_t}$ representa un código de apertura aleatorio variable en el tiempo. Más específicamente, \mathbf{A} puede ser representada como $\mathbf{A}_{::,\ell,t} = \mathbf{C}_{::,\ell,t} \otimes \mathbf{B}_t$ con $\mathbf{C} \in \mathbb{R}^{\frac{N_x}{s} \times \frac{N_y}{s} \times N_\lambda \times N_t}$ como un código de apertura aleatorio variable en el tiempo de baja resolución espacial y

$$(\mathbf{B}_t)_{u,v} = \begin{cases} 1, & \text{si } u = \lfloor \frac{t}{s} \rfloor \text{ y } v = \text{mod}(t, s), \\ 0, & \text{de lo contrario,} \end{cases} \quad (16)$$

para $u = \{0, \dots, S_1\}$ y $v = \{0, \dots, S_1\}$, donde $S \in \mathbb{N}$ representa el tamaño de la ventana temporal.

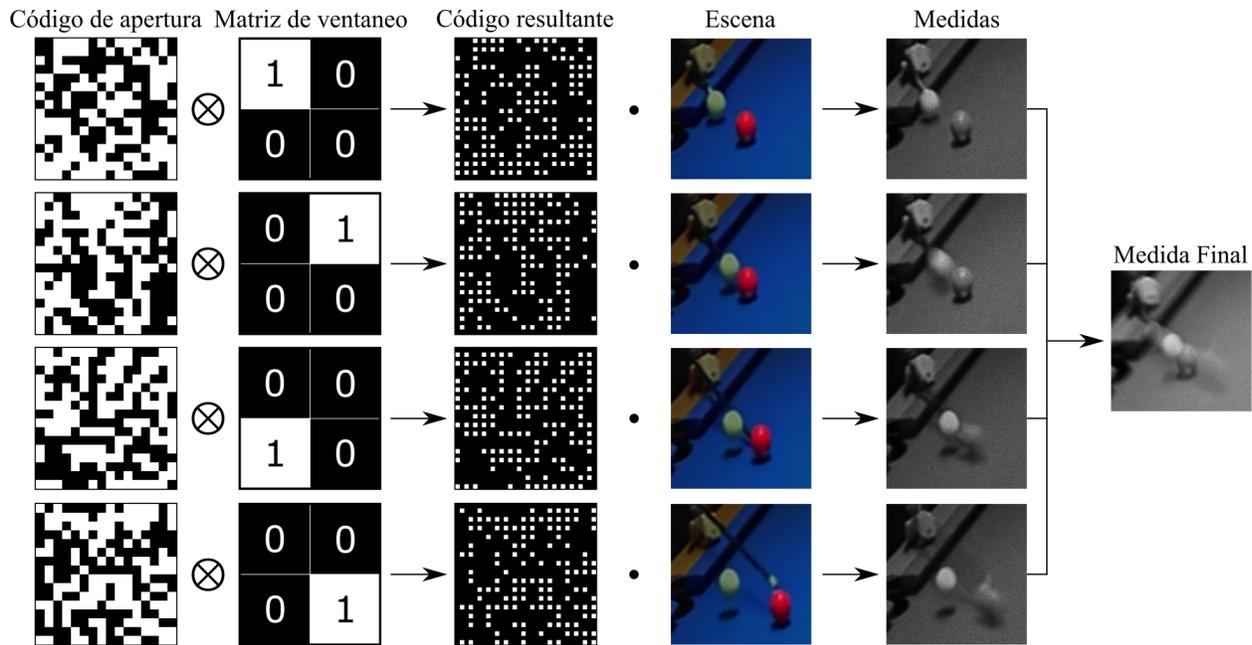


Figura. 37. Construcción del código de apertura, codificación de una escena y adquisición de medidas para el enfoque de ventana espacio-temporal con un tamaño de $S = 2$

La figura 37 muestra una representación del método de codificación propuesto, en esta, se realiza el producto kronecker entre la matriz de baja resolución \mathbf{C} y la ventana temporal \mathbf{B} , generando así, el código de apertura \mathbf{A} . De esta manera, para cada elemento de la matriz de ventaneo \mathbf{B} se obtiene una codificación diferente en la que se puede apreciar la repetición del código de apertura aleatorio y en donde se toman distintos píxeles según el patrón y elemento de la ventana. Realizando esta codificación, se sacrifica resolución espacial dado que no se captura toda la información, pero se alivia la compresión de fotogramas espectrales por pixel debido a que se comprime una serie menor de fotogramas por cada codificación resultante. Por último, la medida generada por la compresión de los fotogramas en cada codificación da como resultado la medida final mostrada en la última columna, que representa la medición capturada en el tiempo de integración.

4.2. Matriz de sensado

El modelo de sensado discreto para adquirir las medidas compresas \mathbf{G} puede establecerse matemáticamente como la proyección lineal

$$\mathbf{g} = \hat{\mathbf{H}}\mathbf{f} + \mathbf{w}, \quad (17)$$

donde $\mathbf{f} \in \mathbb{R}^n$ es la forma vectorial del SV con $n = N_x N_y N_\lambda N_t$, $\mathbf{g} \in \mathbb{R}^m$ es la forma vectorial de las medidas compresas con $m = N_x N_y$ y $\mathbf{w} \in \mathbb{R}^m$ representa un vector aleatorio de ruido de la imagen, denotado como

$$\mathbf{w} = \sigma \mathbf{k}, \quad (18)$$

donde $\mathbf{k} \sim \mathcal{N}_m(0, 1)$ y donde σ se calcula como

$$\sigma = \frac{\|\mathbf{g} - \mu_{\mathbf{g}}\|_F}{\sqrt{10^{snr/10} m}}, \quad (19)$$

en donde $\|\mathbf{g} - \mu_{\mathbf{g}}\|_F$ es la norma frobenius de la resta de las medidas compresas \mathbf{g} con su media $\mu_{\mathbf{g}}$ y snr es el nivel de ruido que se desea agregar.

Además, $\hat{\mathbf{H}} \in \mathbb{R}^{m \times n}$ se establece como la matriz de sensado escasa $\hat{\mathbf{H}} = [\mathbf{H}_0, \dots, \mathbf{H}_{N_t-1}]$ donde cada matriz \mathbf{H}_i tiene tamaño $N_x N_y \times N_x N_y N_\lambda$ conteniendo los códigos de apertura en $\mathbf{A}_{:, :, :, i}$. En la figura 38 se muestra un ejemplo de una matriz de sensado \mathbf{H}_0 para un único fotograma en un instante de tiempo.

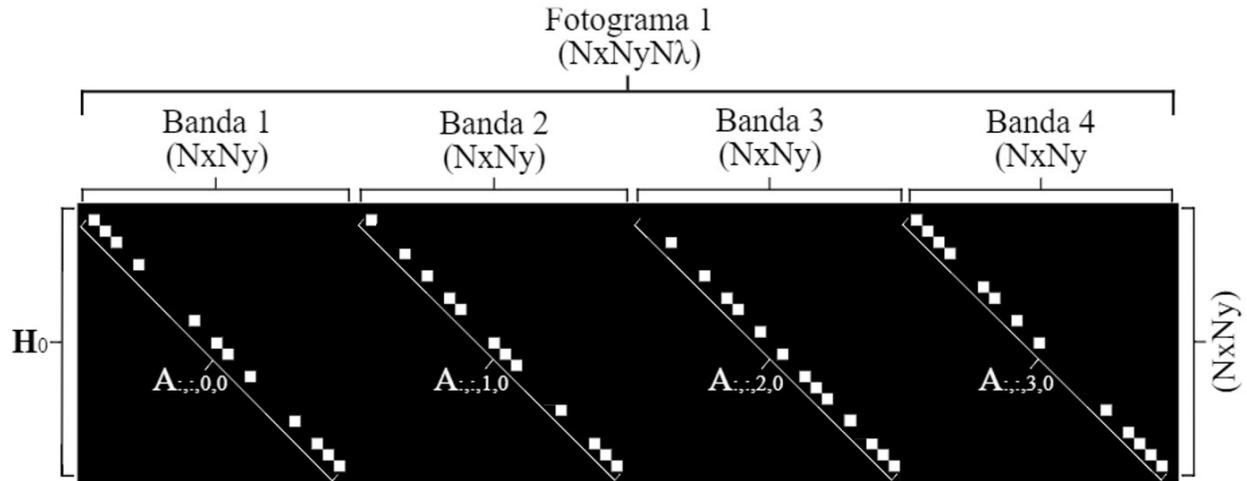


Figura. 38. Ejemplo de matriz de sensado para un código de apertura de tamaño $N_x = 5$, $N_y = 5$, $N_\lambda = 4$, $N_t = 1$.

4.3. Algoritmo de Reconstrucción PNP-ADMM

Una vez obtenidas las medidas compresas, el proceso de reconstrucción es llevado a cabo usando algoritmos tradicionales como el método de dirección alterna de multiplicadores (ADMM, de sus siglas en inglés, *alternating direction method of multipliers*) y su variante como el PNP (de sus siglas en inglés, *plug and play*). En general, el ADMM divide un problema de optimización convexa en problemas más pequeños y fáciles de manejar por lo que estableceremos primero el problema de optimización como:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} z(\mathbf{f}) + \tau w(\mathbf{f}), \quad (20)$$

$z(\mathbf{f})$ es la función objetivo definida como $z(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$, adicionalmente, $w(\mathbf{f})$ es una función de regularización tipo eliminador de ruido y $\tau \in \mathbb{R}^+$ es una constante de regularización. Sin em-

bargo, el problema de optimización definido anteriormente es un problema sin restricciones, por lo que se convertirá a un problema restringido como:

$$(\hat{\mathbf{f}}, \hat{\mathbf{v}}) = \underset{\mathbf{f}}{\operatorname{argmin}} z(\mathbf{f}) + \tau w(\mathbf{v}), \text{ sujeto a } \mathbf{f} = \mathbf{v}, \quad (21)$$

donde $\mathbf{v} \in \mathbb{R}^n$ es una variable de minimización auxiliar. Si usamos el proceso del método de Lagrangiano Aumentado Bertsekas (2014) para resolver este problema, obtenemos la función:

$$\mathcal{L}(\mathbf{f}, \mathbf{v}, \mathbf{u}) = z(\mathbf{f}) + \tau w(\mathbf{v}) + \mathbf{u}^T(\mathbf{f} - \mathbf{v}) + \frac{\rho}{2} \|\mathbf{f} - \mathbf{v}\|^2, \quad (22)$$

donde $\mathbf{u} \in \mathbb{R}^n$ es una variable multiplicadora de Lagrange escalada y $\rho \in \mathbb{R}^+$ es una variable de penalización no negativa. El objetivo es minimizar esta función del lagrangiano aumentado obtenida anteriormente, el ADMM realiza este proceso mediante los siguientes pasos iterativos hasta converger:

$$\hat{\mathbf{f}}^{(k+1)} = \underset{\mathbf{f}}{\operatorname{argmin}} z(\mathbf{f}) + \frac{\rho}{2} \|\mathbf{f} - \tilde{\mathbf{f}}^{(k)}\|^2, \quad (23)$$

$$\mathbf{v}^{(k+1)} = \underset{\mathbf{v}}{\operatorname{argmin}} \tau w(\mathbf{v}) + \frac{\rho}{2} \|\mathbf{v} - \tilde{\mathbf{v}}^{(k)}\|^2, \quad (24)$$

$$\bar{\mathbf{u}}^{(k+1)} = \bar{\mathbf{u}}^{(k)} + \left(\hat{\mathbf{f}}^{(k+1)} - \mathbf{v}^{(k+1)} \right), \quad (25)$$

donde $\bar{\mathbf{u}}^{(k)} = \frac{1}{\rho} \mathbf{u}^{(k)}$ es una variable multiplicadora de Lagrange y tanto $\tilde{\mathbf{f}}^{(k)} = \mathbf{v}^{(k)} - \bar{\mathbf{u}}^{(k)}$ como $\tilde{\mathbf{v}}^{(k)} = \hat{\mathbf{f}}^{(k+1)} + \bar{\mathbf{u}}^{(k)}$ son variables de minimización. Además, $\hat{\mathbf{f}}^{(k+1)}$ es un paso de inversión, $\mathbf{v}^{(k+1)}$ es el paso de eliminación de ruido y $\bar{\mathbf{u}}^{(k+1)}$ es el paso de actualización. Adicionalmente, la calidad de la reconstrucción puede ser mejorada usando la variante del ADMM denominada ADMM PNP la cual sustituye el paso de eliminación de ruido mediante un algoritmo que realiza esta tarea denotado como D_σ .

$$\mathbf{v}^{(k+1)} = D_\sigma \left(\tilde{\mathbf{v}}^{(k)} \right), \quad (26)$$

donde $\sigma = \sqrt{\frac{\tau}{\rho}}$. Este nuevo paso de eliminación de ruido es la ventaja clave del algoritmo PNP y lo convierte en una herramienta poderosa para restaurar imágenes de alta calidad. El algoritmo de eliminación de ruido que es comúnmente utilizado es el filtro de eliminación de ruido recursivo (RF, de sus siglas en inglés *recursive filter*) usado en la restauración de imágenes Gatal and Oliveira (2011).

