

**ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA**

**TRABAJO FIN DE GRADO**

**SISTEMA DE CODIFICACIÓN ESTEGANOGRÁFICO PARA COMUNICACIONES OCC**

**Titulación:** Grado en Ingeniería en Tecnologías de la Telecomunicación

**Autor:** Alberto Ramos Monagas

**Tutores:** Dr. D. Rafael Pérez Jiménez

Dr. Víctor Guerra Yánez

**Fecha:** Julio de 2020



**ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA**

**TRABAJO FIN DE GRADO**

**SISTEMA DE CODIFICACIÓN ESTEGANOGRÁFICO PARA COMUNICACIONES OCC**

**HOJA DE FIRMAS**

**Alumno/a**

**Tutor/a Tutor/a**

**Fecha**: Julio de 2020



**ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA**

**TRABAJO FIN DE GRADO**

**SISTEMA DE CODIFICACIÓN ESTEGANOGRÁFICO PARA COMUNICACIONES OCC**

**HOJA DE EVALUACIÓN**

**Calificación: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Presidente**

**Vocal Secretario/a**

**Fecha**: Julio de 2020

**RESUMEN**

Los recientes esfuerzos en comunicaciones de luz visible a través de enlaces de screen-camera han explotado los recursos de la pantalla para la comunicación de datos. Esas prácticas, aunque convenientes, han dado lugar a controversias entre el espacio asignado a los usuarios y el contenido reservado a los dispositivos, además de su falta de estética visual y de distracción. En este documento proponemos un sistema que permite la comunicación simultánea, de modo dual (entre la cámara y la pantalla), tanto para los usuarios como para los dispositivos sin mostrar ninguna imagen codificada (por ejemplo, códigos de barras). Nuestro sistema codifica los datos en píxeles translúcidos sobre cualquier tipo de contenido de la pantalla utilizando el canal alfa, de modo que cualquier dispositivo equipado con una cámara puede obtener los datos volteando su cámara hacia la pantalla. Además, el aprovechamiento del canal alfa es un concepto muy reconocido por los gráficos por ordenador, ya que se utiliza para codificar los bits en el cambio de translucidez de los píxeles. Al eliminar la necesidad de modificar directamente los valores RGB de los píxeles se permite una comunicación discreta en tiempo real, a la vez que soporta cualquier tipo de contenido en la pantalla. En definitiva, el sistema diseñado aprovecha la propiedad de fusión de parpadeo espaciotemporal del sistema de visión humana y la rápida velocidad de frame de la pantalla moderna. Multiplexa los datos en el contenido del vídeo, fotograma a fotograma, a través de una composición de fotogramas complementarios y una modulación similar a la de la CDMA (del inglés Code Division Multiple Access), basada en técnicas de espectro ensanchado. De este modo se garantiza una comunicación de datos discreta entre la cámara de y la pantalla sin que ello afecte a la experiencia de visualización de los usuarios.

Finalmente, se construyó un modelo utilizando dispositivos inteligentes disponibles en el mercado actual y demostramos su eficacia y robustez en diferentes entornos prácticos.

**PALABRAS CLAVES**

Comunicación pantalla-cámara; Comunicación oculta visible; Comunicación con luz visible; Canal alfa; Comunicación visible de modo dual; Vídeo de fotograma completo; Comunicación por cámara óptica.

**ABSTRACT**

Recent efforts in visible light communications through screen-camera links have exploited the resources of the screen for data communication. These practices, while convenient, have led to controversies between the space assigned to users and the content reserved for the devices, in addition to their lack of visual aesthetics and distraction. In this document we propose a system that allows simultaneous, dual-mode communication (between the camera and the screen) for both users and devices without displaying any encoded images (e.g., bar codes). Our system encodes the data in translucent pixels on any type of screen content using the alpha channel, so that any device equipped with a camera can obtain the data by turning its camera towards the screen. In addition, the use of the alpha channel is a concept well recognized by computer graphics, since it is used to encode the bits in the change of translucency of the pixels. Eliminating the need to directly modify the RGB values of the pixels allows for discreet real-time communication, while supporting any type of content on the screen. In short, the designed system takes advantage of the space-time flicker fusion property of the human vision system and the fast frame rate of the modern screen. It multiplexes the data in the video content, frame by frame, through a composition of complementary frames and a modulation similar to CDMA (Code Division Multiple Access), based on spread spectrum techniques. This ensures discreet data communication between the camera and the screen without affecting the user's viewing experience. Finally, a model was built using intelligent devices available on the market today and we demonstrated their effectiveness and robustness in different practical environments.

**KEY WORDS**

Screen-camera communication; Hidden visible communication; Visible ligth communication; Alpha cannel; Dual-mode visible communication; Full-frame video; Optical Camera Communication.

Índice

[CAPÍTULO 1: INTRODUCCIÓN 9](#_Toc44234972)

[1.1. Antecedentes 9](#_Toc44234973)

[1.2. Justificación e importancia 10](#_Toc44234974)

[1.3. Objetivos y alcance del proyecto 11](#_Toc44234975)

[1.4. Medios utilizados 12](#_Toc44234976)

[1.5. Organización de la memoria 12](#_Toc44234977)

[CAPÍTULO 2: EFECTOS DE LA PERCEPCIÓN VISUAL HUMANA 14](#_Toc44234978)

[2.1. Introducción 14](#_Toc44234979)

[2.2. Espectro visible 15](#_Toc44234980)

[2.3. El ojo humano 16](#_Toc44234981)

[2.4. El sistema visual humano (SVH) 18](#_Toc44234982)

[2.5. El proceso de la percepción visual 19](#_Toc44234983)

[2.6. Mecanismos básicos de la visión 21](#_Toc44234984)

[2.7. El proceso cognitivo y la corteza visual 22](#_Toc44234985)

[2.8. Enmascaramiento visual 23](#_Toc44234986)

[2.9. Percepción del movimiento 24](#_Toc44234987)

[2.10. Percepción del parpadeo de la visión humana (flicker) 25](#_Toc44234988)

[2.11. Factores que determinan la sensibilidad al contraste 27](#_Toc44234989)

[2.12. Sensibilidad temporal 29](#_Toc44234990)

[2.13. Agudeza visual y visión periférica 30](#_Toc44234991)

[2.14. Pantallas y cámaras modernas 31](#_Toc44234992)

[2.15. Comparación entre cámara y ojo 32](#_Toc44234993)

[CAPÍTULO 3: SISTEMAS ESTEGANOGRÁFICOS 32](#_Toc44234994)

[3.1. Introducción 32](#_Toc44234995)

[3.2. Esteganografía en la era digital 34](#_Toc44234996)

[3.3. Estegosistema: descripción y características 35](#_Toc44234997)

[3.4. Introducción a los métodos y técnicas esteganográficas empleadas 39](#_Toc44234998)

[3.5. Técnicas de espectro ensanchado para la esteganografía con imágenes (SSIS) 41](#_Toc44234999)

[CAPÍTULO 4: SISTEMAS DE TRANSMISIÓN SCREEN2CAMERA 43](#_Toc44235000)

[4.1. Introducción 43](#_Toc44235001)

[4.2. Antecedentes de las comunicaciones de cámaras ópticas (OCC) 44](#_Toc44235002)

[4.3. Trabajos previos en sistemas Screen-to-Camera 48](#_Toc44235003)

[4.4. Idea básica del modelo de estudio 50](#_Toc44235004)

[4.5. Justificación y objetivos del modelo 52](#_Toc44235005)

[4.6. Limitaciones y soluciones del modelo 54](#_Toc44235006)

[CAPÍTULO 5: DISEÑO DEL SISTEMA DE TRANSMISIÓN 58](#_Toc44235007)

[5.1. La herramienta MATLAB 58](#_Toc44235008)

[5.2. Descripción del algoritmo propuesto 59](#_Toc44235009)

[5.3. Diagrama de flujo del programa 64](#_Toc44235010)

[5.4. Consideraciones del transmisor 64](#_Toc44235011)

[5.5. Consideraciones del receptor 65](#_Toc44235012)

[5.6. Estructura de los frames 65](#_Toc44235013)

[5.7. Técnicas de codificación empleadas 65](#_Toc44235014)

[5.8. Creación del código 65](#_Toc44235015)

[5.9. Algoritmo de decisión 65](#_Toc44235016)

[5.10. Codificación y multiplexación 65](#_Toc44235017)

[5.11. Implementación 65](#_Toc44235018)

[CAPÍTULO 6: VALIDACIÓN EXPERIMENTAL Y RESULTADOS 66](#_Toc44235019)

[CONCLUSIONES Y RECOMENDACIONES 68](#_Toc44235020)

[BIBLIOGRAFÍA REFERENCIADA 69](#_Toc44235021)

[ANEXOS 81](#_Toc44235022)

# **CAPÍTULO 1: INTRODUCCIÓN**

## **Antecedentes**

En la actualidad, el aumento de los dispositivos conectados a Internet, promovido por la creciente expansión del Internet de las Cosas (IoT del inglés Internet of Things), ha generado una gran demanda de comunicaciones inalámbricas, así como la progresiva saturación del espectro radioeléctrico, que constituye hoy en día un problema real [1]. Este hecho ha obligado a la búsqueda de soluciones alternativas que alivien la carga de los canales de radio convencionales al tiempo que se aseguren velocidades de transmisión altas. Aparecen así las Comunicaciones Ópticas Inalámbricas (OWC del inglés Optical Wireless Communications), que se posicionan como una tecnología alternativa en el sector de las comunicaciones. En concreto, dentro del campo de las Comunicaciones Ópticas, se sitúa la rama de las Comunicaciones por Luz Visible (VLC del inglés Visible Light Comunications) [2], la cual utiliza el amplio espectro no regulado de la luz visible. El ánimo principal de esta tecnología consiste en utilizar la actual la infraestructura de luminarias LED para la transmisión de información, a la par que preservan su función principal como sistema de iluminación. En lo que respecta a los sistemas receptores, se utilizan comúnmente uno o varios fotorreceptores independientes.

El interés por esta tecnología es tal que el Instituto de Ingeniería Eléctrica y Electrónica (IEE del inglés Institute of Electrical and Electronics Engineers) posee un grupo asignado, el IEEE 802.15.7, para la regulación y normalización de estos sistemas, así como sus aplicaciones [3]. A esta rama tecnológica se le conoce como Comunicaciones por Cámara Óptica (OCC del inglés Optical Camera Communications) o Comunicaciones Ópticas basadas en Cámara [4]. Los sistemas OCC comparten con cualquier sistema de comunicación los bloques generales de transmisor, canal y receptor.

En definitiva, las comunicaciones con cámaras ópticas (OCC) han aumentado su relevancia dentro del campo de las comunicaciones de luz visible (VLC) debido a la amplia disponibilidad de cámaras en un gran número de dispositivos (teléfonos inteligentes, tabletas, sistemas de vigilancia, …). La mayoría de estas cámaras se basan en la técnica de persiana (RS), que implementa un proceso de escaneo de imágenes fila por fila. Por lo tanto, una fuente de luz LED, que cambia a una frecuencia superior a la velocidad de obturación, aparecerá en la imagen como una serie de rayas oscuras y brillantes, representando los datos binarios. Por lo tanto, esta arquitectura de la cámara proporciona a los sistemas OCC una velocidad de datos muy superior a la velocidad de frames (fps) [84,85].