4.4. Reconstrucción Espectro-Temporal de SV

El proceso de estimación del SV a partir de sus medidas compresas se muestra en la figura 39, en esta, se observa que la reconstrucción se obtiene a través de dos etapas. La primera etapa realiza la estimación temporal del vídeo en escala de grises y la segunda, alimentada por la etapa anterior, obtiene la información espectro-temporal de todo el SV.

En la primera etapa se utiliza un método de reconstrucción de imágenes compresivas de única captura (SCI, de sus siglas en inglés *snapshot compressive imaging*) en el que se incorpora

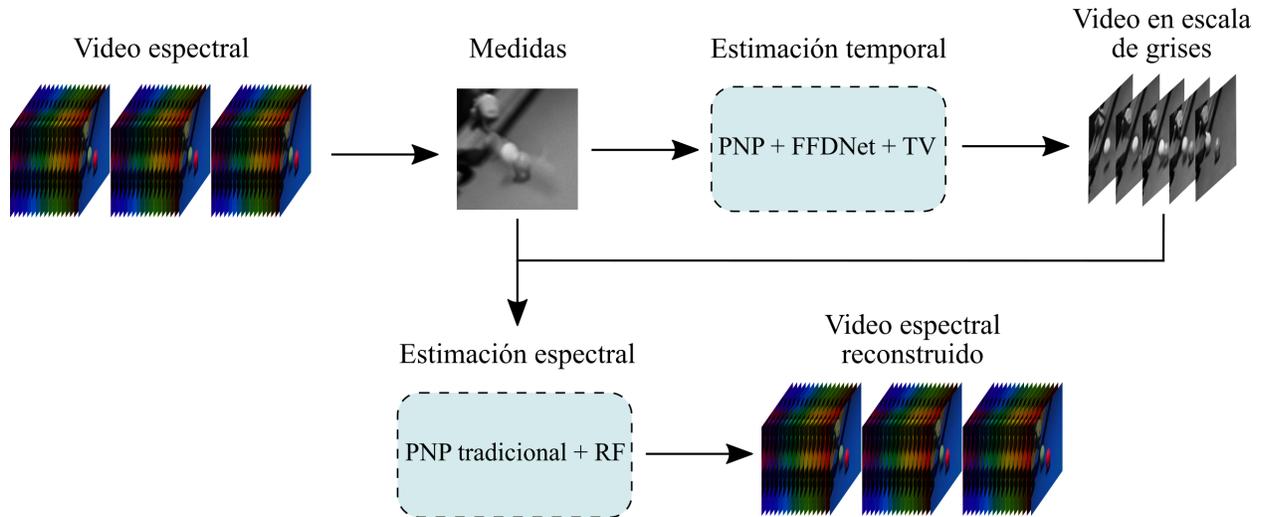


Figura. 39. Flujo del método propuesto: El algoritmo recibe como entrada el SV, posteriormente, se obtienen las medidas comprimidas bajo la metodología propuesta y estas se pasan al algoritmo que implementa la FFDNet y la TV en el marco de trabajo del PNP. Luego, se obtiene como salida un vídeo en escala de grises como aproximación temporal que es utilizado para inicializar un PNP tradicional, el cual a partir de las medidas comprimidas y esta aproximación, reconstruye el vídeo espectral.

la red de eliminación de ruido profunda conocida como FFDNet, presentada en la sección 2.7.4 y un algoritmo como el TV, en el marco de trabajo del Plug and Play Qiu et al. (2021). La principal ventaja de este método es que a pesar de que los enfoques de la FFDNet y el TV por separado no logran obtener un buen rendimiento, la combinación de estas dos técnicas permite que se puedan aprovechar al mismo tiempo sus ventajas y de esta forma que contribuyan entre sí para obtener mejores resultados de reconstrucción. Específicamente, el algoritmo recibe las medidas capturadas en su entrada y realiza 100 iteraciones, en primer lugar, realiza 20 iteraciones únicamente con la FFDNet y posteriormente realiza 80 más con la combinación tanto de la red como de el TV en las que se proporcionan múltiples niveles de ruido. La salida del algoritmo da como resultado la estimación del vídeo en escala de grises.

Para estimar la información espectro-temporal del SV, se implementó el algoritmo PNP tradicional con un RF como algoritmo de eliminación de ruido. En esta etapa se utiliza la estimación temporal del vídeo en escala de grises obtenida previamente para ser usada como inicialización del algoritmo. Esto permite obtener una reconstrucción de mejor rendimiento al suministrarle al PNP mayor información previa, lo que influye directamente en su convergencia. Por esto, la combinación de estas técnicas representa una ventaja frente al uso de un PNP tradicional únicamente para la estimación del SV. La aplicación de esta metodología permite obtener reconstrucciones de SV a una muy alta tasa de compresión, a la vez que conserva la información espectral y temporal de la escena. Es importante resaltar que las medidas compresas usadas en este método se generan a partir del SV y que, además, serán las mismas medidas usadas tanto para la estimación temporal en escala de grises como para la reconstrucción espectro-temporal del SV.

4.5. Captura en el Laboratorio de Vídeos Espectrales

Para la captura de SV en el laboratorio, se empleó el sistema sincronizado que fue presentado en el capítulo 3, compuesto por un grupo de dispositivos optoelectrónicos como la cámara Stingray F-080B, el monocromador TLS-300XR, un microcontrolador Arduino, elementos para crear la dinámica de la escena y el centro de computo como puesto de control para dicha interconexión. A través de la sincronizaron de estos elementos se automatiza y optimiza el proceso de adquisición de SV de baja y alta complejidad dinámica.

El proceso de captura efectuado por dicho sistema comienza por la vinculación de la cámara, el monocromador y el Arduino al computador donde se encuentra el código en el software Matlab, Anexo 1. Después de esto, se inicia el proceso de captura, donde el primer paso es la

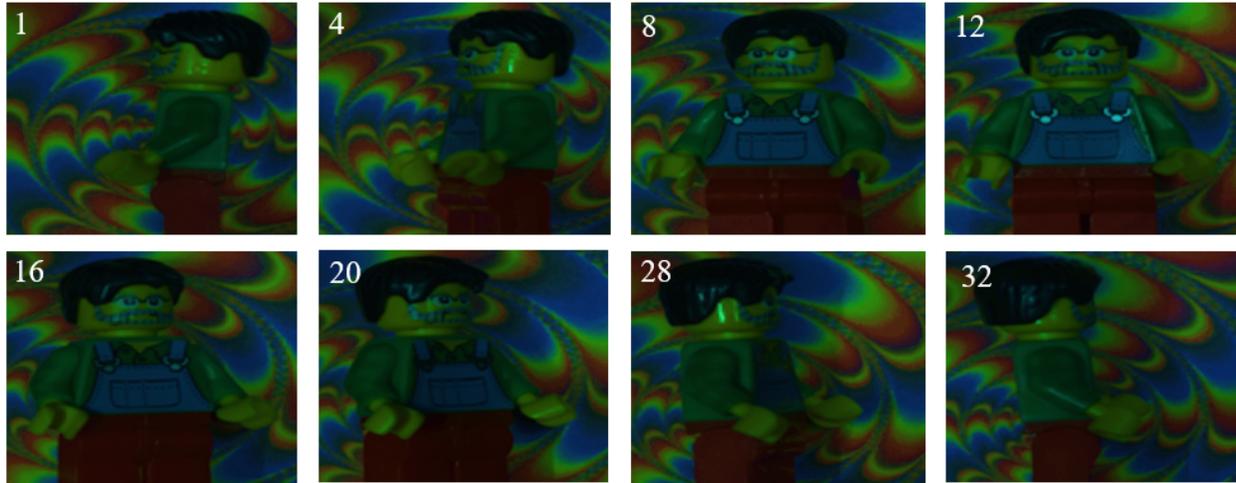


Figura. 40. Captura de SV con 16 bandas espectrales en el rango de 450-700 nm equiespaciadas y 32 fotogramas, a partir del sistema sincronizado en el laboratorio de óptica.

activación del Arduino para de esta forma ubicar la escena en la posición 0° , resaltando que en la figura 40 se adicionó un servomotor para realizar una rotación del fondo el cual se encuentra sincronizado con la rotación de la escena y de esta forma emular escenas con diversas dinámicas espaciales, comprobando de manera general el comportamiento del método propuesto, posteriormente la cámara y el monocromador se activaron debido a una serie de instrucciones programadas, tales como el tiempo de integración, y la luz entregada por el monocromador la cual realizará un barrido desde 400 a 700 nm en pasos de 3 nm, para de esta manera finalizar el ciclo del sistema enviando la orden al Arduino de que realice el giro de 10° de la escena y del fondo, continuando con el siguiente fotograma, concluyendo finalmente con la captura automatizada y repetible de un vídeo espectral por medio de una sincronización por medio de hardware y software de diversos dispositivos ópticos que se encuentran en el laboratorio de óptica.

5. Resultados

A continuación, se presentan los resultados simulados a partir de vídeos obtenidos por medio de la combinación de una red neuronal del estado del arte que permite obtener estimaciones espectrales precisas de vídeos RGB, y por medio de la captura de vídeos espectrales mediante un sistema sincronizado de dispositivos optoelectrónicos implementado en el laboratorio. Se prueba el método propuesto y se usan métricas de evaluación para determinar la calidad de la información espacial, espectral y temporal estimada. Para esto, se definió un tamaño espacial de $N_x = 1024$ y $N_y = 1024$, con $N_\lambda = 16$ como la resolución espectral, $N_t = 32$ como el número de fotogramas y el tamaño de ventana escogido fue de $S = 2$, que permite comprimir 8 fotogramas por elemento de la ventana. Teniendo en cuenta estas dimensiones, la compresión puede calcularse en $C_v = \frac{N_x N_y}{N_x N_y N_\lambda N_t}$ donde N_t es la compresión temporal de la ventana, obteniendo $C_v = \frac{1}{16 \times 32} = 0,78\%$ de la información. Por último, el valor de transmitancia se fijó en 50% según los resultados obtenidos en la sección 5.3 y se agregó un nivel de ruido de $SNR = 20$ dB a las medidas según la ecuación 19.

5.1. Conjunto de Datos de SV

Dado que no hay bases de datos de SV en el estado del arte, para obtener los vídeos espectrales necesarios para la tarea de reconstrucción y su posterior análisis, se utilizó el dataset Need for Speed Kiani Galoogahi et al. (2017). Esta base de datos contiene 100 vídeos RGB capturados con cámaras de alta velocidad de fotogramas, específicamente a 240 fps conteniendo escenas del mundo real con tamaños de imagen desde 1280×720 hasta 1920×1080 .

Primero, para el vídeo seleccionado se toman 32 fotogramas verificando que existan cam-

bios espaciales entre ellos, para cada vídeo un fotograma corresponde a una sola imagen, es decir se utilizan 32 imágenes por vídeo. Posteriormente, se implementa la red de regresión jerárquica (HRNet, de sus siglas en inglés *Hierarchical Regression Network*) Zhao et al. (2020) para obtener la estimación espectral de cada fotograma del vídeo, para hacerlo, las imágenes se ajustan espacialmente a la entrada de red $512 \times 482 \times 3$. La salida de la red consta de 32 cubos de datos tridimensionales, donde cada cubo corresponde a una imagen espectral de 31 bandas con un tamaño de $512 \times 482 \times 31$ que se redimensionan espacialmente a 1024×1024 . Finalmente, se seleccionan 16 bandas igualmente espaciadas para cada cubo y las 32 imágenes espectrales se concatenan sobre la dimensión temporal para obtener un vídeo espectral 4-D con tamaño $1024 \times 1024 \times 16 \times 32$. Dicho proceso se observa en la figura 41

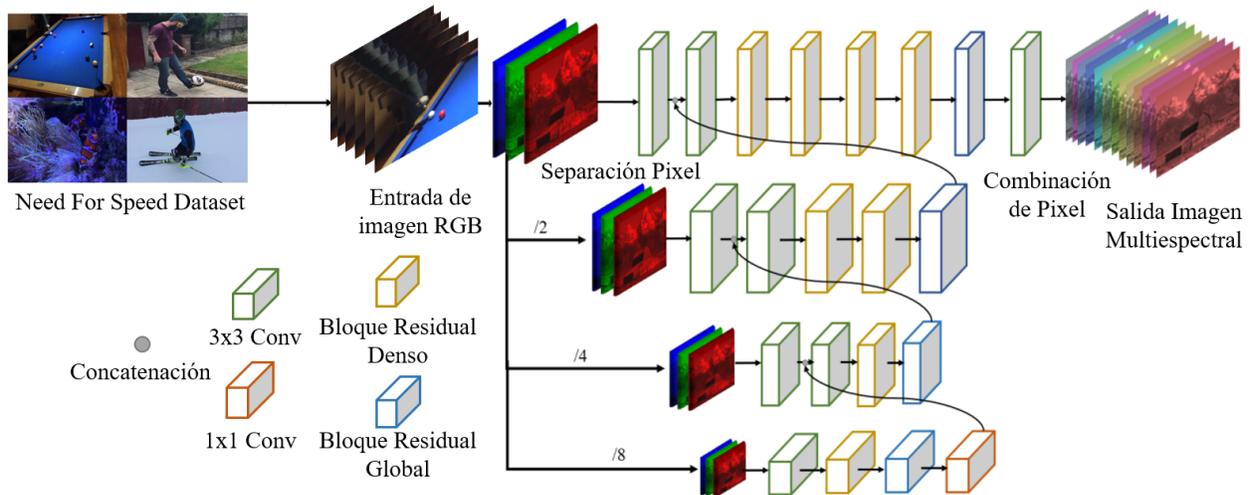


Figura. 41. Red Neuronal HRNet, ingresando vídeos reales RGB con tamaños desde 1280×720 hasta 1920×1080 , finalizando el procesamiento con vídeos espectrales de tamaño $1024 \times 1024 \times 16 \times 32$

De esta forma se logran obtener múltiples SV con diversificación de escenarios reales, que

serán usados para evaluar el desempeño del método propuesto en el caso de la reconstrucción de vídeos a partir de simulaciones.

5.2. Métricas de Evaluación

Con el objetivo de determinar la calidad de las reconstrucciones para un enfoque de CSI, se usan múltiples métricas de evaluación que pueden determinar la calidad de la reconstrucción, su similitud espectral o estructural, entre otras características. Entre estas métricas se destacan la proporción máxima de señal a ruido (PSNR, de sus siglas en inglés *Peak signal-to-noise ratio*), el índice de similitud estructural (SSIM, de sus siglas en inglés *structural similarity index measure*), el error cuadrático medio (MSE, de sus siglas en inglés *mean squared error*) y el mapeador de ángulo espectral (SAM, de sus siglas en inglés *spectral angular mapper*).

El PSNR es una métrica de calidad de imagen que es frecuentemente utilizada en el CSI para estimar la calidad en una reconstrucción de una adquisición realizada a partir de menos medidas. Esta métrica se expresa en la escala logarítmica de decibelios (dB), entre mayor sea el PSNR se tendrá una reconstrucción de mejor calidad. Teniendo a $\mathbf{f} \in \mathbb{R}^n$ como forma vectorial del vídeo espectral y a $\hat{\mathbf{f}} \in \mathbb{R}^n$ como la forma vectorial de las medidas estimadas, podemos expresar el PSNR como:

$$PSNR = 20 \log_{10} \left(\frac{MAX_{\mathbf{f}}}{\sqrt{MSE(\mathbf{f}, \hat{\mathbf{f}})}} \right), \quad (27)$$

donde MAX_f es el valor máximo de los píxeles en el vídeo espectral \mathbf{f} y $MSE(\mathbf{f}, \hat{\mathbf{f}})$ es el error cuadrático medio definido como:

$$MSE(\mathbf{f}, \hat{\mathbf{f}}) = \frac{1}{n} \sum_{i=0}^{n-1} (\mathbf{f}_i - \hat{\mathbf{f}}_i)^2. \quad (28)$$

Adicionalmente, como una métrica para apoyar los análisis de resultados y tener un estudio más detallado de las reconstrucciones, el SSIM suele ser utilizado como otro indicador de desempeño basado en el análisis estructural de las medidas estimadas respecto a las reales.

El SSIM es una métrica perceptual que cuantifica la diferencia estructural entre dos medidas, el vídeo espectral original y el vídeo espectral reconstruido a partir de menos medidas. Esta métrica tiene un rango de $[0, 1]$ donde entre mayor sea el valor obtenido mejor será la reconstrucción, se calcula según la ecuación:

$$SSIM(\mathbf{f}, \hat{\mathbf{f}}) = \frac{(2\mu_{\mathbf{f}}\mu_{\hat{\mathbf{f}}} + C_1) + (2\sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} + C_2)}{(\mu_{\mathbf{f}}^2 + \mu_{\hat{\mathbf{f}}}^2 + C_1) (\sigma_{\mathbf{f}}^2 + \sigma_{\hat{\mathbf{f}}}^2 + C_2)}, \quad (29)$$

donde $\mu_{\mathbf{f}}$ y $\mu_{\hat{\mathbf{f}}}$ son el promedio de \mathbf{f} y $\hat{\mathbf{f}}$ respectivamente, además, $\sigma_{\mathbf{f}}^2$ y $\sigma_{\hat{\mathbf{f}}}^2$ son la varianza de \mathbf{f} y $\hat{\mathbf{f}}$ respectivamente, por último, $\sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}}$ corresponde a la covarianza entre ellos. Adicionalmente, $C_1 = (k_1P)^2$ y $C_2 = (k_2P)^2$ corresponden a variables que aseguran la estabilidad en el caso de que el denominador sea 0, y P es un rango dinámico de los valores de los píxeles con $k_1 = 0,01$ y $k_2 = 0,03$ por defecto.

La métrica SAM representa la similitud espectral entre dos muestras determinando el ángulo entre sus espectros, para hacerlo usa las medidas de comparación como si fueran vectores y calcula el ángulo entre el vector del espectro de la imagen de referencia y el vector del espectro de la imagen estimada. Dado que se trata de hallar la distancia entre dos vectores los ángulos más

pequeños representan una mejor similitud. El SAM puede calcularse como:

$$SAM = \cos^{-1} \left(\frac{\sum_{l=0}^{L-1} \mathbf{F}_l \hat{\mathbf{F}}_l}{\sqrt{(\sum_{l=0}^{L-1} \mathbf{F}_l)} \sqrt{(\sum_{l=0}^{L-1} \hat{\mathbf{F}}_l)}} \right) \quad (30)$$

5.3. Análisis de Transmitancia

La transmitancia se define como la proporción de luz o cantidad de energía que pasa a través de una muestra a cierta longitud de onda. El valor de la transmitancia óptica un objeto se puede establecer cómo

$$\mathbf{T}_r = \frac{I}{I_0}, \quad (31)$$

donde I es la cantidad de píxeles del SV, es decir, su resolución y I_0 es la cantidad de píxeles a través de los cuales puede pasar la luz de la muestra. En otras palabras, la transmitancia es el porcentaje de píxeles que permanecen activados en el código de apertura, es decir, si el valor del píxel es 1, permite capturar la información del SV y si es 0, bloqueará la luz.

Al momento de implementar un método de adquisición y estimación de SV utilizando un enfoque de CS, la transmitancia es uno de los principales factores que afectan el rendimiento de la reconstrucción Huang et al. (2013). Por esto, se realizó un análisis de transmitancia usando una variación porcentual para 25 %, 50 % y 75 % en la tarea de reconstrucción de 10 SV del conjunto de datos obtenido en la sección 5.1, empleando las métricas PSNR y SSIM. Los resultados obtenidos para cada vídeo se representan en la figura 42 utilizando un diagrama de cajas para determinar el

valor óptimo de transmitancia que obtiene la mejor reconstrucción.

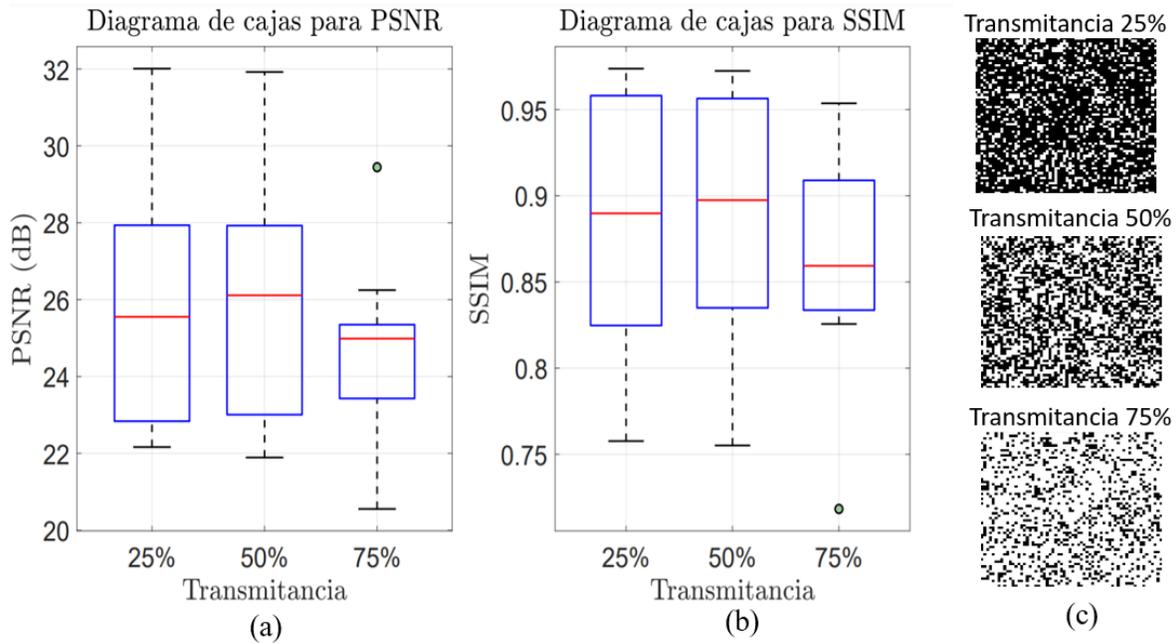


Figura. 42. Resultados de diagramas de caja para 10 reconstrucciones de video espectrales usando (a) PSNR y (b) SSIM. En (c) se muestran los códigos de apertura a diferentes valores de transmitancia.

La figura 42(a) muestra que el rango intercuartílico (IQR, de sus siglas en inglés interquartile range) correspondientes a los datos entre los cuartiles uno y tres, se encuentran aproximadamente entre 23 dB y 28 dB para transmitancias del 25% y 50%, mientras que para el 75% se observan valores inferiores, que van tan solo de 23.5 dB a 25.5 dB incluyendo un dato atípico por encima de su máximo. Para el máximo calculado, se obtienen nuevamente valores similares entre el 25% y 50% llegando a 32 dB y de la misma forma, la transmitancia del 75% obtiene un valor mucho más bajo, ligeramente por encima de los 26 dB. Adicionalmente, para los valores mínimos se aprecia un comportamiento similar en el que el 25% y 50% obtienen al rededor de 22 dB y el 75% sigue con su bajo rendimiento, obteniendo cerca a los 20.5 dB. Finalmente, a pesar de que los resultados

de los diagramas de cajas para transmitancias del 25 % y 50 % fueron muy cercanos entre ellos, la del 50 % presenta una mayor mediana, de más de 26 dB, frente a 25.5 dB para el 25 %.

En la figura 42(b) se muestra que el IQR para la transmitancia del 25 % se encuentra entre 0.83 y 0.96, mientras que para el 50 % está entre 0.84 y 0.96, y para el 75 % muestra un menor desempeño, ubicándose entre 0.83 y 0.91 con un valor atípico muy alejado de su mínimo. Además, a pesar de que la transmitancia del 75 % tiene un mínimo más alto obteniendo 0.82 frente al 0.76 obtenido por el 25 % y 50 % aproximadamente, tiene una mediana más baja que las demás, de 0.86, mientras la transmitancia del 50 % tiene el valor de mediana más alto con casi 0.9 y la transmitancia del 25 % obtiene 0.89. Finalmente, el máximo para el 25 % y 50 % se ubicó en 0.97 y para el 75 % fue más bajo, ligeramente superior al 0.95.