## **Justificación e importancia**

A lo largo de este proyecto se pretenderá evaluar la capacidad de un sistema de comunicación OCC, donde la señal transmitida se encuentra incrustada en las imágenes de vídeo emitidas por un televisor, mientras que la recepción será realizada mediante una cámara integrada en un dispositivo móvil.

Al tratarse de una tecnología altamente novedosa, no hay muchas referencias sobre las técnicas a utilizar y las prestaciones y capacidades de este tipo de sistemas. En esta propuesta se pretenden analizar y buscar diferentes opciones y estudiar cuáles son sus características y parámetros más importantes.

Diversos autores propusieron [5] un sistema en tiempo real basado en el canal alfa (transparencia) para introducir información espacialmente distribuida sobre cualquier imagen. El algoritmo sugerido por los autores dividía la pantalla en una retícula uniforme y decidía, en función de la secuencia de frames, qué cantidad de información incluir o, incluso, si descartar incrustar información (frames negros, por ejemplo) [6].

Otro ejemplo de esteganografía en vídeo es el uso de la transformada discreta del coseno (DCT del inglés Discrete Fourier Transform). Sin embargo, esta técnica ampliamente utilizada parte de la base de que tanto el emisor del mensaje esteganográfico como el receptor, disponen del fichero sin pérdidas [7]. En el caso que nos ocupa, el canal de transmisión modificaría la información al tratarse de una captura *in situ* del vídeo con información imperceptible en reproducción.

En esta propuesta se pretenden analizar y proponer diferentes opciones y estudiar cuáles son sus características y parámetros más importantes. El modelo que se va a evaluar consta de los siguientes componentes:

* **Sistema de transmisión OCC basado en vídeo**. Consiste en una pantalla de televisión convencional que emite imágenes de un vídeo pre-procesado al que se le ha aplicado técnicas esteganográficas para embeber la señal de datos. Dicha integración debe permitir incorporar la señal de información de tal manera que no se altere de forma apreciable por el usuario la calidad de imagen en el televisor. En este sentido se tendrá en cuenta, en primer lugar, las posibles propiedades de la imagen (iluminación, contraste, …) susceptibles de ser modificadas y los límites de dichas modificaciones que aseguren el mantenimiento de la calidad de las imágenes. A continuación, se procederá al estudio de las posibles codificaciones o modulaciones a realizar sobre dichas propiedades.
* **Sistema de recepción OCC.** Aunque el objeto de este proyecto no es el receptor, es conveniente conocerlo para trabajar con él. Consiste en una cámara de un móvil convencional que realizará la captación y procesado de la imagen de vídeo. Este tratamiento digital permitirá extraer la información embebida en la imagen. En este punto habrá que tener en cuenta aspectos tales como la zona de imagen a analizar o la complejidad computacional del procesado de señales a implementar.

Como resultado de este proyecto, se pretenden analizar las posibles soluciones de transmisión clasificándolas en función de su complejidad y prestaciones. Asimismo, las técnicas de detección asociadas y sus parámetros más relevantes como velocidad de transmisión, distancia o tiempo de adquisición de datos serán igualmente estudiadas.

## **Objetivos y alcance del proyecto**

Realizando una investigación e implementación del presente proyecto con la herramienta de desarrollo MATLAB® se llegará a determinar el diseño de un aplicativo que permitirá efectuar la codificación esteganográfica del vídeo para llevar a cabo una comunicación síncrona entre transmisor y receptor.

Se realizará el análisis de los factores que influyen en la percepción visual humana, que serán de vital importancia en el próximo punto a tratar, el sistema de transmisión Screen-to-camera (también conocido como S2C, nombre que recibirá a lo largo de este documento). En el siguiente punto anteriormente mencionado se efectuará un análisis de los S2C utilizados en la actualidad que serán de gran ayuda a la hora de diseñar nuestro sistema desde el punto de vista técnico. Además, se hará un especial hincapié en la forma en la que los datos irán codificados, por lo que se deberá realizar un estudio de las técnicas de codificación empleadas en dichos sistemas para obtener un esquema de codificación robusto frente a errores e interferencias. Se incluirá además una breve explicación de las características de los estegosistemas que nos permitirán conocer qué conjunto de métodos pueden ser utilizados para ocultar información.

Referente al aplicativo en software se podrá contar con una interfaz completamente didáctica, práctica y amigable para cualquier tipo de usuario, la cual posee una opción que permite seleccionar el tipo de vídeo que va a ser codificado.

En relación a los objetivos, el presente proyecto tiene como meta el estudio de los sistemas de codificación esteganográficos para las comunicaciones OCC. Los objetivos fundamentales que se propusieron para este Trabajo Fin de Grado han sido:

1. Estudio de los sistemas Screen To Camera (S2C).
2. Evaluación sistemas de codificación que permitan utilizar una televisión/pantalla como fuente de datos sin afectar a la percepción visual.
3. Diseño y evaluación de un sistema de test.

## **Medios utilizados**

Para la realización del código de este proyecto únicamente fue necesario disponer de un ordenador portátil con conexión a internet y del entorno MATLAB. Por el contrario, desde el punto de vista de las comunicaciones ópticas, ha sido necesaria la utilización de una pantalla que visualice el vídeo con el código incrustado (puede ser la misma pantalla del ordenador, tableta, televisor, …) y un móvil con un trípode que capture en forma de grabación dicho vídeo contenido con el fin de comprobar que en el receptor se ha realizado correctamente la comunicación.

## **Organización de la memoria**

Este documento se estructurará en siete capítulos, siendo el primero el actual, en el que se a cabo una introducción inicial, a modo de motivación, para comprender los motivos que han llevado a la elaboración de este proyecto, además de definirse la importancia y necesidad del tema que se aborda, y se dejarán explícitos los elementos de la investigación y diseño. Los capítulos sucesivos abarcan contenidos bastante diferenciados, los cuales son comentados a continuación.

* **Capítulo 2**: En el capítulo 2 se plasmará el estado del arte de la percepción visual humana, describiendo el sistema visual humano desde el punto de vista biológico y físico, destacándose especialmente los aspectos referentes al enmascaramiento visual y al proceso de visión humana. Asimismo, se pretende realizar una breve comparación entre los sistemas de pantalla actuales y cámaras modernas con el ojo humano.
* **Capítulo 3**: El capítulo 3 recoge un estudio del estado del arte de la esteganografía resaltando su aplicación en diversas ramas de la ciencia y la técnica. Se hace un análisis de los estegosistemas, caracterizándolos y clasificando los procesos y esquemas que intervienen, particularmente en el esteganografiado digital con imágenes.
* **Capítulo 4:** En el capítulo 4 se describirá el estado del arte de los S2C, aplicándole especial atención a las comunicaciones OCC que serán fundamentales en el diseño de nuestro sistema. Además, se describirán los principales objetivos del sistema a tratar junto con los principales problemas e inconvenientes del uso de esta tecnología, así como las soluciones planteadas.
* **Capítulo 5:** Este capítulo documentará el diseño de un método esteganográfico implementado en MATLAB, que pone en práctica la teoría matemática para el incrustado de la información. El algoritmo resultante se validará en los siguientes capítulos, donde se determinarán los logros alcanzados en cuanto a la perceptibilidad, señalando además el incremento de la robustez y seguridad del esteganograma obtenido en el proceso de codificación. Del mismo modo, en este capítulo se describirá el diseño del transmisor y del receptor de nuestro sistema, así como las decisiones tomadas en cada uno de ellos.
* **Capítulo 6:** A lo largo de este capítulo tendrá lugar una descripción del sistema implementado, utilizado para la realización de pruebas de funcionamiento y validación.
* **Capítulo 7:** Tras la realización de las ya mencionadas pruebas, los resultados obtenidos serán expuestos durante este capítulo 7.

Las conclusiones reflejarán los logros obtenidos con el diseño de S2C empleado, a partir de un análisis de los resultados y de los experimentos probados. Las recomendaciones se harán teniendo en cuenta las limitantes del método descrito, para su posterior perfeccionamiento por otros diseñadores e investigadores del tema; así como para el enriquecimiento de algunas de las temáticas abordadas. Además, se incluirán las referencias bibliográficas haciendo un listado, de toda la bibliografía consultada y siguiendo las normas establecidas. Por otro lado, la totalidad del código se encontrará en el apartado del anexo.

# **CAPÍTULO 2: EFECTOS DE LA PERCEPCIÓN VISUAL HUMANA**

## **Introducción**

El sistema visual humano (SVH) se ha desarrollado y perfeccionado después de millones de años de adaptación al medio que lo rodea, permitiendo al hombre interactuar de manera efectiva y precisa con su hábitat. El espectro visible, la visión central y periférica, las diferentes clases de células fotorreceptoras, y la visión binocular son algunas de estas adaptaciones. A lo largo de los años, los científicos han intentado definir modelos que expliquen lo mejor posible el funcionamiento del SVH. Un modelo muy conocido sobre la capacidad de la percepción de detalles en la visión es la *Función de Sensibilidad al Contraste* (del inglés *Contrast Sensitivity Function* – *CSF*), definida entre otros por Mannos y Sakrison [86], esta función intenta modelar la capacidad del sistema de visión de distinguir las diferentes frecuencias espaciales en función del contraste para un estímulo dado. Este modelo ha sido estudiado y. modificado a lo largo de los años, para contemplar, por ejemplo, variables como la velocidad y excentricidad.

El SVH tiene muchas limitaciones que pueden ser explotadas a la hora de representar diversas aplicaciones. El mundo real no es exactamente como lo percibimos. Aunque el proceso óptico y físico del sistema visual funciona prácticamente igual en todos los individuos, la percepción visual es mucho más compleja, puesto que consiste en la interpretación del estímulo recibido, y esta interpretación depende, en parte, de cada individuo.

Adquirimos conciencia del mundo que nos rodea a través de los sentidos. Los estímulos desencadenan sensaciones, pero la organización, interpretación y análisis de estas no depende exclusivamente de los sentidos, sino también del cerebro. A partir de los estímulos recogidos por los sentidos organizamos y recreamos la realidad y adquirimos conciencia de ella por medio de la percepción. El estímulo pertenece al mundo exterior y causa un efecto o sensación, mientras que la percepción es el proceso psicológico de la interpretación y depende, en gran parte, del «mundo interior» de cada individuo.

Si limitamos el estudio al campo visual, la percepción se define por el estímulo que produce la luz que, a su vez, nos crea una sensación que es analizada e interpretada en nuestro cerebro. Aunque el acto perceptivo tenga lugar de forma automática, es realmente complejo y tiene múltiples implicaciones. El mundo real no tiene que ser exactamente lo que percibimos por los ojos.

Los filósofos fueron los primeros en estudiar el origen de las percepciones: unos mantienen la teoría de que se trata de una reacción intuitiva e innata, mientras que otros creen que es fruto del aprendizaje y de la acumulación de experiencias. Finalmente, el proceso óptico y físico de la percepción visual, como se verá en los siguientes apartados, funciona mecánicamente de modo parecido en todos los humanos de vista sana. Las diferencias fisiológicas de los órganos visuales prácticamente no afectan al resultado de la percepción humana. Estas diferencias empiezan con la interpretación de la información que se recibe. Es decir, diferente cultura, educación, edad, memoria, inteligencia, e incluso estado emocional, pueden alterar el resultado final de interpretación. Se trata de una lectura de una interpretación inteligente de señales cuyo código no está en los ojos sino en el cerebro [22-26].