Por último, en la figura 42(c) se muestran los diferentes códigos de apertura generados aleatoriamente con los valores de transmitancia estudiados, concluyendo así que la transmitancia del 25 % presenta menor transmisión de información, y el código de apertura del 75 % presenta una transmisión totalmente contraria a la anterior, infiriendo de esta manera que un valor intermedio de transmitancia del 50 % debe ser el óptimo, y por medio del análisis anteriormente realizado es resaltable que el promedio en términos de PSNR obtenido por las transmitancias de 25 %, 50 % y 75 % fue de 26.03, 26.13 y 24.71, junto con un promedio de SSIM de 0.885, 0.89 y 0.863, respectivamente. Implicando que el uso de una transmitancia del 50 % producirá un PSNR y SSIM más alto en la reconstrucción.

5.4. Resultados de Simulación

En esta sección se muestran múltiples análisis y resultados para la reconstrucción de dos SV obtenidos bajo el procedimiento de la sección 5.1 y dos SV capturados experimentalmente.

5.4.1. Análisis Espacio-Temporal. A continuación se presentarán los resultados de reconstrucción espacio-temporal obtenidos mediante el método propuesto. Primero se muestran los resultados obtenidos en la figura 43 para la reconstrucción de un SV con una escena de un juego de billar para la primera etapa de estimación temporal. Para esto, en la primera fila se muestra la escena original en escala de grises y en la segunda se encuentra la escena reconstruida para los fotogramas 4, 12, 20 y 28 de manera que pueda observarse el movimiento espacial de la misma. Adicionalmente, en la parte inferior izquierda de cada fotograma reconstruido se encuentran los valores de PSNR y SSIM obtenidos con el objetivo de realizar un análisis cuantitativo de la calidad de la estimación.

Para la figura anteriormente mencionada, puede observarse que la dimensión temporal de la escena se estima correctamente, notándose el movimiento entre cada fotograma. Por ejemplo, las bolas y el taco de billar de la reconstrucción describen un movimiento y posición igual al dado en la escena original incluso para detalles más pequeños como las sombras de los objetos. Adicionalmente, los valores de PSNR resultan adecuados, obteniendo entre 25.87 dB a 27.94 dB mientras que los resultados de SSIM son más bajos, de 0.544 a 0.563, debido a características no recuperadas como el taco del billar del fotograma 28.

Por otro lado, en la figura 44 se presentan los resultados adquiridos para la reconstrucción

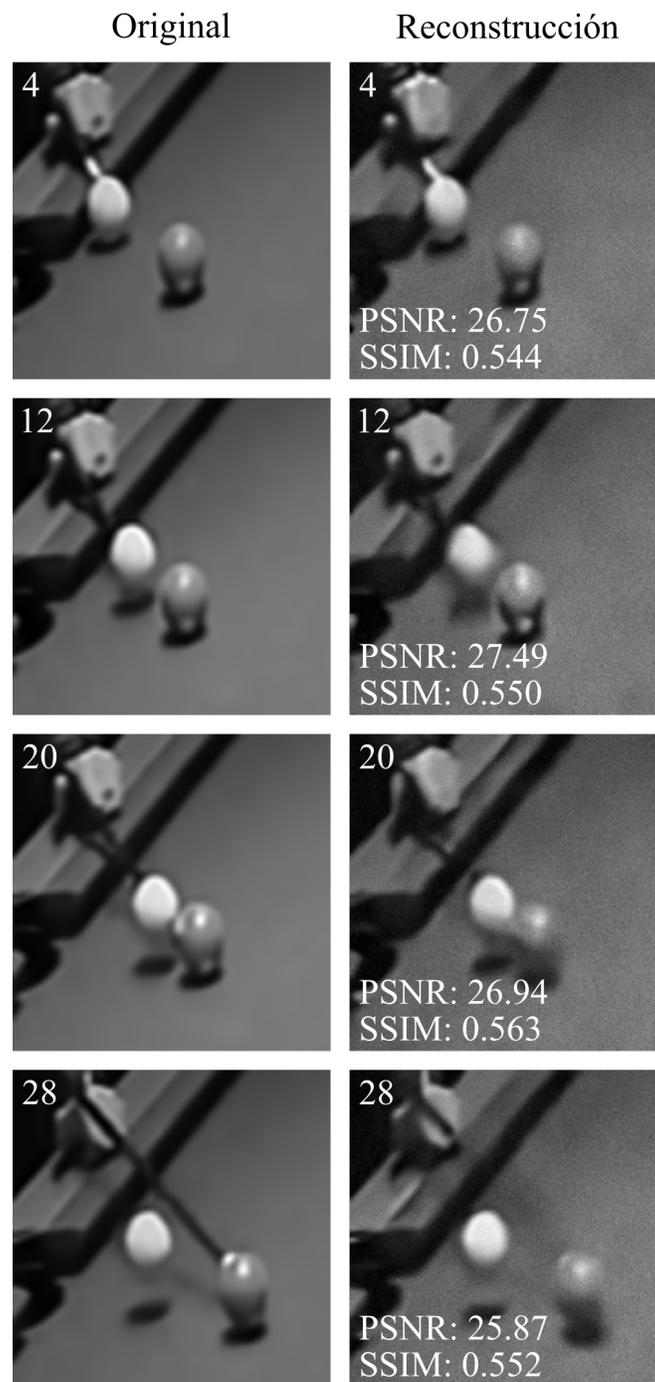


Figura. 43. Resultados para un SV de un juego de billar en donde se aprecian los fotogramas originales y la reconstrucción temporal en escala de grises. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente.

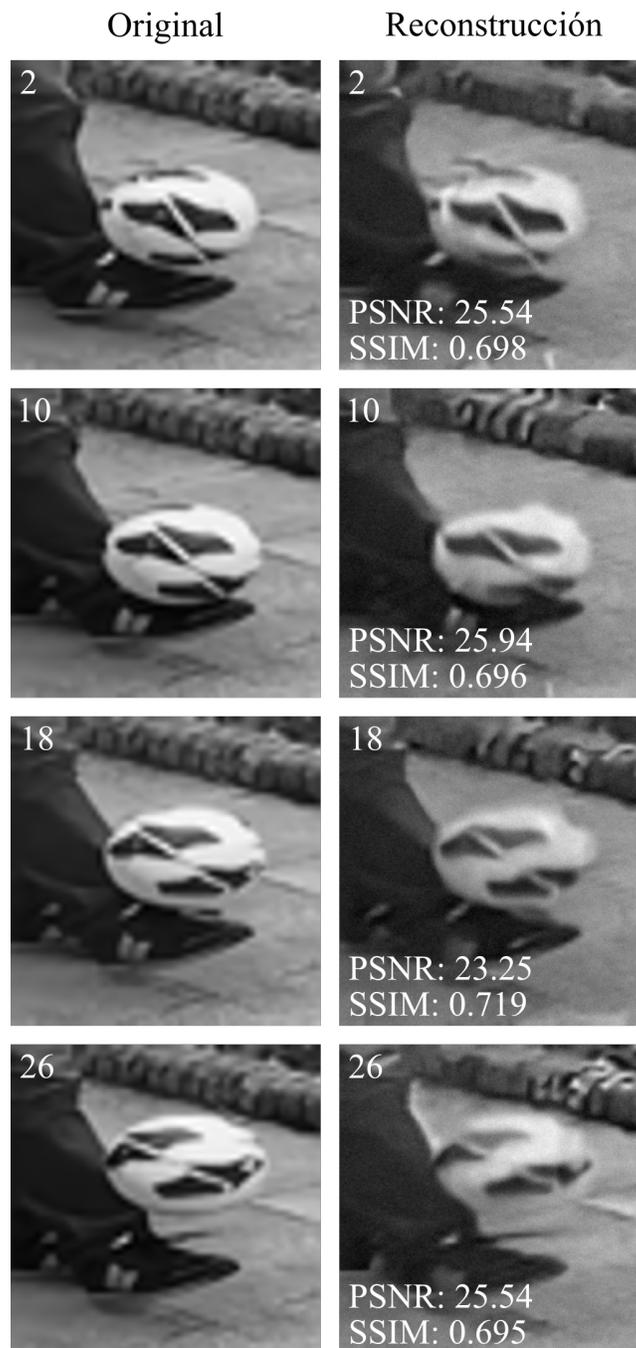


Figura. 44. Resultados para un SV de un balón de fútbol rebotando en donde se aprecian los fotogramas originales y la reconstrucción temporal en escala de grises. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente.

de un SV con la escena de un balón de fútbol rebotando, en primer lugar, se presentará análisis de estimación temporal. Inicialmente se muestra la escena original en escala de grises en la primera fila y para la segunda se muestra la escena reconstruida en los fotogramas 2, 10, 18, 26 con el objetivo de visualizar el movimiento entre cada fotograma. Al igual que en la figura 43, se presentan los valores de PSNR y SSIM que permiten realizar un análisis cuantitativo de la calidad de la estimación del método propuesto, donde en mencionada escena se puede notar que el movimiento tanto del balón como del pie de la persona se estima correctamente, por ejemplo, se puede notar que el balón sube y baja de la misma manera que en la escena original. Además, detalles pequeños como el cordón del zapato se conservan, como en el fotograma 26. Para las métricas de PSNR y SSIM se obtiene un máximo de 25.94 dB y 0.719 y un mínimo de 23.25 dB y 0.695, respectivamente, lo que muestra una buena aproximación que servirá para complementar la reconstrucción del SV.

5.4.2. Análisis Espectro-Temporal. Para la reconstrucción espectro-temporal se obtienen resultados satisfactorios en el rendimiento del método, por lo tanto se prosigue a realizar un análisis en el área de reconstrucción espectro-temporal del método.

Como primer análisis a realizar, en la figura 45 se presentan resultados espectro-temporales para el SV reconstruido en su representación RGB usando los mismos fotogramas de la figura 43, es decir, la escena original y reconstruida con iguales métricas de evaluación.

Al igual que para la estimación en escala de grises, en la figura 45 se observa como se recupera correctamente el movimiento de la escena entre los fotogramas, notándose detalles finos y delgados como el taco del billar, cuya forma se conserva así como preserva los bordes en estructuras como la mesa. De la misma manera, se conservan características como la forma y color de

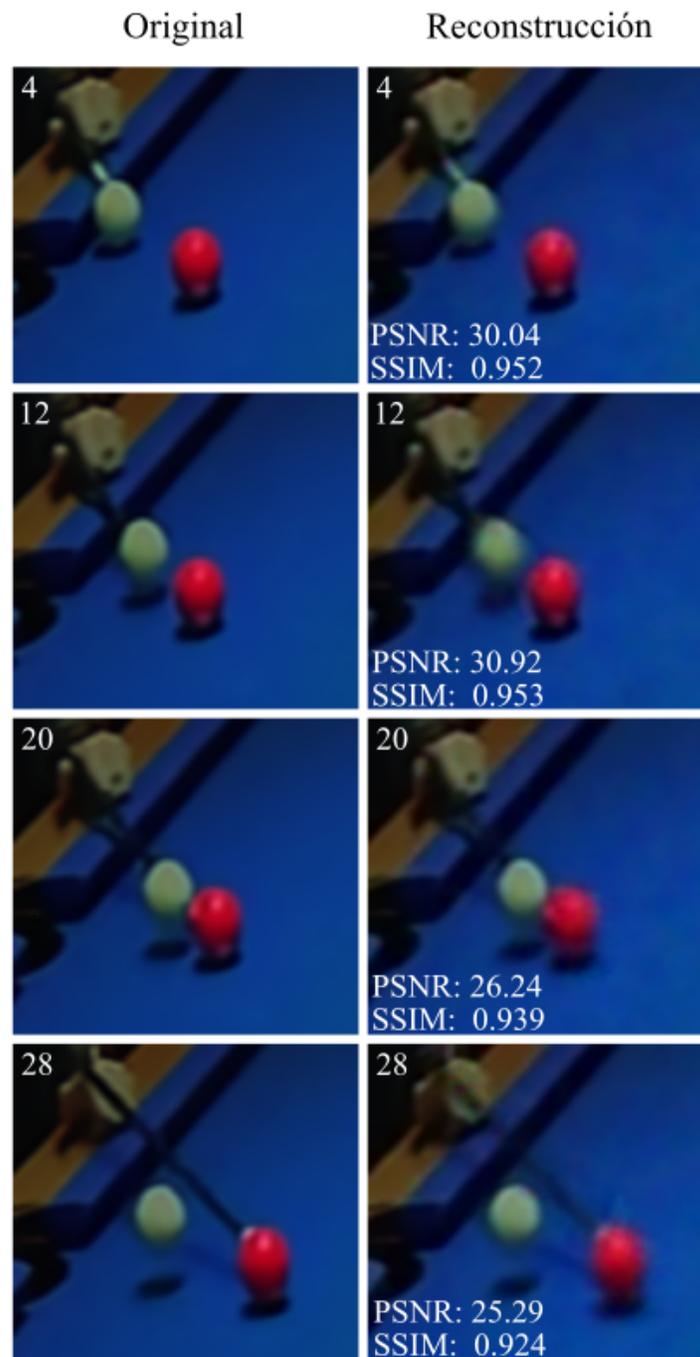


Figura. 45. Resultados espectro-temporales SV de un juego de billar, donde se muestran los fotogramas originales y la reconstrucción temporal en representación RGB. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente.

las bolas y la mano del jugador. Adicionalmente, a pesar de que como se había mencionado anteriormente, el fotograma 28 de la figura 43 no estimaba correctamente el taco de billar, se puede observar que para la reconstrucción del SV su forma, tamaño y movimiento se reconstruyen de mejor manera. Para las métricas de PSNR y SSIM se alcanzan los 30.92 dB y 0.953 en estos fotogramas, respectivamente, se puede observar una mejoría notoria, principalmente para la métrica SSIM que había sido baja en la estimación de escala de grises.