## **Espectro visible**

La capa de ozono actual filtra casi toda la radiación (ultravioleta) por debajo de los 280 nm (nanómetros), que, por ser radiación ionizante, es capaz de causar gran daño biológico en los tejidos externos. Al absorber la mayoría de esta radiación emanada por el Sol, la capa de ozono se convierte en un componente crucial de nuestro ecosistema.

La mayor intensidad de la radiación del Sol que penetra en nuestra atmósfera está localizada en la parte visible del espectro electromagnético. La máxima intensidad de luz solar a la que estamos expuestos está en la región amarilla del espectro visible (560 a 590 nm), lo que corresponde a la sensibilidad espectral de los conos fotorreceptores de la retina y responsables de la visión diurna.

La porción de rayos ultravioleta que llega hasta nosotros (de 280 a 400 nm) está en gran parte reducida por el ozono, pero viene en suficiente cantidad para causar reacciones biológicas familiares como un incremento en la producción de melanina y daños genéticos como cáncer de piel. Dado su potencial destructivo, es lógico en términos evolutivos que hayamos alejado nuestra sensibilidad de la luz ultravioleta.

El rango del espectro electromagnético visible para el humano está entre 400 y 750 nanómetros, pero no es casualidad que nuestro sistema visual pueda ver ese rango de frecuencias, sino otra adaptación. El medio en el que nos movemos (el aire) es prácticamente transparente a esta porción del espectro. La radiación visible es capaz de transmitir información desde los objetos distantes hasta nuestros ojos. La Figura 1 muestra el espectro visible al ojo humano, y su relación con el resto de las ondas electromagnéticas conocidas.



Figura 1. Espectro de ondas electromagnéticas.

En resumen, nuestras capacidades visuales son el resultado directo de una larga evolución en un entorno relativamente estable bajo la atmósfera y el Sol.

## **El ojo humano**

La información o estímulo viaja hasta nuestro sistema visual en forma de luz a través de un medio físico, comúnmente aire. La luz es transformada (enfocada) por las lentes de cada ojo y proyectada en la retina en la parte trasera de cada ojo. La retina consta de fotorreceptores, denominados conos y bastones, que convierten la radiación electromagnética (luz) en actividad neuronal (química y eléctrica). Los nervios ópticos conectados a los ojos transmiten esta información hasta la corteza visual del cerebro, en ambos hemisferios.

El campo de visión (FOV) está limitado obviamente por el movimiento que puedan realizar nuestros ojos (con la cabeza fija), por nuestra nariz y las cejas. Además, existe un gran solapamiento entre las imágenes que genera cada ojo, que el cerebro debe luego ordenar y juntar para generar una imagen coherente.

La estructura del ojo es manejada por 3 pares de músculos, y su equilibrio posicionamiento resultan en un constante e imperceptible movimiento, que hace que la información que recibe la retina cambie, aunque sea muy poco, cada 100 ms. Existen muchas imperfecciones en el sistema visual humano, como las aberraciones cromáticas, distorsiones y hasta fluidos en el ojo, que hacen que la luz no sea igualmente procesado, sin embargo el funcionamiento de la visión humana pareciera indicar que existen mecanismos temporales en el cerebro que permiten construir una imagen apropiada incluso ante la presencia de estos problemas.

La retina tiene aproximadamente 125 millones de fotorreceptores, que convierten la energía electromagnética en energía eléctrica y química. Los fotorreceptores se clasifican en conos y bastones. Los bastones son muy sensibles a la luz y a los cambios de luminosidad, pero se saturan fácilmente ante altos niveles de luz. Los bastones son más efectivos en el proceso de visión en condiciones de baja iluminación, llamada visión escotópica, con en la noche. Existen 120 millones de bastones, estos son más sensibles a longitudes de onda cercanas a los 500 nm y son hasta 10 veces más sensibles que los conos. Los conos, aproximadamente 6 millones, son mucho más efectivos en condiciones de alta iluminación (durante el día), en la visión fotópica, y son sensibles al color. Los conos se dividen en tres tipos, cada uno de ellos más sensible a diferentes longitudes de onda, desde las cortas (420 nm), las medias (530 nm) hasta las largas (560 nm) [87].

Los conos y bastones no están uniformemente distribuidos en la retina. Los conos están ubicados en una zona en el centro de la mácula lutea, llamada fóvea, de aproximadamente 1,5 metros cuadrados. Los fotorreceptores de la fóvea son los responsables de la agudeza visual en el proceso. Cuando se hace “foco” en un objeto, se centra la proyección de la luz que emite este justo en el centro de la fóvea, la foveola, de aproximadamente 0,33 mm de diámetro. En la fóvea, existe una convergencia 1:1 con las células ganglionares (del nervio óptico), por ello es la zona donde se obtiene mayor resolución o nitidez visual. El ser humano hace foco cambiando la forma de la lente (el poder focal en humanos es de 15 dioptrios).

* **Proceso temporal**. En cada fotorreceptor, ocurre un proceso químico que dura algunos milisegundos. Todos los fotones que recibe cada fotorreceptor durante este proceso contribuyen al resultado, dando lugar a una respuesta promedio en el tiempo que dura el proceso. El efecto de ese promedio se denomina “temporal smoothing”, y se puede comprobar fácilmente al observar luces que parpadean a gran velocidad. Cuando la luz parpadea lentamente, se percibe cada parpadeo individual correctamente, pero al aumentar la velocidad, y superar la tasa denominada CFF (*critical flicker fusion*) el flash se fusiona en una imagen continua. Esta CFF está alrededor de los 60 Hz. Bajo condiciones de iluminación bajas, el CFF disminuye levemente [88].
* **Proceso espacial**. Desde los más de 120 millones de fotorreceptores los estímulos deben transmitirse por aproximadamente un millón de células ganglionares. Estas últimas reciben los impulsos de múltiples fotorreceptores. La relación entre los fotorreceptores y estas células es muy diferente entre los conos y los bastones. Los conos, en la fóvea, tienen una relación de casi 1:1, requiriendo estímulos con suficiente energía para poder transmitir los impulsos, mientras que los bastones, en su mayor parte en la periferia de la fóvea, tienen una relación mucho mayor y muchos se conectan a la misma terminación de una célula ganglionar. La relación de muchos bastones a una sola célula ganglionar genera un efecto de amplificación del estímulo de la luz, por eso es que en la noche sólo los bastones participan en la visión, captando la poca luz disponible.

## **El sistema visual humano (SVH)**

Los ojos son el instrumento central en el sistema de visión humana, el cual reacciona a la luz y así obtiene la percepción de las formas de los objetos, los colores, la profundidad y el movimiento, etc. La estructura de un ojo es similar a la de una cámara [13], en el sentido de que un ojo también tiene una lente (lente cristalina) y una placa sensorial (retina). Las unidades sensoriales (células receptoras) de la retina son mayores en el centro y menores en los bordes, creando así una visión central (también llamada fóvea) y periférica. La visión centralizada es responsable de la alta capacidad de agudeza visual, esencial para capturar los principales detalles visuales. El área periférica más grande está caracterizada por sensores de luz altamente sensibles (células de varilla) y por lo tanto es adecuada para detectar el movimiento, y se vuelve más efectiva en la oscuridad. A la vez que ofrece abundante información del mundo físico, nuestros ojos desnudos tienen límites físicos inherentes en la resolución, tanto espacial como temporal. En concreto, nuestra percepción visual posee tres características ópticas, las cuales son respuesta centralizada, la fusión de parpadeo paso bajo y el efecto de matriz fantasma, que se relacionan con la resolución espacial, temporal y la sensibilidad al movimiento, respectivamente.

* **Respuesta centralizada**. Los ojos humanos no pueden discriminar los detalles demasiado finos, debido al límite físico de la mínima diferencia entre las células sensoriales adyacentes en la retina central. Típicamente, la excelente resolución espacial de los ojos humanos es de 0,07º, lo que corresponde de 1,2mm a 1 metro de distancia de observación [14]. La resolución espacial también se ve afectada por el contraste espacial. El campo receptivo de los ojos humanos tiene diferentes respuestas en el centro y en los alrededores. Los campos receptivos pequeños (o grandes) son estimulados por altas (o bajas) frecuencias espaciales. Típicamente, nuestra percepción espacial sigue una característica de paso de banda (cercana a la de paso bajo): la mayor frecuencia es de unos 2-4 ciclos por ángulo visual (equivalente a un ciclo de contraste dentro de 5mm a 1m de distancia de observación). En resumen, los ojos humanos no pueden distinguir cosas demasiado pequeñas, y los detalles más finos se fusionan y aparecen como una especie de promedio. Un ejemplo es el control de la vista: cuando los detalles finos exceden los del ojo, uno sólo verá la forma aproximada.
* **Fusión de parpadeo paso bajo**. Las fluctuaciones temporales de la intensidad de la luz no son perceptibles para el ojo humano cuando esta supera una cierta frecuencia, denominada frecuencia de parpadeo crítica (CFF para abreviar en adelante)[15, **16**]. En cambio, el ojo humano sólo percibe la luminancia media. El comportamiento temporal del sistema de visión humano puede aproximarse como un filtro lineal de paso bajo a una frecuencia superior a la CFF. Un ejemplo común es la observación de la rueda de un coche girando. Cuando gira lo suficientemente rápido, dará como resultado la percepción semitransparente de esta. Cabe destacar que la CFF se ve afectada por muchos factores, incluyendo los contrastes de color, el movimiento, las formas de onda de luminancia, entre otros muchos. En escenarios típicos, se cree que la CFF de los ojos humanos es de unos 40-50Hz según los resultados de una amplia investigación sobre la visión [17, 18**, 15**]. Por ejemplo, los parpadeos que presenta un monitor de rayos catódicos (CRT para abreviar) de 60Hz no son perceptibles.
* **Efecto de matriz fantasma**. Describe otra característica espacial del ojo humano: la sensibilidad al movimiento. Si bien los objetos en movimiento rápido se acercan a la vista (ya sea por el movimiento del objeto o por el movimiento del ojo, como los ojos que giran), se puede notar el parpadeo incluso cuando la frecuencia de visualización es mucho más alta [19]. Por ejemplo, podemos observar el parpadeo de un LED a una frecuencia muy superior a la típica de un CFF si el LED se mueve en un entorno oscuro, pero sólo vemos el LED constantemente encendido si está quieto. Cabe destacar que, a diferencia de la fusión del parpadeo, el origen del efecto de la matriz fantasma no ha podido ser determinado aún hoy en día. Estudios recientes descubren que a una menor amplitud de parpadeo se produce un mayor ciclo de trabajo y un mayor tamaño del haz lo hacen menos visible [20, 21].