Posteriormente, en la figura 46 se presentan los resultados obtenidos para la estimación espectro-temporal del SV para los fotogramas 2, 10, 18, 26 en su representación RGB. Aquí, se puede notar una mejoría en la calidad de la imagen reconstruida, por ejemplo, se logran recuperar detalles como los inscritos en la pelota, conservando su forma y silueta, así como se mantiene la forma del pantalón, sus pliegues e incluso los cordones del zapato de la persona. Además, en las zonas estáticas de la escena las características son aún más visibles, el muro anaranjado ubicado en la parte superior derecha de la escena es recuperado con detalles muy similares a la escena original, donde se conservan sus grietas y estructura, a diferencia de la figura 44 en la que esto no pasaba.

5.4.3. Análisis Espectral. Con el objetivo de complementar el análisis de los resultados presentados anteriormente, se estudiarán espectralmente las reconstrucciones obtenidas para un fotograma de cada SV recuperado. Para esto, se tomaron 6 de las 16 bandas espectrales en el rango de 400 a 700 nm de forma equiespaciada y se usaron las métricas de PSNR y SSIM, de manera que se pueda observar y determinar la calidad de la estimación de la respuesta espectral del método propuesto. La figura 47 muestra los resultados obtenidos para el fotograma 4 de la figura 45 en donde la primera columna corresponde a las bandas espectrales originales y la segunda

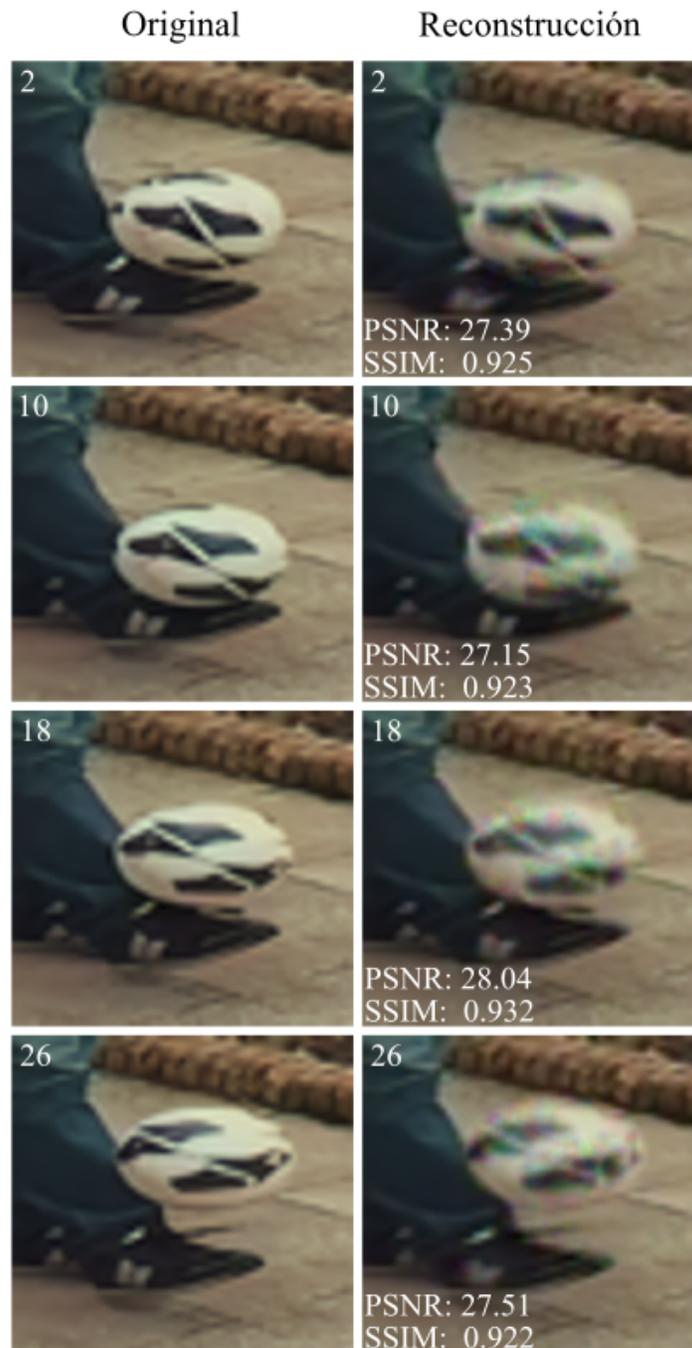


Figura. 46. Resultados espectro-temporales para un SV de un balón de fútbol rebotando en donde se muestran los fotogramas originales y la reconstrucción temporal en representación RGB. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente.

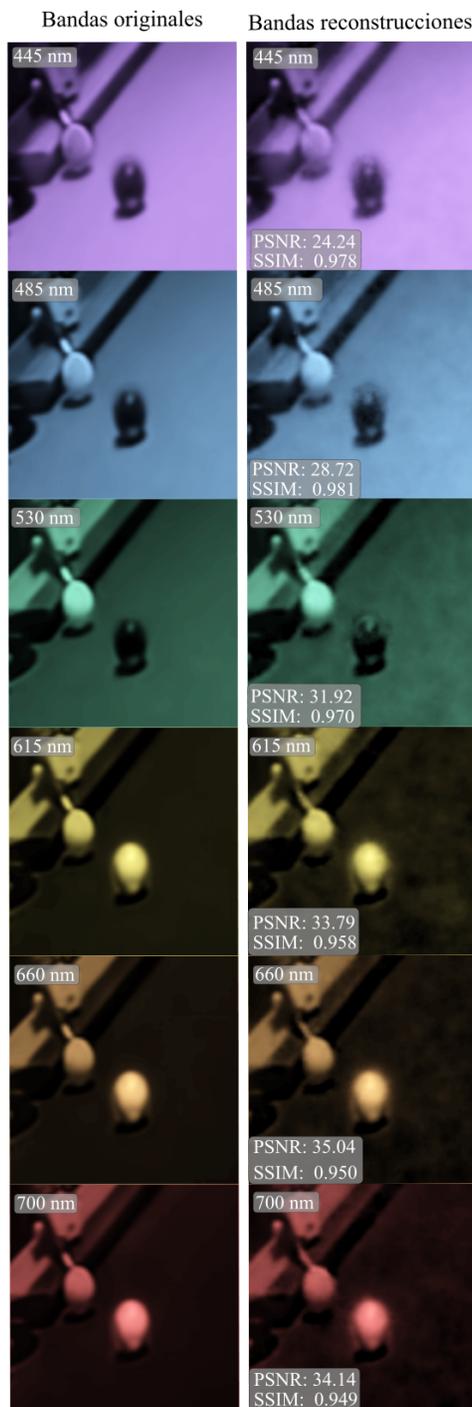


Figura. 47. Resultados espectrales del SV con la escena de un juego de billar en la que se observan 6 bandas espectrales para la escena original y la reconstrucción. Se incluyen las métricas de PSNR y SSIM con las longitudes de onda de cada banda correspondiente.

columna corresponde a las reconstruidas.

De los resultados de reconstrucción espectral en la figura 47, se puede notar que las bandas que se encuentran en las longitudes de onda más cortas obtienen resultados de PSNR menores con 24.24 dB en los 445 nm mientras que para el SSIM se obtienen mayores valores, llegando a 0.981 en los 485 nm. En comparación, las longitudes de onda más largas obtienen valores mayores de PSNR con 35.04 dB en los 660 nm y menores valores en SSIM llegando a 0.949 para la misma longitud de onda. Además, se puede observar que para el caso de la bola roja de billar, se presentan mayores intensidades en el rango de los 600 nm como era de esperarse debido a su color, a su vez, para el tapiz azul se observan que las mayores intensidades están en las longitudes de ondas del espectro azul nuevamente como se esperaba, mostrando una correcta estimación de los valores espectrales en la imagen.

Adicionalmente, se presenta un análisis de las respuestas espectrales de la reconstrucción en la figura 45 para el fotograma 4 con el objetivo de analizar el rendimiento espectral del método propuesto. Para esto, se escogieron tres regiones de interés (ROI, de sus siglas en inglés region of interest) R_1 , R_2 y R_3 , donde cada una corresponde a una grilla de 4×4 píxeles que es promediada para obtener un único valor de intensidad por banda. Se utilizan diferentes ROI para cada escena en las que se incluye la métrica SAM obtenida, obteniendo los resultados mostrados en la figura 48.

Donde se observa que en la firma espectral de la región R_1 se obtiene una mayor intensidad en las longitudes de onda más grandes, esto debido a que su riqueza espectral se encuentra en los colores rojos para rangos mayores de 600 nm. Además, puede notarse que la firma se ajusta de

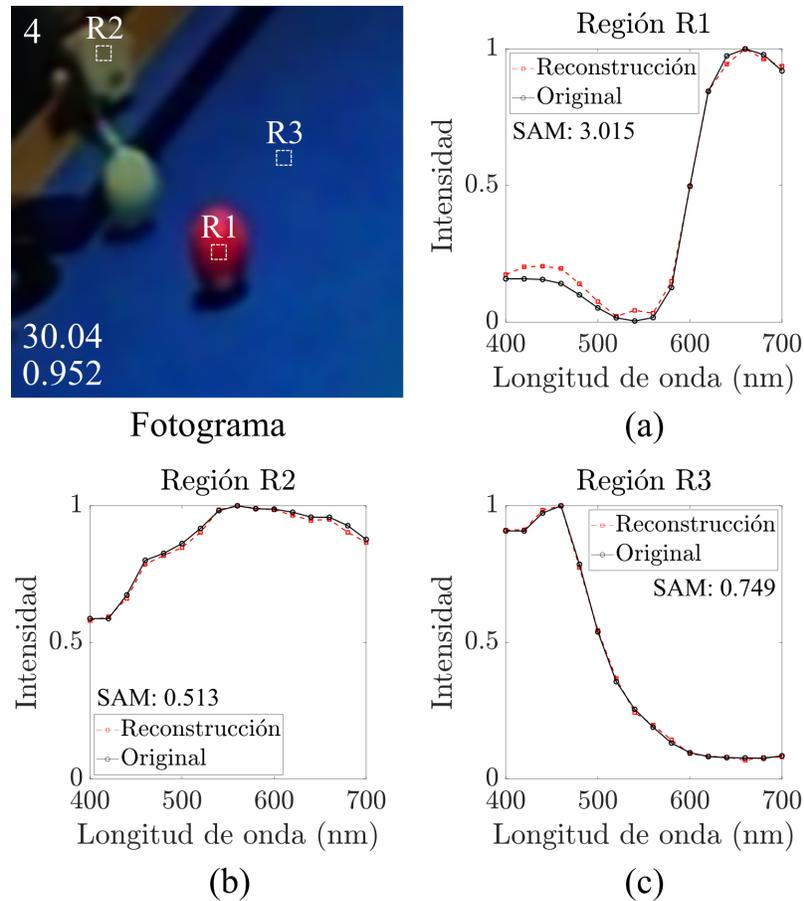


Figura. 48. Respuesta espectral de 3 regiones de interés de un fotograma espectral para la escena de un SV de un juego de billar. Se muestra la firma espectral original con su reconstrucción y la métrica SAM.

manera correcta para estas longitudes de onda más grandes mientras que para las longitudes de onda más pequeñas difiere en algunos de sus puntos, obteniendo un SAM de 3.015. Asimismo, en la región R_2 se estima una firma espectral muy similar con la firma original, se puede notar como la curva describe un comportamiento igual a la original por lo que se obtiene un SAM bajo, de 0.513. Finalmente, para la última región denominada R_3 se percibe una firma espectral con mayor intensidad en el rango de los 450 nm, debido a que el punto escogido toma el tapiz de la mesa de



Figura. 49. Resultados espectrales del SV con la escena de un balón de fútbol en la que se observan 6 bandas espectrales para la escena original y la reconstrucción. Se incluyen las métricas de PSNR y SSIM con las longitudes de onda de cada banda correspondiente.

billar, de color azul. Una vez más los resultados son altamente fiables debido a que la estimación se ajusta de forma similar a la firma original, obteniendo un SAM de 0.749.