## **El proceso de la percepción visual**

El sistema visual presenta un alto grado de especialización funcional, en el cual sus canales de entrada procesan de manera independiente y diferente el estímulo visual que se percibe. Se pueden distinguir dos etapas claras en el proceso de visión, una es la que detecta los cambios de menor resolución en una gran área visual, que es la visión periférica, y otra de mucho mayor resolución y sensibilidad al color, que permite identificar objetos y reconocerlos con detalle, la visión central.

La teoría más aceptada hoy en día sobre cómo funciona la visón, es el modelo multicanal. Desarrollada por Enroth-Cugell y Robson [90] y Campbell [89], esta teoría propone esencialmente que el sistema visual procesa la imagen de la retina de manera simultánea a varias escalas espaciales diferentes. La mayoría de las escenas naturales que se pueden considerar tienen información visual a diferentes escalas. Por ejemplo, en el caso de un bosque, la silueta de todos los árboles puede dar un nivel de detalle bajo de la escena, aunque también se puede observar cada árbol en particular, aumentándose así la resolución de los detalles, llegando hasta las hojas el máximo detalle en la escena. La teoría de este modelo multicanal sugiere que el sistema visual extrae toda esta información “multiresolución” simultáneamente, y que después, en alguna etapa del proceso de visión de más alto nivel, se reúne y combina para formar la escena tal cual como la percibe el hombre.

Esta teoría concuerda con el conocimiento actual del diseño neuronal del sistema de visión humano. El tamaño de campo receptivo de una neurona define el tamaño del estímulo al cual esta es altamente sensible. En las diferentes etapas tempranas de la visión se encuentran células con un amplio rango de campos receptivos, capaces de detectar cada una, diferentes niveles de detalle. Según este modelo, cada canal que procesa la información visual puede verse como una clase de neurona ajustada a un campo de recepción de un tamaño determinado. Del mismo modo, estudios [91] indican que el procesado del color, la agudeza visual, la velocidad de movimiento y el contraste también realizan el paralelo por diferentes partes del sistema visual.

Los objetos emiten o reflejan radiaciones luminosas de distinta frecuencia e intensidad que penetran en el interior del globo ocular a través de la pupila. La pupila se dilata o contrae en función de las condiciones lumínicas por la acción del iris. Después, la señal luminosa pasa por la córnea, el cristalino y la cámara interior acuosa hasta llegar a la retina, la parte fotosensible del ojo, dónde se encuentran las células ganglionares, bipolares y fotorreceptoras (los conos y los bastoncillos, las únicas células sensibles a la luz). Hay también, otros dos tipos de células: las células horizontales, que conectan entre sí a los conos y los bastoncillos, y las células amacrinas que conectas a las células bipolares con las ganglionares. La retina es un tejido fotorreceptor que cubre la mayor parte de la superficie interior del ojo y constituye el plano sobre el que se proyectan las imágenes de forma invertida. En la retina, los fotorreceptores (conos y bastones) transforman la luz en energía electroquímica que se transmite al cerebro a través del nervio óptico.

La energía electromagnética que incide sobre los conos y los bastoncillos se transforma en impulsos nerviosos que llegan hasta las células ganglionares, cuyos axones se unen para formar el nervio óptico en el disco óptico, llamado punto ciego porque carece de células fotorreceptoras y no es sensible a la luz. Los impulsos nerviosos, que proceden de la retina llegan a través del núcleo geniculado lateral (NGL) del tálamo hasta la corteza visual del cerebro, situada en el lóbulo occipital, donde se produce la propia percepción [22-26].

Los haces nerviosos de cada ojo se encuentran en el quiasma óptico donde parte de ellos se cruzan para ir a parar al hemisferio cerebral opuesto. Las fibras que salen del lado izquierdo de ambas retinas (y que corresponden al lado derecho del campo visual) se proyectan hacia el hemisferio izquierdo, y las que salen del lado derecho de ambas retinas (y que corresponden al lado izquierdo del campo visual) se proyectan hacia el hemisferio derecho, como se representa en la Figura 2.



Figura 2. La luz penetra en el interior del glóbulo ocular y se convierte en impulsos nerviosos. Los haces nerviosos de cada ojo se encuentran el quiasmo óptico, donde parte de ellos se cruzan para ir al hemisferio cerebral opuesto (en rojo, en la parte superior del cerebro). Los impulsos nerviosos llegan al tálamo y finalmente hasta la corteza visual (en rojo, en la parte inferior del cerebro).

## **Mecanismos básicos de la visión**

Hay tres mecanismos que intentan explicar el proceso básico de la visión. Estos son: los campos receptivos, la inhibición lateral y la magnificación cortical. Las células ganglionares se conectan con las células fotorreceptoras (los conos y los bastones) y funcionan de acuerdo con la estimulación recibida en el llamado campo receptivo. Se distinguen dos tipos de respuesta de los campos receptivos: de centro encendido y de centro apagado.

Las células ganglionares se activan al máximo cuando las células receptoras del centro se activan y las de la periferia se mantienen inactivadas (centro encendido). En sentido contrario, las células ganglionares de los campos receptivos de centro apagado se activan cuando son activados los receptores de la periferia y los receptores del centro del campo quedan inactivados (centro apagado). Es un mecanismo antagónico que permite una percepción óptima del contraste.

Es importante distinguir que los conos son muy eficaces en la percepción de contornos, contrastes y colores en condiciones de mucha luminosidad, mientras que los bastones funcionan mejor en condiciones de baja luminosidad. Este fenómeno explica las diferencias entre las conexiones de las células ganglionares con los conos y los bastones. Es decir, un cono se conecta con una célula ganglionar y varios bastones son conectados con una misma célula ganglionar.

Otro mecanismo básico de la visión es la inhibición lateral. Este fenómeno se produce cuando un estímulo induce efectos opuestos en el centro y la periferia de un campo receptivo. Estas regiones antagonistas compiten entre si y la célula ganglionar a la que están conectadas se mantiene prácticamente inactivada. De este modo se explica la percepción del contraste.

El tercer proceso es el que se conoce como magnificación cortical. La organización del sistema visual en el córtex del cerebro mantiene una «cierta» fidelidad con los objetos externos. Es decir, en el córtex visual se proyecta en cada momento un mapa topológicamente equivalente a la imagen externa. Sin embargo, la zona de enfoque principal de la imagen externa requiere una proyección mayor en la zona cortical que las zonas periféricas de la imagen [22-26].

## **El proceso cognitivo y la corteza visual**

Toda la información que se recibe a través de los ojos llega al cerebro, dónde se procesa. Se conocen unas 30 áreas visuales localizadas en los lóbulos occipitales, parietal, temporal y frontal de la corteza cerebral. Cada área extrae diferentes tipos de información de la señal de entrada visual; desde los rasgos más elementales como la frecuencia espacial, orientación y contraste (áreas del lóbulo occipital), hasta los rasgos más complejos tales como el movimiento, el color o la forma de los objetos (propio de las regios parietal temporal y frontal de la corteza cerebral).

Existe una organización jerárquica entre las áreas visuales. Todas las áreas están altamente interconectadas, pero cada una de ellas está especializada en una parte del análisis funcional de la información. La cognición visual es el resultado de interacción recurrentes entre las distintas áreas visuales. Sin embargo, hay que tener en cuenta que generalmente no se mira una escena de forma estática, los ojos se mueven buscando las partes «interesantes» de una escena para construir un mapa mental de la realidad observada.

El ojo humano hace movimientos oculares rápidos (conocidos como movimientos sacádicos) debido a que únicamente la parte central de la retina, la fóvea, tiene una alta concentración de conos. La fóvea es la parte de la retina encargada de la visión en alta resolución. El resto de la retina está tapizada básicamente por bastones, células fotosensibles monocromáticas, especialmente buenas en la detección del movimiento. Es decir, el ojo capta pequeñas partes de una misma escena para optimizar sus recursos. Si toda la escena fuera vista como lo que entendemos por alta resolución, el diámetro del nervio óptico debería ser incluso mayor que el propio globo ocular. Además, un procesado de toda la escena requeriría también un cerebro varias veces superior al que tenemos. Por este motivo, en el proceso de percepción visual primero tiene lugar la detección de los bordes del objeto que miramos, después se inicia el proceso de estereoscopia, el fusionado de las dos imágenes procedentes de los dos ojos. Este proceso tiende a construir la superficie de la imagen, añadiéndole textura, profundidad y orientación, captando movimiento y color. A continuación, se lleva a cabo el rellenado cognitivo del fondo, es decir, el cerebro construye la totalidad de la escena a partir de las pequeñas partes que visualiza mediante rápidos movimientos oculares prácticamente imperceptibles.

En el tramo final de todo estímulo visual se ha descrito la existencia de señales neuronales del tipo retroalimentación en el córtex visual. Estas señales de modulación tardía se asocian con varias zonas cognitivas como la conciencia visual, la atención y la memoria visual. Algunos estudios recientes [22-26] han demostrado que la interpretación que hacemos de los estímulos generados en nuestros ojos está condicionada por las presunciones que hacemos previamente sobre el entorno. Por lo que, en el proceso de la percepción visual, el cerebro parece que construye, en parte, lo que vemos.

Como se ha comentado en apartados anteriores, es posible visualizar mejor aquellas cosas a las que dirigimos la mirada directamente, pero, puesto que el cerebro construye parte de la realidad, las suposiciones que se hacen acerca del entorno condicionarán la percepción final. Por ejemplo, cuando confiamos en las sombras juzgando la forma de un objeto, lo que se están haciendo en realidad es suponer la reflectancia de la superficie del objeto, así como su iluminación.

Algunos estudios [22-26] sugieren que, sin este tipo de presunciones, las posibilidades de interpretar los estímulos visuales serían muy limitadas. Así, si parte de la percepción depende de nuestras propias presunciones, está claro que lo que nos muestra el mundo real no tiene por qué ser lo que percibimos. Los ojos sólo son responsables de una parte de la percepción visual, el cerebro hace el resto. En definitiva, se podría afirmar que se trata de un acto de reconstrucción interpretativa.

## **Enmascaramiento visual**

El enmascaramiento visual es la disminución o eliminación de la visibilidad de un estímulo, llamado objetivo, por la presencia de otro estímulo, llamado máscara, que está cerca del objetivo en el espacio y/o el tiempo [32]. El enmascaramiento visual se produce típicamente cuando el objetivo y la máscara tienen una orientación, una frecuencia espaciotemporal, un movimiento, un color u otro atributo similares [33]. Por ejemplo, la energía local de alta frecuencia en una imagen reduce la visibilidad de otras características de alta frecuencia como el ruido, reduciendo la importancia perceptiva de las distorsiones. Las distorsiones de compresiones JPEG y el ruido son muy visibles en regiones de luminancia suave como caras o cielos azules, mientras que pueden ser casi imperceptibles en áreas de alta textura como el pelo, la hierba o las flores [34]. Esto se llama enmascaramiento de contraste [35].

Se han utilizado modelos de enmascaramiento visual espacial de contraste (enmascaramiento) de textura para predecir la percepción de degradaciones en imágenes estructurales [36] y las diferencias de imágenes visibles [37]. La normalización de las respuestas neuronales también ha demostrado que reduce significativamente las dependencias estadísticas entre las respuestas y afectan a la visibilidad de la distorsión, además de proporcionar una explicación del efecto de enmascaramiento del contraste [38,39].

Para entender el enmascaramiento visual temporal se han realizado un gran número de experimentos psicofísicos utilizando destellos de luz [40], ondas sinusoidales [41], estímulos de vernier [42], cambios de vista [43], aglomeraciones [44], y cambiar los puntos de cierre [45-47]. En las investigaciones sobre el procesamiento de vídeo se ha comprobado que los observadores tienen dificultades para percibir una reducción temporal de los detalles espaciales en las señales de televisión inmediatamente antes y después de los cambios de escena [48]. Netravali y otros [49] investigaron la percepción del ruido de cuantificación durante las transiciones de luminancia. Haskell y otros colaboradores [50] estudiaron la tolerancia de los observadores a las distorsiones de las imágenes en movimiento. Otros investigadores [51] diseñaron un codificador de vídeo adaptativo utilizando la visibilidad del ruido en áreas planas, texturas y bordes. Girod [52] destacó el significado teórico del enmascaramiento temporal. Johnston y otros [53] construyeron un cuantificador de vídeo no lineal utilizando el enmascaramiento temporal.

Se han propuesto varias ideas para explicar el enmascaramiento temporal en los algoritmos de compresión de vídeo. Por ejemplo, [54,55] utilizaron los cálculos de la diferencia de trama global JND para tener en cuenta el enmascaramiento temporal localizado en el tiempo causado por los cambios de escena, mientras que, en una línea similar, [56] utilizaron el parpadeo temporal global para enmascarar la distorsión espacial.

## **Percepción del movimiento**

La percepción del movimiento es el proceso de inferir la velocidad y la dirección de los objetos en movimiento. La percepción del movimiento se produce desde la retina a través del núcleo genético lateral (LGN) y la corteza visual primaria (V1) hasta el área visual temporal media (MT) y más allá [61]. Las señales visuales se modelan comúnmente como un paso de banda espacial filtrado por las células ganglionares en la retina y el filtrado temporal en el LGN para reducir la entropía temporal. En V1, se produce una descomposición de los datos visuales en múltiples escalas y orientaciones a lo largo del espacio y el tiempo. Más adelante, el área extra cortical MT crea respuestas espaciotemporales orientadas localmente y ajustadas al movimiento [62,63].

Watson y Ahumada [64] propusieron un modelo de percepción de la velocidad, en el que el movimiento local se modeló como movimiento local translacional. En el dominio de la frecuencia, los movimientos complejos en segmentos de vídeo sin cambios de escena pueden ser analizados usando los espectros localizados espaciotemporalmente de los patrones de imagen que se supone que están siendo traducidos. Específicamente, supongamos que a(x, y, t) es un fragmento arbitrario de vídeo espacio-temporal en la coordenada espacial (x, y) y el tiempo t. Supongamos que 𝜆 y 𝜙 denotan las componentes de velocidad horizontal y vertical de un fragmento de imagen que se transforma. Cuando un pedazo de imagen se convierte a una velocidad constante [𝜆, 𝜙], la secuencia de vídeo en movimiento se convierte en a (x -𝜆𝑡, y - 𝜙𝑡, t). El espectro de un fragmento de imagen estacionario se encuentra en el plano u, v, mientras que la transformada de Fourier de un fragmento de imagen que se traslada se corta en un plano oblicuo a través del origen. Tal plano puede ser expresado como: 𝜆𝑢 + 𝜙𝑣 + 𝑤 = 0, donde u, v y w son variables de frecuencia espacial y temporal correspondientes a (x, y) y t, respectivamente. La orientación de este plano indica la velocidad y la dirección del movimiento.

## **Percepción del parpadeo de la visión humana (flicker)**

Los ojos humanos perciben el cambio temporal de la intensidad de la luz, es decir, el parpadeo, de una manera de paso bajo. Cuando la intensidad de la luz fluctúa lo suficientemente rápido, los ojos humanos sólo perciben la intensidad media en lugar del parpadeo. Este fenómeno se conoce como efecto de fusión del parpadeo. La frecuencia más baja que causa la fusión del parpadeo se llama frecuencia de parpadeo crítica (del inglés Critical Flicker Frequency - CCF). CCF no es un valor constante y varía con muchos factores. A continuación, señalamos algunas características importantes de la fusión de parpadeo:

* El CCF depende de la amplitud del parpadeo. Con un flicker más débil, la CCF también será más baja.
* La visión humana es más sensible al cambio de luminancia que al cambio cromático (color). El CCF cromático es de unos 25 Hz, sólo la mitad del CCF de luminancia [80].
* El CCF se considera generalmente inferior a 60 Hz en la mayoría de las circunstancias [81]. Los dispositivos de pantallas están diseñados para refrescarse a más de 60 Hz de frecuencia para evitar el parpadeo visible [82].

Se han propuesto diversos métodos para predecir la visibilidad del parpadeo, incluida la suma de las diferencias al cuadrado [57,58]. Sin embargo, estos métodos dependen en gran medida del contenido y requieren varios umbrales entre los marcos y se limitan a una precisión por bloques, sin tener en cuenta el enmascaramiento visual temporal de las distorsiones de parpadeo locales.

Aunque el ser humano no puede percibir el parpadeo de alta frecuencia, si es posible grabar dicho parpadeo con cámaras. La diferencia entre la visión humana y la exposición de la cámara permita que los métodos de comunicación pantalla-cámara basados en VLC logren una transmisión discreta con incrustación entre frames.

En relación al enmascaramiento visual temporal de las distorsiones de parpadeo, se realizaron una serie de estudios subjetivos en videos sobre la calidad, movimiento de objetos, frecuencia de parpadeo, y excentricidad [59,60]. Los resultados muestran que la visibilidad del parpadeo local está fuertemente reducida por la presencia de grandes y coherentes movimientos de objetos, de acuerdo con los experimentos psicofísicos humanos sobre estímulos sintéticos [46,47]. El impacto de los movimientos de los objetos en la visibilidad de las distorsiones de parpadeo es significativo cuando la calidad de un video de prueba es pobre.

La fusión de parpadeo puede ser también descrita por el espacio de color CIE XYZ, por la regla de la fusión del color. El espacio de color CIE XYZ representa cada color percibido con tres valores triestímulos (X, Y, Z). Específicamente, el componente Y mide la luminancia de un color. La regla de fusión de colores es una conclusión directa de la ley de Grassmann de igualación de colores [114]. Esta ley establece que si dos colores L1 = (X1, Y1, Z1) y L2 = (X2, Y2, Z2) se alternan en la pantalla a una frecuencia constante (más alta que la CCF), la visión humana percibe el color fusionado como:

, )

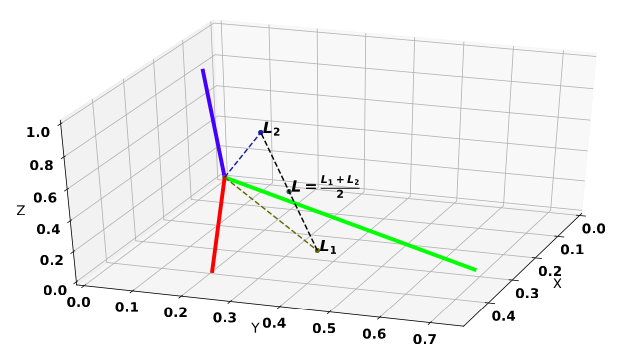
Llamamos al color L1 y L2 como un par de fusión de L, que puede verse como una descomposición de L en el espacio de color CIE XYZ (Figura 3). Especialmente, si mantenemos invariable la componente de luminancia Y. Por ejemplo, Y1 = Y2, entonces L1 y L2 forman un par de fusión lumínica-invariante de L, que puede verse como una descomposición de L en el plano XZ.

Figura 3. Ejemplo de fusión de colores en el espacio de color CIE XYZ. L1 y L2 forman un par de fusión de L. Las líneas rojas, azules y verdes corresponden a los colores primarios RGB.

## **Factores que determinan la sensibilidad al contraste**

Se pueden encontrar varias definiciones del contraste en la literatura, aquí se mencionan las tres más empleadas:

* **Weber**. El contraste según Weber, directamente deducido de la ley general de Weber, intenta establecer una relación cuantitativa entre la magnitud de estímulo físico y cómo éste es percibido. La ley general de Weber establece que “*el menor cambio discernible en la magnitud de un estímulo es proporcional a la magnitud del estímulo*”. En el caso del contraste, la función derivada de dicha ley es la siguiente:

donde e representan la luminancia de las características y del fondo respectivamente. Este modelo es usado normalmente cuando se encuentran pequeñas características sobre un fondo de luminancia constante.

* **Michelson**. El contraste de Michelson [92] se define como:

con y las luminosidades máximas y mínimas respectivamente. El contraste C está en el intervalo {0 ≤ *C* ≤ 1}. El contraste de Michelson se utiliza en patrones en los cuales las características oscuras y claras son equivalentes y ocupan proporciones similares en la imagen.

* **RMS**. El contraste RMS no depende de la frecuencia espacial ni de la distribución de contraste de la imagen. Se define como el desvío estándar de las intensidades de los píxeles [93] y viene dado de la forma:

siendo la intensidad de cada píxel de una imagen de *M* por *N*, y como el promedio de intensidades de la imagen.

Durante los años 60, algunos investigadores [86] desarrollaron algunas ideas sobre la habilidad del hombre de percibir los detalles, e intentaron encontrar los límites del sistema de visión humano. El resultado de estos estudios es un modelo muy conocido denominado *Contrast Sensitivity Function* (*CSF*) presentado por Mannos y Sakrison entre otros.

La función CSF indica los límites de percepción a diferentes frecuencias y contrastes de un estímulo visual. Si la frecuencia del estímulo visual es muy alta (detalles imperceptibles) entonces no se podrá reconocer más el patrón del estímulo, y la percepción será solamente de un color uniforme. Estos límites de resolución espacial están directamente relacionados con el número y el tamaño de nuestros fotorreceptores. La función CSF a la que llegaron Mannos y Sakrison es la siguiente:

y en la Figura 4se puede observar una representación de tal función. La curva delimita lo que las personas, con visión normal, pueden ver y lo que no pueden ver, usando la visión central.

Hasta hace poco se creía que la CSF de una persona no podía mejorar, y que a medida que el sujeto avanza en su edad su CSF indicaría una pérdida de agudeza visual. Estudios recientes de la Universidad de Rochester [93] mostraron que jugadores de videojuegos tienen mejor percepción de pequeños detalles (altas frecuencias) que personas que no juegan frecuentemente, demostrando que el SVH puede adaptarse y mejorar, en función de los estímulos cotidianos.



Figura 4. Función de sensibilidad al contraste.

La sensibilidad al contraste depende de muchos factores que tienen un origen tanto óptico como neural. Entre los factores ópticos más relevantes, debido a que determinan la función de transferencia de la modulación óptica, están el diámetro de la pupila [27,28], la excentricidad [29,30] y, naturalmente, el estado refractivo del sujeto [27,31].

En definitiva, entre los factores neurológicos, que son los que dan más información sobre la naturaleza de los mecanismos fisiológicos que intervienen en la detección de los patrones acromáticos, consideraremos particularmente tres: la luminancia media, el tamaño del estímulo y la excentricidad. La luminancia media determina el estado de adaptación del sistema visual, el tamaño determina el número de ciclos que comprende el estímulo y la excentricidad determina las características de las matrices de sensores que actúan en cascada para realizar una determinada tarea visual.