Al igual que con la SV de las bolas de billar presentado anteriormente, se presentarán los resultados obtenidos en la reconstrucción espectral para el fotograma 26 del SV de la escena del balón en la figura 49. Nuevamente se tomaron 6 de las 16 bandas espectrales en el rango de 400 a 700 nm de forma equiespaciada y se usaron las métricas de PSNR y SSIM en donde la primera columna corresponde a las bandas espectrales originales y la segunda columna corresponde a las reconstruidas.

Para la reconstrucción espectral de la escena se obtienen resultados notablemente buenos, se puede notar que la respuesta a lo largo de todo el espectro del fotograma es uniforme, obteniendo entre 27.30 dB y 27.58 dB para el PSNR, así como de 0.948 a 0.958 en SSIM. Además, se puede notar en la figura 49 que las intensidades en cada una de las bandas muestran un comportamiento adecuado, por ejemplo, para el muro anaranjado de la parte superior derecha de la escena se puede observar que sus intensidades son bajas para las primeras bandas mientras que para las longitudes de onda cercanas al rojo son mayores. De la misma manera, el pantalón de la persona tiene color azul y puede notarse más claro para las longitudes de onda cercanas a los 400 nm mientras que para las más lejanas se oscurece, lo que muestra una correcta estimación de los espectros.

Adicionalmente, se presenta un análisis de las respuestas espectrales de la reconstrucción en la figura 46 para el fotograma 25 con el objetivo de analizar la eficiencia espectral del método propuesto. Para esto, se escogieron tres ROI, R_1 , R_2 y R_3 , donde al igual que con el SV anteriormente presentado, cada región corresponde a una grilla de 4×4 píxeles que es promediada para obtener

un único valor de intensidad por banda. Asimismo, utilizan diferentes ROI para cada escena en las que se incluye la métrica SAM obtenida, obteniendo los resultados mostrados en la figura 50.

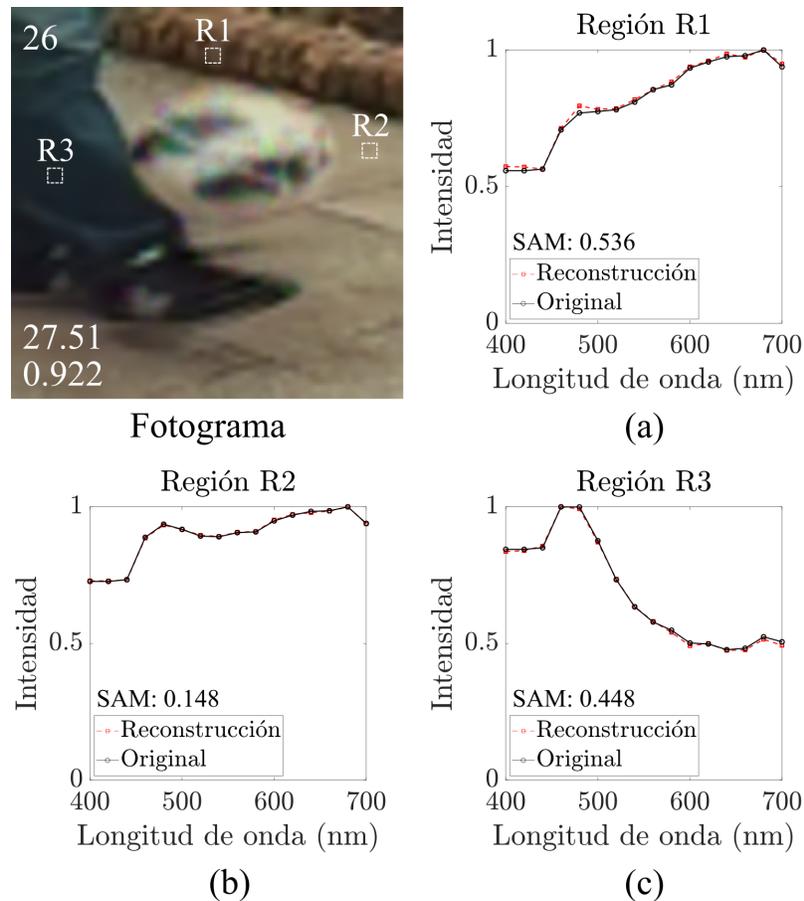


Figura. 50. Respuesta espectral de 3 regiones de interés de un fotograma espectral para la escena de un SV de un balón de fútbol. Se muestra la firma espectral original con su reconstrucción y la métrica SAM.

Destacando que en la firma espectral de la región R_1 se obtiene una tendencia de mayor intensidad hacia las longitudes de onda más grandes, esto debido a que su riqueza espectral tiende a enfocarse en los colores rojos para rangos mayores de 600 nm, a pesar de que no sea perceptible a simple vista, obteniendo un SAM de 0.536. Igualmente, en la región R_2 se estima una firma

espectral idéntica con la firma original dado que las dos describen un comportamiento similar, por lo que se obtiene un SAM bajo de 0.513. Por último, para la región denominada R_3 se percibe una firma espectral con mayor intensidad en el rango de los 450-500 nm, debido a que el punto escogido toma el pantalón del jugador el cual es de color azul. Nuevamente se reafirma que los resultados son altamente fiables debido a que la estimación se ajusta de forma semejante a la firma original, obteniendo un SAM de 0.448.

5.4.4. Análisis para capturas experimentales de SV. Teniendo en cuenta que se realizó la captura de vídeos espectrales a partir del sistema optoelectrónico sincronizado en el laboratorio, expuesto en el capítulo 3, se llevó a cabo la tarea de reconstrucción espectro-temporal usando la metodología propuesta con el objetivo de obtener una comparación del rendimiento y comportamiento de este método para los vídeos espectrales obtenidos en la sección 5.1, como para los vídeos espectrales capturados experimentalmente.

En la figura 51 se exponen los resultados adquiridos para la reconstrucción de un SV con una escena de un lego girando para la primera fase la cual es de estimación temporal. Por dicha razón, en la primera fila se presenta la escena original en escala de grises y en la segunda se encuentra la escena reconstruida para los fotogramas 5, 13, 21 y 29 de manera que pueda apreciarse el movimiento espacial de la misma. Además de esto, en la parte inferior izquierda de cada fotograma reconstruido se encuentran los valores de PSNR y SSIM obtenidos con el objetivo de realizar un análisis cuantitativo de la calidad de la estimación.

Donde se puede observar una difuminación en cada fotograma, esto ocurre debido a que a diferencia de las escenas anteriormente presentadas en esta existe mayor dinámica debido a que

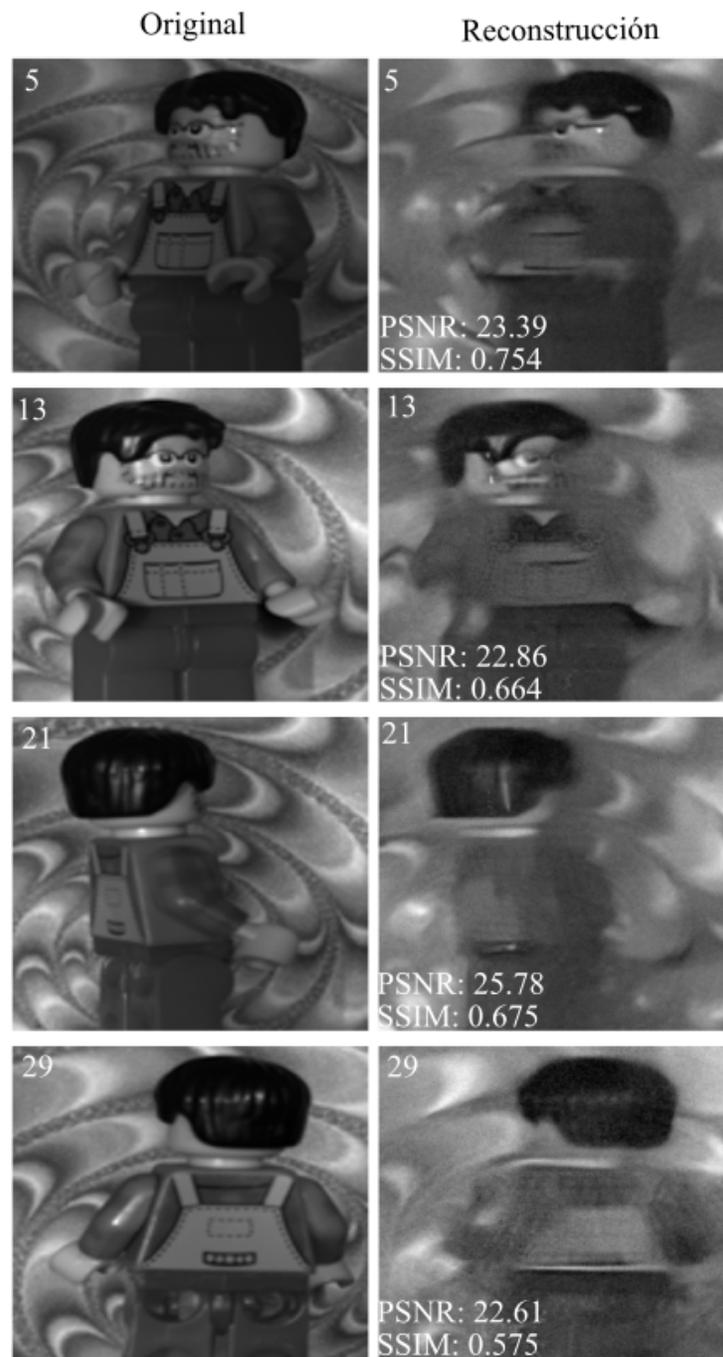


Figura. 51. Resultados para un SV capturado experimentalmente de un Lego girando en donde se aprecian los fotogramas originales y la reconstrucción temporal en escala de grises. Se incluyen las métricas de PSNR y SSIM con los números de cada fotograma correspondiente.

el fondo se encuentra en movimiento al igual que el objeto principal, a pesar de esto se logran diferenciar pequeños detalles del lego como su traje y su cara, y en general la dimensión temporal se estima razonablemente. Adicionalmente, los valores de PSNR resultan adecuados, obteniendo entre 22.61 dB a 25.78 dB mientras que los resultados de SSIM son menores, de 0.575 a 0.754, debido sus altas características dinámicas de la escena como lo es la rotación del fondo.

Continuando con el análisis temporal, en la figura 52 se presentan resultados espectro-temporales para el SV reconstruido en su representación RGB utilizando los mismos fotogramas de la figura 51, utilizando como métricas el SSIM y SAM, obteniendo resultados satisfactorios en el rendimiento del método. Al igual que para la estimación grayscale, en la figura 52 se observa como se recupera correctamente a grandes rasgos el movimiento de la escena entre los fotogramas incluyendo nuevamente el difuminado creado por la complejidad de la escena, pero aun así reconstruyendo pequeños detalles como la cara del lego y el traje, asimismo se mantienen características como la forma y color del lego y el fondo. Adicionalmente, es remarcable la reconstrucción de la parte posterior de la escena incluso apreciándose mejor resultado que en el lego. Para las métricas de SSIM y SAM se alcanzan los 0.893 y 21.77 en estos fotogramas, respectivamente, en comparación con la figura 51 se puede observar una mejoría notoria para la métrica SSIM que había sido menor.