## **Sensibilidad temporal**

El SVH puede distinguir menos detalles cuando el objeto que se proyecta directamente en la fóvea está en movimiento. El resultado de esto es la sensación de emborronado a medida que el objeto pasa frente a el punto de enfoque o cuando el sujeto mueve la cabeza para enfocar otro punto.

Murphy [94] propuso que la incapacidad del ojo de seguir objetos en movimiento de manera precisa es la causa del efecto de emborronado que se percibe, sin embargo, estudios más recientes hechos por Tyler [95] y Nakayama [96] sugieren que los fotorreceptores limitan la sensibilidad temporal a los detalles, concluyendo que el proceso de detectar movimiento en sí mismo implica la integración de energía, en el tiempo durante el movimiento del objeto, y es justamente éste proceso de integración temporal y espacial el que puede destruir la información visual de determinadas características de alta frecuencia del objeto.

Daly [97] extendió el modelo de la CSF para agregar la variable velocidad. Esta versión se puede aplicar bien a imágenes estáticas y enfocadas con la visión central (fóvea), es decir, cual es la percepción del estímulo cuando éste está en movimiento y el sujeto está siguiéndolo con su vista.

Figura 5. (a) CSFs para velocidades de 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64 y 128 grados (derecha a izquierda). (b) La frecuencia espacial detectable más alta posible en función de la velocidad.

En la Figura 5 se pueden ver dos gráficas representadas por Reddy [98] que muestran como la CSF se desplaza en el eje de la frecuencia espacial al aumentar la velocidad del estímulo. La mayoría de los autores coinciden que al incluir en el modelo el movimiento del estímulo, el resultado es que la CSF sufre desplazamientos y escalados.

## **Agudeza visual y visión periférica**

La visión periférica es una parte del proceso de visión de visión humano que ocurre fuera del centro de la fóvea. Comparada con otros animales, la visión periférica de los seres humanos es pobre, especialmente a la hora de distinguir colores y formas. En el sistema de visión humano la densidad de los fotorreceptores en la retina es mucho mayor en la parte central que en los bordes.

Los modelos anteriores de CSF no se aplican bien cuando se considera la agudeza visual en la visión periférica. Las principales funciones de la visión periférica son, reconocer estructuras y formas conocidas sin la necesidad de enfocarlas, identificar formas similares y movimientos (leyes de la psicología de Gestalt) y proporcionar sensaciones al sistema visual que conforman el trasfondo de la visión detallada.

El SVH no percibe las diferencias visuales como si fuera una función binaria, sino que lo hace como una función de probabilidad en función del ángulo visual en grados, que permite distinguir la característica. Un umbral sobre esa función de probabilidad define si el objeto o la característica es detectada por el sistema o no. Los diferentes modelos propuestos a lo largo de las últimas décadas son intentos de definir estas funciones que describan lo más preciso posible el comportamiento del sistema de visión. Está claro que la sensibilidad y agudeza visual decrecen en la periferia del ojo, aunque las hipótesis varían entre los autores.

## **Pantallas y cámaras modernas**

Con el rápido avance de la tecnología en los últimos años, las pantallas y cámaras están mejorando rápidamente. La mayoría de las pantallas LCD, especialmente las que tienen capacidad 3D, soportan una frecuencia de refresco de 120 Hz o superior. Por ejemplo, el monitor del Macbook Air de 13 pulgadas, que se utiliza en nuestro trabajo, soporta 60 fotogramas por segundo (FPS). Otros monitores como el LG 47LH55 incluso soporta físicamente 240 FPS. Además, con la aparición de las pantallas basadas en diodos emisores de luz orgánica (OLED), se soporta un tiempo de respuesta mucho más bajo (menos de 0,01 ms) en comparación con el LCD, lo que indica su gran capacidad para soportar una tasa de refresco bastante alta.

Las cámaras también están evolucionando de forma espectacular. Las cámaras de alta resolución para fines concretos pueden producir imágenes de un gigapíxel y alcanzar una velocidad de cuadro de captura de vídeo de más de 2.000 FPS. Incluso las cámaras de los teléfonos inteligentes convencionales pueden soportar imágenes de alta resolución a una velocidad de captura rápida, que supera lo que la retina está buscando. Por ejemplo, el Samsung Galaxy S5 admite una resolución de 16 megapíxeles y una tasa de captura de 120 FPS, mientras que el iPhone 6 admite una resolución de 8 megapíxeles y una tasa de captura de hasta 240 FPS. En el caso de nuestro proyecto, usaremos como dispositivo móvil el Huawei Mate 10, el cuál soporta una grabación de 30 FPS y 60 FPS en el modo dual.

## **Comparación entre cámara y ojo**

Una diferencia importante entre el ojo y la cámara es que el ojo no contiene un obturador. Como resultado, no hay un proceso de exposición seguro mientras vemos las cosas, ni vemos las cosas en marcos (o instantáneas). Esto también explica la propiedad de fusión del parpadeo. Por el contrario, la cámara toma instantáneas discretas, cada detalle de la instantánea puede ser examinado. Además, mientras que el sistema de visión humana está todavía mucho más avanzado que una cámara, según la métrica específica de la resolución espacial y temporal, ya es superado ampliamente por la cámara.

Las propiedades distintivas del SVH y las cámaras, y su diferencia de capacidades proporcionan un espacio de diseño potencial que podemos explorar. Considerando el hecho de que el SVH se ha mantenido en gran medida constante durante mucho tiempo mientras que las tecnologías de pantalla y cámara siguen avanzando a pasos agigantados, la brecha de capacidad, y por lo tanto la zona de diseño, se hará más grande en el futuro.

# **CAPÍTULO 3: SISTEMAS ESTEGANOGRÁFICOS**

## **Introducción**

Hasta el momento se ha hablado de la esteganografía como el conjunto de métodos para ocultar información. En realidad, esta definición no es del todo correcta, ya que a este conjunto de métodos se le conoce precisamente como métodos de ocultación de la información (del inglés, information hiding), siendo la esteganografía un subconjunto de los mismos. La ocultación de la información es la ciencia que engloba cualquier método que permita esconder cualquier tipo de información, sea cual sea su naturaleza, sus medios, o sus fines. En definitiva, la esteganografía es la rama de la criptología que trata sobre la forma de escribir información secreta de tal modo que sólo quien la envía y quien la recibe sepan de su existencia.

En el marco de las telecomunicaciones, la esteganografía, el arte del enmascaramiento de la información, adquiere gran utilidad e importancia para el anonimato y la seguridad de transmisiones secretas. Gracias a la digitalización, los avances tecnológicos del software, el hardware computacional y la expansión de las redes de telecomunicaciones, se genera y transmite a diario un gran torrente de información, el cual hace que sea posible el ocultamiento de metadatos, así como el transporte desapercibido de información confidencial. La esteganografía es desarrollada con diferentes técnicas y variantes, desde tiempos remotos, con el objetivo de proveer privacidad en las comunicaciones a distancia; así pues, un sistema esteganográfico será aquel que enmascara el contenido en un medio para evitar que el mensaje sea percibido, de modo tal que ninguna parte fuera del emisor y el receptor conozcan la existencia de este.

Actualmente existe un creciente número de aplicaciones que utilizan esta tecnología, no sólo para llevar a cabo comunicaciones privadas a través de Internet, sino también en la protección de la propiedad intelectual (copyright), en votaciones digitales, en transacciones electrónicas y para el espionaje en entornos militares o económico-industriales. Diferentes esquemas esteganográficos compiten entre sí para lograr mayor robustez al ruido, alta capacidad de almacenamiento de información, seguridad ante la detección no autorizada e imperceptibilidad de la imagen esteganografiada.

El principal objetivo de los sistemas esteganográficos es garantizar la total imperceptibilidad del mensaje oculto; con este propósito se han desarrollado diversas técnicas con alto grado de complejidad y éxito, pero ninguna es totalmente invulnerable, debido a la continua evolución de las técnicas de esteganoanálisis, orientadas a la adquisición de la información.

Debido a la importancia que ha adquirido esta ciencia en los últimos tiempos, este trabajo surge con motivo de documentar los algoritmos y procesos que llevan a cabo la esteganografía en imágenes y realizar una descripción del diseño e implementación de un método para el cifrado de la información de manera imperceptible, segura y robusta; para ello se analizan y comparan las principales técnicas aplicadas en los sistemas esteganográficos, comenzando por las más simples, desarrolladas en el dominio espacial, y finalizando con el estudio de los complejos algoritmos que requieren específicas transformaciones de la señal. De esta forma, se da margen a la aparición de nuevos y más avanzados métodos que en las situaciones requeridas y gracias a la existente tecnología logran su cometido.

En todo sistema esteganográfico se pueden identificar los siguientes elementos:

* *La carga útil, mensaje o marca*: son los datos o mensajes que se desean transportar de forma oculta. Es el fichero u objeto que se pretende embeber. Este fichero puede ser texto, imagen, audio o vídeo, bien sea cifrado o sin cifrar.
* *La portadora o fichero que cubre a la carga útil (estegomedio)*: es el fichero que oculta la carga útil. Es el objeto visible en el que el mensaje embebido se oculta. Análogamente, puede ser de texto, imagen, audio o vídeo.
* *El paquete u objeto esteganográfico, señal esteganografiada o fichero con la marca digital (esteganograma)*: es el resultado de incrustar la carga útil en la portadora. Es la portadora con la carga útil, mensaje o marca de agua oculta embebida.
* *Clave secreta (llave) o par de claves pública-privada esteganográficas*: es la clave o claves esteganográficas utilizadas para controlar el proceso de ocultar la información y extraerla intacta en el receptor. Aportan seguridad al sistema.

## **Esteganografía en la era digital**

El desarrollo de la tecnología ha propiciado la creación de un escenario adecuado para elevar la esteganografía a un nivel superior. Los avances en computación proporcionan medios para calcular rápidamente los cambios necesarios en la ocultación de un mensaje, e Internet proporciona los medios para transportar grandes cantidades de información a cualquier punto del planeta [65]. La esteganografía actual se basa en esconder datos binarios en ficheros. Los bits que componen el mensaje a ocultar se introducen en la cubierta digital ya existente (bien sea añadiéndolos, o realizando operaciones aritméticas con los originales), procurando que el fichero resultante después de realizar los cambios parezca el original.

El proceso de ocultamiento de la información en un sistema esteganográfico comienza por identificar el medio de recubrimiento, el estegomedio, que no es más que le medio en donde el mensaje será ocultado. Los diferentes tipos de estegomedios son [66]: archivos de imagen, sonido, texto o vídeo, archivos ejecutables, páginas web, campos no usados de paquetes de redes (TCP(IP)), espacios no utilizados de discos y particiones escondidas.

Cualquier tipo de fichero es apropiado para ocultar datos en su interior, pero los más utilizados han sido las imágenes y ficheros de sonido. Las imágenes digitales son adecuadas para esconder datos por muchas razones; están siempre presentes en Internet, ilustrando un sitio Web, o de adjunto en un correo electrónico; lo que no llama la atención. Aún con las técnicas modernas de compresión, las imágenes son todavía relativamente grandes y pueden variarse de forma imperceptible. Existen varios métodos desarrollados para la esteganografía con imagen, más que para cualquier otro estegomedio [67].