Igualmente, para enriquecer el análisis de los resultados presentados anteriormente, se estudiarán espectralmente las reconstrucciones obtenidas para un fotograma del SV recuperado. Para esto, se tomaron 6 de las 16 bandas espectrales en el rango de 400 a 700 nm de forma equidistante y se usaron las métricas de PSNR y SSIM, de manera que se pueda apreciar y establecer la calidad

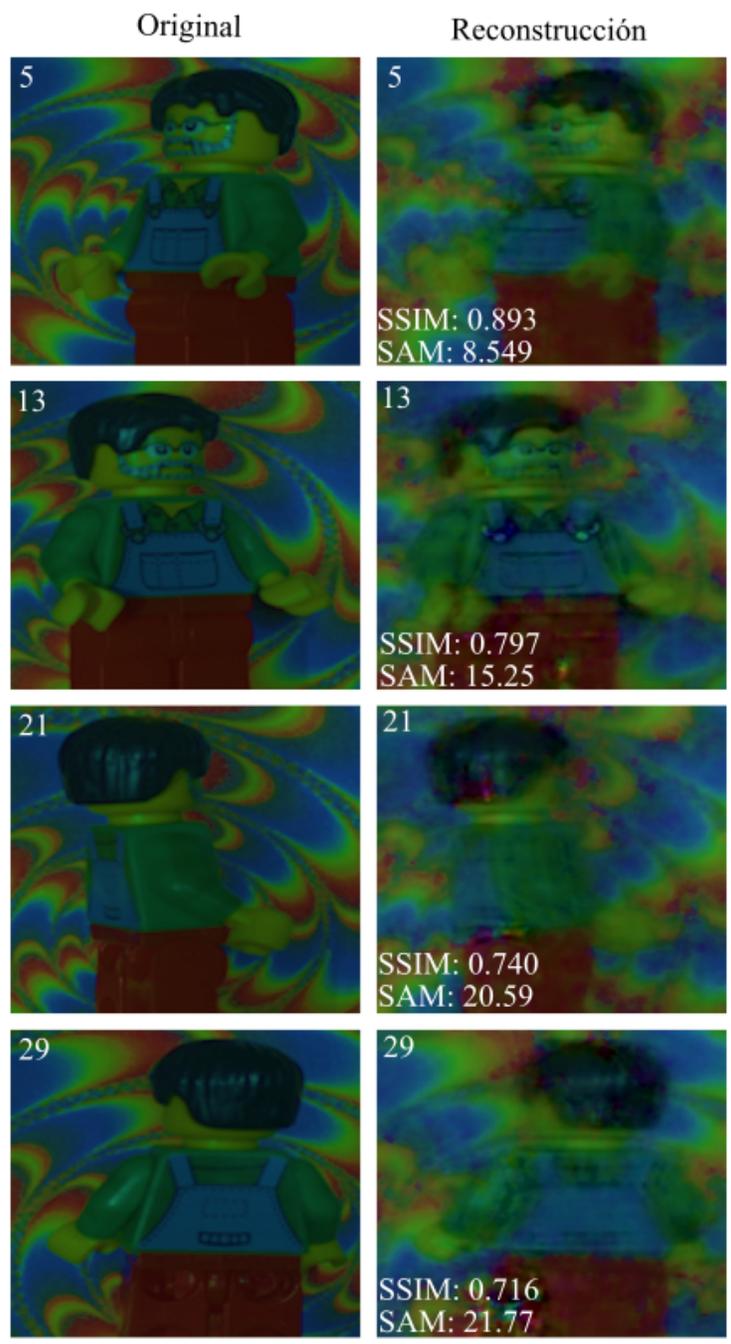


Figura. 52. Resultados espectro-temporales para un SV capturado experimentalmente de un Lego girando en donde se muestran los fotogramas originales y la reconstrucción temporal en representación RGB. Se incluyen las métricas de SSIM y SAM con los números de cada fotograma correspondiente.

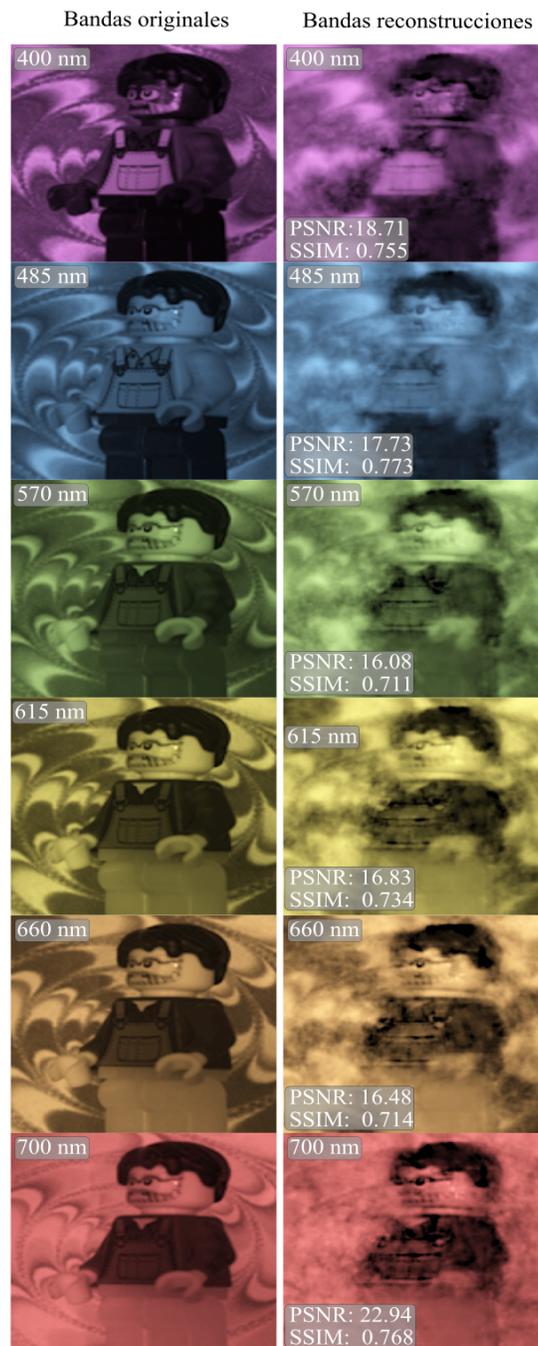


Figura. 53. Resultados espectrales del SV con la escena de un lego girando en la que se observan 6 bandas espectrales para la escena original y la reconstrucción. Se incluyen las métricas de PSNR y SSIM con las longitudes de onda de cada banda correspondiente.

de la estimación de la respuesta espectral del método propuesto en este tipo de SV. La figura 53 muestra los resultados obtenidos para el fotograma 5 de la figura 52 en donde la primera columna corresponde a las bandas espectrales del SV original y la segunda columna corresponde a las reconstruidas.

De los resultados de reconstrucción espectral en la figura 53, es resaltable que las bandas que se encuentran en las longitudes de onda extremas obtienen los resultados de PSNR mas altos con 18.71 dB en los 400 nm y 22.94 dB en los 700 nm, igualmente, para el SSIM se obtienen mayores valores, llegando a 0.755 en los 400 nm y 0.768 en los 700 nm, remarcando que el mayor de SSIM se adquirió en los 485 nm con un valor de 0.773. Además, se puede observar de forma cualitativa que a pesar de la complejidad espectral de la escena se obtiene una reconstrucción aceptable, por ejemplo, se aprecia como la respuesta espectral en cada color es adecuada, para las primeras bandas las mayores intensidades se concentran en el chaleco de la figura, luego, para las longitudes de onda intermedias se ven intensidades más grandes en los brazos del personaje y finalmente para las longitudes de onda del espectro rojo se observan mayores intensidades en los pantalones del Lego. Esto permite que al momento de combinar todas las bandas en un fotograma se obtenga la reconstrucción espectral que se observa en la figura 52.

Adicionalmente, para complementar el análisis temporal se presenta un análisis de las respuestas espectrales de la reconstrucción en la figura 52 para el fotograma 5 con el objetivo de analizar la eficiencia espectral del método propuesto en SV emulados en el laboratorio. Para esto, se escogieron tres ROI, R_1 , R_2 y R_3 , donde al igual que con los SV anteriormente presentados, cada región corresponde a una grilla de 4×4 píxeles que es promediada para obtener un único valor de

intensidad por banda. Asimismo, se posicionan en diferentes ROI para cada escena en las que se incluye la métrica SAM obtenida, obteniendo los resultados mostrados en la figura 54.

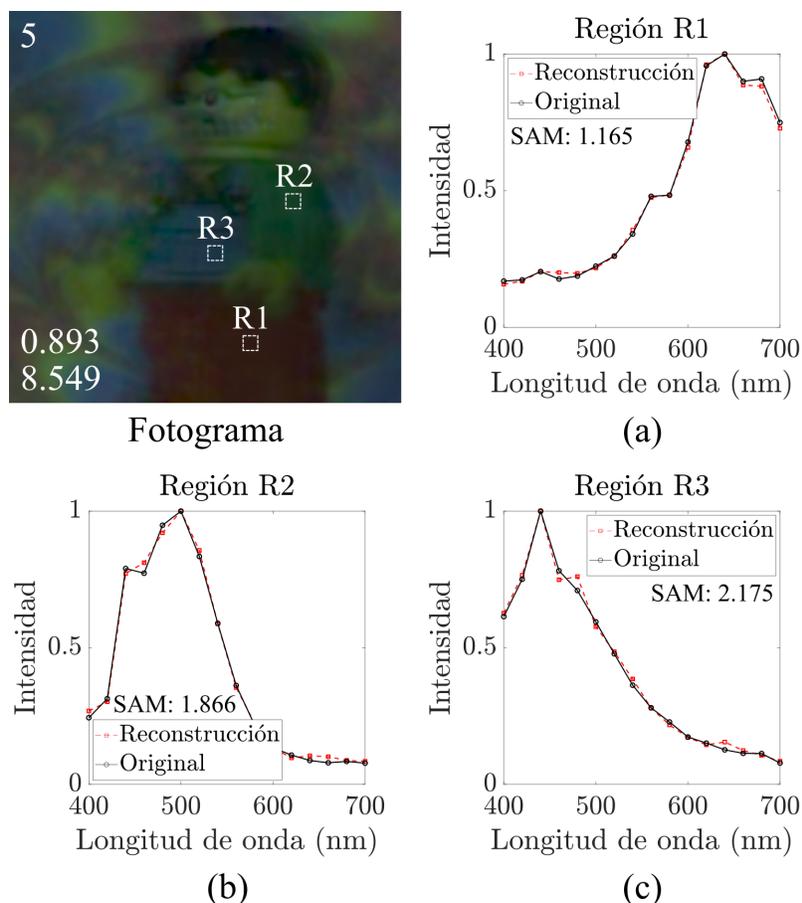


Figura. 54. Respuesta espectral de 3 regiones de interés de un fotograma espectral para la escena de un SV de un Lego girando. Se muestra la firma espectral original con su reconstrucción y la métrica SAM.

A través de la figura 54, se observa que en la firma espectral de la región R_1 se adquiere una mayor intensidad en las longitudes de onda más grandes, debido a que la región escogida corresponde al pantalón de la figura de lego, lo que muestra una correcta estimación obteniendo un SAM de 1.165. Igualmente, en la región R_2 se estima una firma espectral muy similar con la firma

original, aquí, los valores de intensidad se concentran en torno a las longitudes de onda cerca a los 500 nm como es de esperarse, debido a que se seleccionó una zona de color verde, obteniendo un SAM de 1.866. Finalmente, para la última región denominada R_3 se obtiene una firma espectral con mayor intensidad en el rango de los 450 nm, debido a que el punto escogido toma el chaleco del lego el cual posee un color azul y obteniendo un SAM de 2.175 debido a algunos puntos de la firma que no se ajustaron correctamente a las intensidades. En conclusión, la metodología de adquisición y reconstrucción obtiene resultados aceptables para SV capturados experimentalmente, incluso cuando contienen una escena de alta complejidad, con formas irregulares, bastante movimiento entre fotogramas y múltiples colores.

5.4.5. Análisis Comparativo del método propuesto. Para finalizar el análisis del desempeño del método propuesto se realizó la reconstrucción de los 10 SV usados en la sección 5.3 usando el método tradicional en la que toda la información se comprime en un solo disparo y el SV se estima usando un PNP tradicional. Para esto, se utilizaron los mismos parámetros de resolución con $N_x = 1024$, $N_y = 1024$ espacialmente, $N_\lambda = 16$ como el número de bandas y $N_t = 32$ para los fotogramas comprimidos en la medida. Los resultados obtenidos se muestran en la tabla 1 en donde se compara el promedio alcanzado en las métricas de PSNR, SSIM y SAM para cada uno de los métodos.