Cada tipo de formato de fichero gráfico se puede utilizar para desarrollar métodos esteganográficos concretos [68]; pero todos ellos tienen características comunes que permiten su uso activo en la esteganografía:

* Contienen una cabecera que identifica el tipo de fichero del que se trata, que posee información para interpretarlo como puede ser el tamaño de la imagen o el número de colores.
* Una vez descomprimidos los datos de un fichero mediante el algoritmo concreto de cada formato, los datos del fichero indican el color específico de cada píxel de la imagen.

En función de los colores de la imagen se utilizarán más o menos bits que indicarán las cantidades de rojo, verde y azul que se usará para representar cada color en cada píxel. Debido a que la paleta de colores puede ser modificada según la imagen, es necesario almacenarla en un fichero. De los 24 bits de cada píxel, se utilizan 8 bits por color; en el caso de imágenes a escala de grises, cada píxel utiliza sólo 8 bits para establecer un valor entre 0 y 255 que representa el nivel o luminancia de este.

## **Estegosistema: descripción y características**

Como se mencionó con anterioridad, la esteganografía consiste en ocultar un mensaje secreto en un medio de cobertura para que sólo el destinatario previsto sea consciente de la existencia del mensaje oculto. Para mayor seguridad, el mensaje suele ser codificado antes de ser incrustado en el medio de cobertura. El resultado de este procedimiento de codificación es una secuencia pseudo-aleatoria de 0 y 1, que luego se incrusta en la cubierta. Al incrustar los bits del mensaje en una imagen, normalmente se introducen algunos artefactos visuales y se cambian las estadísticas de la imagen. Las estadísticas de la imagen son analizadas por métodos de estegoanálisis, mientras que los artefactos visuales pueden ser detectados por el ojo humano.

La Figura 6 representa los pasos de inserción y extracción en un sistema esteganográfico general, en el cual el transmisor (también llamado codificador) aplica primeramente alguna transformación a la imagen portadora, en la que el mensaje **m** será enmascarado, obteniendo el vector **X** que representa a los valores ya transformados de la imagen. Esta transformación (DFT - Transformada discreta de Fourier, DCT- Transformada discreta del coseno, DWT-Transformada discreta de Wavelet) es opcional en dependencia del método que se aplique. Seguidamente, se lleva a cabo el algoritmo esteganográfico del esquema en cuestión que puede ser tres tipos esencialmente: los que modifican el bit menos significativo, los de espectro extendido y los que aplican modulación del índice de cuantificación; al cual se le introduce la llave de codificación **k**, útil también en el proceso de descifrado, y es obtenida como salida la señal esteganografiada **S**, que con una transformación inversa resulta ser la imagen con el mensaje o marca incluida.

De esta manera, el esteganograma viaja a través de un determinado canal, el cual introduce ruido o es objeto de ataques intencionados (para obtener la información escondida o simplemente destruirla) o ataques no intencionados (manipulaciones al fichero como es la compresión o filtrado), hacia el receptor. Una vez realizada o no, la misma transformación que en el codificador, es obtenido el esteganograma con ruido **S’**, que junto a la original llave **k** forma las entradas del esquema de encubrimiento inverso para de esta forma extraer el mensaje. El estegosistema anterior puede contener técnicas criptográficas para mayor seguridad, o códigos de corrección de errores agregando redundancia al mensaje para incrementar la robustez de este.



Figura 6. Pasos de inserción y extracción del mensaje en un sistema esteganográfico.

La esteganografía adaptativa utiliza las características de la imagen y la información del contenido para mejorar el proceso de incrustación. Por ejemplo, se pueden introducir artefactos menos detectables en la imagen de la portada por el método de incrustación. En algunos casos, estos artefactos se detectan con dificultad por los métodos de esteganalización debido a los cambios triviales introducidos en las estadísticas de las imágenes. Sin embargo, los artefactos pueden ser perceptibles visualmente para el ojo humano. Por lo tanto, la indetectabilidad perceptiva es muy importante en la esteganografía. Esto significa que la aplicación de un método de esteganografía a una imagen no debe dejar ningún fenómeno visualmente perceptible. Para apoyar esta idea y mejorar la robustez de los métodos de marca de agua, se propusieron modelos perceptivos para describir el sistema visual humano.

Las características de un estegosistema, representante de un esquema determinado por un algoritmo específico, dan una medida de la eficiencia de este, por lo que se usan para hacer comparaciones entre los métodos existentes y establecer pautas de superioridad o deficiencias. Estas son:

* *Imperceptibilidad*: Mide la diferencia que existe entre la imagen original (el estegomedio) y la imagen esteganografiada con el mensaje oculto (esteganograma), la cual no debe ser distinguible por el sistema visual humano (SVH). Se mide en términos de la relación señal a ruido pico, PSNR. Un PSNR mayor que 30 dB, implica que la diferencia entre ambas no es distinguible por el SVH.

Las siglas MSE representan el error de distorsión cuadrático medio existente debido a la inserción y los vectores **Io** y **Iw** son la imagen original y marcada, respectivamente, y los valores **N1**, **N2** representan el tamaño de las imágenes.

* *Robustez*: Número de modificaciones que el medio puede soportar antes de que se pierda la información oculta (antes de que el ruido del canal o un adversario destruya o modifique la información escondida).

Se mide en BER (razón de error de bit). Un número sustancial de bits erróneos es dado típicamente por un BER de 0.15.

Muchas técnicas intensifican la robustez de sus mensajes implementando un codificador de control de error cíclico, para la recuperación de estos, y para la detección de errores en la transmisión.

* *Capacidad de inserción*: Cantidad de información que puede ser ocultada en determinado estegomedio. La capacidad se puede medir en bytes de información, respecto al estegomedio, determinándose, de esta forma, el número de bits insertados en un byte o píxel (para el caso de imágenes a color) de la cubierta.
* *Seguridad*: Mide la dificultad de detección de la información escondida. La seguridad va a aumentar con el nivel de dificultad del proceso de cifrado esteganográfico, generalmente, y se puede incrementar aún más con procesos criptográficos incluidos en cada uno de los algoritmos realizados.

## **Introducción a los métodos y técnicas esteganográficas empleadas**

Los métodos esteganográficos se clasifican según la forma de inserción en [66]:

* *Métodos de adición*: Se oculta el mensaje secreto en las secciones del medio portador que pueden ser ignoradas por la aplicación que lo procesa.
* *Métodos de sustitución*: Se modifican ciertos datos del medio portador por los datos del mensaje secreto.
* *Métodos de generación*: Se crea el esteganograma a partir de la información secreta, sin contar con un medio portador previamente.

Los esquemas de sustitución son los más aplicados para la esteganografía con imágenes, a causa de la gran variedad de imágenes existentes con diferentes matices y tonalidades pictóricas; y gracias a que contienen una mayor información que el propio mensaje a empotrar, el resultado de los algoritmos aplicados es muy similar a la imagen original de la cubierta utilizada en el proceso de codificación. Debido a la gran diversidad de formas de llevar a cabo este “arte”, los métodos se pueden dividir en dos grandes grupos según el dominio de la imagen utilizado [68]:

* *Basados en el domino espacial de la imagen*: abarcan técnicas que aplican la modificación de los bits menos significativos de cada píxel, o modificándolos mediante expresiones aritméticas, con puertas lógicas o mediante la manipulación del ruido.
* *Basados en el dominio de la transformación*: están relacionados con algoritmos de modificación y transformación de la imagen. Ocultan los mensajes en las áreas más significativas de la cubierta y pueden manipular las propiedades de la cubierta como es la luminosidad. Utilizan técnicas como son la transformada discreta del coseno (DCT), la transformada discreta de Fourier (DFT), o de Wavelet (DWT), entre otras.

Estas transformadas se pueden desarrollar por bloques, o sobre la imagen entera. De la primera forma se divide la imagen en bloques más pequeños (los dos tamaños más populares de estos bloques son 8x8 y 16x16), y la transformada se realiza individualmente en cada bloque. La ventaja de usar el dominio transformado es que es generalmente más fácil balancear la distorsión introducida por el ocultamiento y da mayor robustez contra el ruido o el ataque, que en el dominio espacial.

La esteganografía es un arte complejo y con muchos matices. Sin llegar a la combinación de esteganografía y criptografía, es posible el uso de determinadas técnicas avanzadas que permiten aumentar la eficacia de una información oculta [65], como son:

* *Uso de múltiples claves*: técnica heredada directamente de la criptografía, pero con distinta forma de aplicación. Consiste en usar distintas codificaciones para cada porción arbitraria del mensaje a ocultar. Así en el dominio espacial, por ejemplo, una frase de cinco palabras puede tener una clave de codificación para cada una de las palabras: en la primera se resta una unidad en los ceros y se suma una unidad en los unos, en la segunda se realiza lo mismo, pero invirtiendo el orden de los bits, en la tercera se realiza una XOR de los bits.

Las claves, también reconocidas como llaves esteganográficas, son muy utilizadas por los esquemas para incrementar seguridad. Si el receptor no tiene exactamente la misma llave usada en el algoritmo de empotrado de la información, entonces no podrá obtener los datos ocultos, aunque conozca el procedimiento inverso correcto. Estas pudieran ser también datos sobre la imagen original, sobre la naturaleza del mensaje oculto (si es una imagen: sus dimensiones, si es un fichero: el tipo o cabecera del mismo), o simplemente algún dato adicional del algoritmo codificador.

* *Esteganografía en capas*: establece una relación lineal entre los elementos ocultos. Por ejemplo: la codificación de la segunda palabra o letra de un mensaje depende de la primera palabra o letra (puede depender del último valor de la cifra, del último valor modificado, de la posición). De esta forma el esquema, que aplica la esteganografía en capas, establece un orden estricto de decodificación que impide obtener completamente el mensaje sin la primera parte, con lo cual únicamente se debe comunicar la pauta a seguir para encadenar los fragmentos en el decodificador.
* *Adición de ruido*: en un esteganograma todo el fichero es considerado ruido, pero se puede añadir ruido en el proceso de esteganografiado. Además de modificar los bits necesarios para inyectar los datos, se modifican unos cuantos bits aleatorios del mensaje de forma que, aún teniendo el fichero original, un posible atacante deba conocer el sistema de codificación usado.
* *Uso de distintas magnitudes*: aunque lo habitual es variar en 1 bit de determinado byte de la imagen original, nada impide variarlo en más bits. Así, se establecen algoritmos más complejos que incrementan la seguridad del mecanismo de empotrado, como, por ejemplo: en el procesamiento de una frase de cinco palabras, para ocultar la primera de las palabras se suma 1 bit en la codificación de la primera letra, 2 bits en la codificación de la segunda, 3 bits en la tercera, hasta que vuelva a aparecer una modificación de 1 bit, que significará el inicio de otra palabra.