De acuerdo con los datos expuestos, es importante resaltar que el método propuesto obtiene resultados superiores en cada una de las métricas usadas en la comparación. Específicamente, para el PSNR se obtiene un promedio de 26.1726 dB frente a 19.7482 dB obtenidos por el método tradicional, además, para el SSIM se alcanza un 0.89061 en nuestra metodología mientras que

	Tradicional	Propuesto	Ganancia
PSNR	19.7482	26.1726	+ 6.4244
SSIM	0.69392	0.89061	+ 0.1967
SAM	15.2367	4.04503	- 11.192

Tabla 1. Comparación de los resultados obtenidos para la reconstrucción de 10 SV usando el método tradicional y la metodología propuesta. Se usan los promedios alcanzados para las métricas de PSNR, SSIM y SAM y se muestra la ganancia que se obtiene en cada una de las métricas al usar el enfoque propuesto.

de la forma tradicional se llega a tan sólo 0.69392, finalmente, se obtiene una gran mejoría en el SAM con 4.04503 en comparación con 15.2367 entre los métodos presentados. En conclusión, el rendimiento de la metodología de adquisición presentada en este trabajo es capaz de superar los métodos tradicionales por más de 6 dB en PSNR, al rededor de 0.2 en SSIM y logra reducir el SAM en más de 11 grados, mostrando un desempeño destacable.

6. Trabajo Futuro

En este trabajo se propuso un nuevo sistema óptico para la adquisición de medidas comprimidas de videos espectrales en el que se utilizan diversos elementos ópticos como un DMD, una cámara CCD y un filtro tuneable como modulador de luz. Por esto, sería interesante emplear el método propuesto para la reconstrucción de medidas obtenidas experimentalmente en un sistema real implementado en un laboratorio.

Además, la red neuronal FFDNet usada puede ser entrenada con un conjunto de datos propios que sean adquiridos en el laboratorio para incrementar el rendimiento de la labor de reconstrucción en el problema específico, así como puede adaptarse para que en su salida se obtenga una estimación espectro-temporal en lugar de una espacio-temporal.

7. Conclusiones

Se propuso una metodología para la adquisición y reconstrucción de vídeo espectral compresivo basada en un enfoque de ventana espacio-temporal basado en un sistema óptico compuesto principalmente por un filtro recursivo y un DMD que permite generar codificaciones de apertura de color variantes en el tiempo y usar un único sensor. Además, se presentó la combinación de diversas técnicas de reconstrucción como el uso de un algoritmo plug and play que incluye en su marco de trabajo una red rápida y flexible y la variación total para la eliminación de ruido junto con un plug and play tradicional alimentado por la salida del método anterior que permitió obtener resultados destacables.

Adicionalmente, el método propuesto fue probado sobre un conjunto de datos de escenarios del mundo real y capturas experimentales realizadas en un laboratorio a través de la sincronización de múltiples elementos optoelectrónicos. Mediante las pruebas de reconstrucción en los vídeos simulados y capturados se muestra que es posible estimar una escena de vídeo espectral de forma fiable cuando aproximadamente solo el 0,78 % de la información de la escena se toma para $N_x = 1024$, $N_y = 1024$, $N_\lambda = 16$ y $N_t = 32$ con un tamaño de ventana de $S = 2$ y un código de apertura aleatorio con 50 % de transmitancia. Los resultados obtenidos demuestran que la metodología propuesta presenta un rendimiento de estimación espectro-temporal robusto incluso en presencia de 20 dB de ruido, además, puede estimar escenas de alta complejidad espectro temporal logrando reconstruir detalles finos y bordes así como el movimiento de las mismas.

El desarrollo de este trabajo permitió la presentación de un artículo de conferencia que

fue aceptado y sustentado en el XXIII “Simposio de Imagen, Procesamiento de Señales y Visión Artificial” (STSIVA 2021).

8. Apéndice

Apéndice A. Función de Sincronización de elementos optoelectronicos.

```

1 L = 2;
2 T = 100;
3 wave_lits = floor(linspace(450,700,L));
4 dataset = zeros(776, 1032, size(wave_lits,2));
5 %%
6 tic
7 Ct = 0;
8 for t=0:180/T:180
9     Ct = Ct + 1;
10    for w = 1:length(wave_lits)
11        %% Barrido de espectro, Monocromador
12        comand = ['GOWAVE ' num2str(wave_lits(w))];
13        fprintf(s, comand)
14        pause(1)
15        for Rt = 1:5
16            %% Captura de fotograma
17            trigger(vid1);
18            pause(1)
19            Cx(:, :, Rt) = squeeze(getdata(vid1));
20        end
21        dataset(:, :, w) = mean(Cx(:, :, 2:end), 3);

```

```
22     imagesc(dataset(:,:,w)); colorbar; title(num2str(wave_lits(w)))
23     colormap(gray)
24     drawnow
25     save(int2str(Ct), 'dataset', '-v7.3')
26 end
27 ArduinoFunc(t)
28 end
29
```

Apéndice B. Función de Sincronización de Servomotor al Sistema.

```
1 function ArduinoFunc(step)
2     Tcap = 0.5; %%Tiempo para rotacion de la escena.
3     a = arduino('COM7', 'Uno');
4     conex=servo(a, 'D4');
5     writePosition(conex, step);
6     pause(Tcap);
7 end
8
```

Referencias Bibliográficas

- and Simulation, D. (2020). Design-and-simulation/cacti-simulation: Simulation for cacti (coded aperture compressive temporal imaging) system.
- Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S. (2013). Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115.
- Area Tecnología (2020). Servomotores 360° - [noviembre, 2021].
- Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic press.
- Bi, X., Wei, Y., Xiao, B., and Li, W. (2019). Rru-net: The ringed residual u-net for image splicing forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0.
- Buades, A., Coll, B., and Morel, J.-M. (2011). Non-local means denoising. *Image Processing On Line*, 1:208–212.
- Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Carin, L., and Brady, D. J. (2016). Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108.
- Chambolle, A., Caselles, V., Cremers, D., Novaga, M., and Pock, T. (2010). An introduction to

- total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, pages 263–340. de Gruyter.
- Chang, C.-I. (2016). Hyperspectral target detection. In *Real-Time Progressive Hyperspectral Image Processing*, pages 131–172. Springer.
- Chen, J. G., Shah, V., and Liu, L. (2021). Performance of a u-net-based neural network for predictive adaptive optics. *Optics Letters*, 46(10):2513–2516.
- Control Automático Educación (2020). Como usar un servomotor con arduino (180° y 360°) - [febrero, 2022].
- Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2006). Image denoising with block-matching and 3d filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, volume 6064, page 606414. International Society for Optics and Photonics.
- Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095.
- Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306.
- Gastal, E. S. and Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. In *ACM SIGGRAPH 2011 papers*, pages 1–12. Association for Computing Machinery.

- Gat, N. (2000). Imaging spectroscopy using tunable filters: a review. In *Wavelet Applications VII*, volume 4056, pages 50–64. International Society for Optics and Photonics.
- Gehm, M. E., John, R., Brady, D. J., Willett, R. M., and Schulz, T. J. (2007). Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 15(21):14013–14027.
- Gupta, R. and Hartley, R. I. (1997). Linear pushbroom cameras. *IEEE Transactions on pattern analysis and machine intelligence*, 19(9):963–975.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.
- Huang, G., Jiang, H., Matthews, K., and Wilford, P. (2013). Lensless imaging by compressive sensing. In *2013 IEEE International Conference on Image Processing*, pages 2101–2105. IEEE.
- Huang, J. and Nowack, R. L. (2020). Machine learning using u-net convolutional neural networks for the imaging of sparse seismic data. *Pure and Applied Geophysics*, 177(6):2685–2700.
- Hwang, S., Park, J., Kim, N., Choi, Y., and Kweon, I. S. (2015). Multispectral pedestrian detection: Benchmark dataset and baselines. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Jiao, L., Huo, L., Hu, C., and Tang, P. (2020). Refined unet: Unet-based refinement network for cloud and shadow precise segmentation. *Remote Sensing*, 12(12).
- Kiani Galoogahi, H., Fagg, A., Huang, C., Ramanan, D., and Lucey, S. (2017). Need for speed:

- A benchmark for higher frame rate object tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1125–1134.
- Leitner, R., De Biasio, M., Arnold, T., Dinh, C. V., Loog, M., and Duin, R. P. (2013). Multi-spectral video endoscopy system for the detection of cancerous tissue. *Pattern Recognition Letters*, 34(1):85–93.
- Liu, T., Li, Y.-F., Liu, H., Zhang, Z., and Liu, S. (2019a). Risir: Rapid infrared spectral imaging restoration model for industrial material detection in intelligent video systems. *IEEE Transactions on Industrial Informatics*.
- Liu, Z., Cao, Y., Wang, Y., and Wang, W. (2019b). Computer vision-based concrete crack detection using u-net fully convolutional networks. *Automation in Construction*, 104:129–139.
- Llull, P., Liao, X., Yuan, X., Yang, J., Kittle, D., Carin, L., Sapiro, G., and Brady, D. J. (2013). Coded aperture compressive temporal imaging. *Optics express*, 21(9):10526–10545.
- Marquez, M., Meza, P., Arguello, H., and Vera, E. (2019). Compressive spectral imaging via deformable mirror and colored-mosaic detector. *Optics express*, 27(13):17795–17808.
- Monno, Y., Teranaka, H., Yoshizaki, K., Tanaka, M., and Okutomi, M. (2018). Single-sensor rgb-nir imaging: High-quality system design and prototype implementation. *IEEE Sensors Journal*, 19(2):497–507.
- Nassif, N., Cense, B., Park, B., Pierce, M., Yun, S., Bouma, B., Tearney, G., Chen, T., and De Boer,

- J. (2004). In vivo high-resolution video-rate spectral-domain optical coherence tomography of the human retina and optic nerve. *Optics express*, 12(3):367–376.
- Qiu, H., Wang, Y., and Meng, D. (2021). Effective snapshot compressive-spectral imaging via deep denoising and total variation priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9127–9136.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Rueda, H., Arguello, H., and Arce, G. R. (2015). Dmd-based implementation of patterned optical filter arrays for compressive spectral imaging. *JOSA A*, 32(1):80–89.
- Scebba, G., Da Poian, G., and Karlen, W. (2020). Multispectral video fusion for non-contact monitoring of respiratory rate and apnea. *IEEE Transactions on Biomedical Engineering*, 68(1):350–359.
- Strong, D. and Chan, T. (2003). Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165.
- Talo, M. (2019). Convolutional neural networks for multi-class histopathology image classification. *ArXiv*, abs/1903.10035.
- Tsai, T.-H., Llull, P., Yuan, X., Carin, L., and Brady, D. J. (2015). Spectral-temporal compressive imaging. *Optics letters*, 40(17):4054–4057.

- Vane, G., Green, R. O., Chrien, T. G., Enmark, H. T., Hansen, E. G., and Porter, W. M. (1993). The airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 44(2-3):127–143.
- Wagadarikar, A., John, R., Willett, R., and Brady, D. (2008a). Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51.
- Wagadarikar, A. A., Pitsianis, N. P., Sun, X., and Brady, D. J. (2008b). Spectral image estimation for coded aperture snapshot spectral imagers. In *Image Reconstruction from Incomplete Data V*, volume 7076, page 707602. International Society for Optics and Photonics.
- Wang, Y. W., Reder, N. P., Kang, S., Glaser, A. K., and Liu, J. T. (2017). Multiplexed optical imaging of tumor-directed nanoparticles: a review of imaging systems and approaches. *Nanotheranostics*, 1(4):369.
- Wu, Y., Mirza, I. O., Arce, G. R., and Prather, D. W. (2011a). Demonstration of a dmd-based compressive sensing (cs) spectral imaging system. In *CLEO: Applications and Technology*, page ATuE2. Optical Society of America.
- Wu, Y., Mirza, I. O., Arce, G. R., and Prather, D. W. (2011b). Development of a digital micromirror device (dmd)-based snapshot spectral imaging (dmd-ssi) system. In *IEEE Photonic Society 24th Annual Meeting*, pages 149–150. IEEE.
- Wu, Y., Mirza, I. O., Arce, G. R., and Prather, D. W. (2011c). Development of a digital micromirror

- device (dmd)-based snapshot spectral imaging (dmd-ssi) system. In *IEEE Photonic Society 24th Annual Meeting*, pages 149–150. IEEE.
- Wu, Y., Wu, J., Jin, S., Cao, L., and Jin, G. (2021). Dense-u-net: Dense encoder–decoder network for holographic imaging of 3d particle fields. *Optics Communications*, 493:126970.
- Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155.
- Zhang, K., Zuo, W., and Zhang, L. (2018a). Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622.
- Zhang, Z., Liu, Q., and Wang, Y. (2018b). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753.
- Zhao, Y., Po, L.-M., Yan, Q., Liu, W., and Lin, T. (2020). Hierarchical regression network for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423.