En nuestro caso de trabajo, para la esteganografía en imágenes existen diferentes enfoques, como la esteganografía de conservación de estadísticas, la esteganografía basada en modelos y la incrustación de máscaras como métodos de procesamiento natural, como bien se han mencionado algunos con anterioridad. El principal logro en la esteganografía en imágenes es que el mensaje secreto no sea reconocido por terceros que no tengan derechos de acceso. El receptor debe extraer la información oculta mediante el algoritmo que utiliza el emisor.

En la estenografía basada en modelos, se asume un modelo para la imagen de la portada [99]. El procedimiento de incrustación cambia los elementos de la imagen (valores de los píxeles o coeficientes de transformación) para que no violen el modelo de cubierta asumido. Por ejemplo, Sallee [100] propuso una distribución Cauchy generalizada (GCD) para modelar la distribución de cada coeficiente de la transformada del coseno discreto (DCT) y asumió que todas las imágenes obedecen a este modelo. La idea de utilizar un modelo para la estadística de imágenes en su enfoque era una idea novedosa. Pero, desafortunadamente, los métodos de estegoanálisis detectan fácilmente las imágenes estego-generadas por este método [101], [102]. Otra de las técnicas más empleadas en la esteganografía de imágenes es la LSB (del inglés Least Significant Bit), en la que el mensaje secreto es cambiado a binario para cubrir el interior de la imagen.

## **Técnicas de espectro ensanchado para la esteganografía con imágenes (SSIS)**

SSIS es un esquema ciego donde la imagen original no se necesita para recuperar la información oculta [130]. El destinatario sólo necesita una llave para revelar el mensaje. Este sistema proporciona la habilidad de esconder una cantidad significante de bits de información dentro de las imágenes, evitando la detección por un observador. Este objetivo defiende la maximización de la capacidad y la minimización de la perceptibilidad.

Técnicas de comunicación de espectro ensanchado, codificación para el control de errores, y procesamiento de imágenes se combinan para lograr el esquema SSIS. El proceso del sistema de transmisión se encuentra proyectado en la Figura 7. Dentro del sistema, el mensaje, después de la encriptación opcional, es codificado con una baja razón de código de corrección para no incrementar el mensaje debido a redundancias, produciendo el mensaje codificado **m**. El remitente introduce la llave 2 en un generador pseudo-aleatorio de ruido de banda ancha, generando una secuencia extendida **n**. Luego, llegamos al esquema de modulación que se usa para extender el espectro de banda estrecha de **m** en la secuencia **n**, componiendo la señal **s**, la cual es introducida en un entrelazador espacial. Los entrelazadores también pueden usar una llave para dictar el algoritmo de entrelazado. La señal resultante se combina con la imagen de la cubierta para con la imagen de la cubierta para producir la imagen esteganografiada la cual es apropiadamente cuantificada para conservar el rango dinámico inicial de la imagen, obteniendo así la parte transmisora del estegosistema.

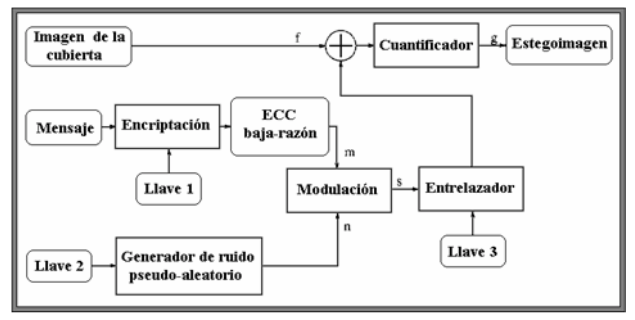


Figura 7. Proceso de codificación SSIS.

En el receptor, el sistema decodificador, mostrado en la Figura 8, usa técnicas de restauración de imagen y desentrelazado para construir una estimación de la señal con la información **s^**, de la imagen recibida. El destinatario mantiene la misma llave 2 que usó el remitente, regenerando con ella la secuencia extendida **n**. El mensaje codificado se desmodula y se construye una estimación **m^** de este mensaje codificado. La estimación del mensaje se descifra con el decodificador de errores, y entonces desencripta si es necesario, para que los datos ocultos sean revelados.

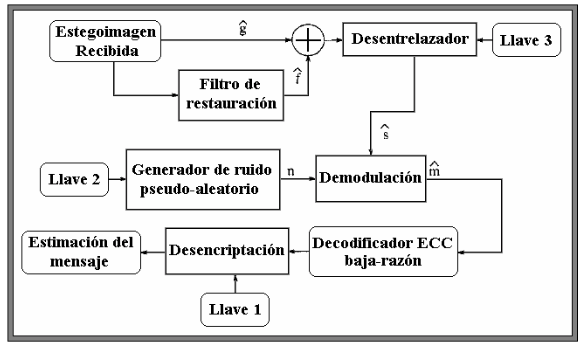


Figura 8. Proceso de decodificación SSIS.

# **CAPÍTULO 4: SISTEMAS DE TRANSMISIÓN SCREEN2CAMERA**

## **Introducción**

En los últimos años hemos sido testigos del rápido crecimiento de las pantallas electrónicas desplegadas en el mundo cibernético. En efecto, estamos rodeados de tales dispositivos en varios factores de forma, que van desde la pantalla de un teléfono o de una tableta, el monitor de un ordenador, un televisor, un tablero electrónico de publicidad y pantallas aún más grandes. Estas pantallas se han convertido en una fuente primaria de información desde el punto de vista del usuario. Por ejemplo, la reproducción de vídeo ha contribuido al 79% del tráfico de Internet [12]. En estos escenarios de uso común, la función principal de una pantalla es transmitir en formación información a los ojos humanos. Así, la pantalla establece la comunicación entre ella misma y los ojos. En estos escenarios de uso común, la función principal de una pantalla es transmitir en formación a los ojos humanos. Se establece así la comunicación de pantalla a ojo. A menudo es muy deseable transmitir aún más cierta información a un usuario mientras mira la pantalla. Un ejemplo de la vida real es remitir al espectador a una página web de películas para obtener información adicional. Otro caso similar es el de los vendedores que presentan información añadida (por ejemplo, especificaciones de productos y ventas) en los anuncios de televisión. Con la creciente prevalencia de las cámaras, la comunicación visible de pantalla a cámara ha surgido rápidamente como un canal de comunicación conveniente e improvisado

Durante este periodo se ha producido la adopción universal de los enlaces visibles entre pantalla y cámara como un valioso canal lateral para la comunicación entre dispositivos [8-11]. Así pues, se introducen códigos de barras dinámicos, por ejemplo, códigos de respuesta rápida (QR), para aumentar la capacidad de transporte de datos por el canal visible entre la pantalla y la cámara de vídeo. Sin embargo, debido al espacio limitado de la pantalla, el contenido visual favorable para el ser humano (normalmente en forma de imágenes) y el contenido visual favorable para los dispositivos (por ejemplo, los códigos QR) se ha disputado la asignación del espacio del monitor y han dado lugar a desafíos estéticos. En consecuencia, hemos visto que un código QR únicamente puede tomar una pequeña área de todo el espacio. Esto no sólo limita su capacidad de transporte de información, sino que también induce a esfuerzos adicionales para poder capturar el código (por ejemplo, estar suficientemente cerca del código QR). La experiencia del usuario para la representación de la pantalla también se considera una distracción. Por lo tanto, surge una pregunta clave: ¿Podemos eliminar esta diferencia restaurando la visualización del cuadro completo para los usuarios y estableciendo simultáneamente la comunicación de datos entre el dispositivo de pantalla y el cuadro completo? Fundamentalmente, exige un nuevo paradigma de comunicación en modo dual, que permita la entrega simultánea de contenido de vídeo primario y otra información sin perjudicar la experiencia de visualización del usuario.

En este proyecto, se presenta un sistema que permite la comunicación visible en modo dual, de cuadro completo, mediante la multiplexación del canal visible para transportar datos sobre el contenido de vídeo normal. La Figura 9 ilustra el concepto de multiplexación de cuadro completo, donde la pantalla muestra el contenido de video multiplexado del video original y los cuadros de datos (por ejemplo, códigos QR). El usuario aún puede ver el video como de costumbre, sin notar los marcos de datos incrustados. La cámara captura los datos transportados por los cuadros multiplexados y decodifica la información relevante. La información se codifica en patrones visuales y se muestra en la pantalla. Los dispositivos equipados con cámaras capturan posteriormente la imagen mostrada por el monitor y recuperan la información de los datos.



Figura 9. Concepto de multiplexación de fotograma completo: Screen-to-eye para videos y Screen-to-camera para la comunicación de datos a través del mismo canal visible simultáneamente.

## **Antecedentes de las comunicaciones de cámaras ópticas (OCC)**

La demanda de comunicaciones móviles de datos está creciendo enormemente con la omnipresente conectividad del Internet de las Cosas (del inglés Internet of Things - IoT), tal y como se menciona en capítulos anteriores, y el crecimiento de los servicios digitales como los medios sociales y los contenidos de vídeo. Según el último informe de Ericsson, se espera que el tráfico móvil total aumente con una tasa de crecimiento anual del 42% [70]. En 2023, el tráfico de datos móviles globales mensuales superará los 100 exabytes (EB). Para satisfacer esta demanda que aumenta exponencialmente, se debe utilizar ya sea el aumento del ancho de banda o la mejora de la eficiencia espectral. Sin embargo, el aumento de la eficiencia espectral es lento y no puede satisfacer esta demanda insaciable. La explotación del nuevo espectro se convierte en una única solución, ya que se demanda una banda de espectro mucho más amplia de terahercios (THz). En esas circunstancias, las comunicaciones ópticas inalámbricas han suscitado un gran interés de investigación en los últimos años debido a sus muchas propiedades deseables, entre ellas una gran cantidad de espectro disponible (de 350 nm a 1550 nm), una gran eficiencia energética, una normativa independiente y una seguridad de las comunicaciones debidamente controlada [71,72,73].

De ahí a que las comunicaciones inalámbricas ópticas (del inglés Optical Wireless Communications - OWC) aparecieran como una alternativa y una opción complementaria a las comunicaciones de radiofrecuencia (RF) existentes. Las OWC cubren las bandas de ultravioleta (UV), infrarroja (IR) y visible. Basándose en la utilización del espectro, la información a tratar sobre las OWC terrestres pueden dividirse en cuatro categorías: comunicaciones ópticas en el espacio libre (FSO), comunicaciones de luz visible (VLC), fidelidad a la luz (Li-Fi), y comunicaciones de cámaras ópticas (OCC) [74]. Los sistemas de comunicación FSO consisten en un transmisor de diodo láser (LD) y un receptor de fotodiodo (PD). Típicamente se basa en bandas UV o visibles, ofrece una alta tasa de transmisión a larga distancia, y puede ser usado para la reconfiguración de redes de comunicación. Sin embargo, las comunicaciones de los FSO necesitan una alineación estricta, y, en consecuencia, su costo en los equipos es alto. El VLC basado en bandas visibles es un transmisor de LD o diodo emisor de luz (LED) y un receptor de PD basado en tecnología de comunicación de medio alcance. VLC es capaz de ofrecer una alta tasa de datos dentro de un rango de decenas de metros, pero no considera el acceso de múltiples usuarios. A diferencia de VLC, Li-Fi es una tecnología de red de luz basada en un transmisor LED que implica el acceso de múltiples usuarios, comunicaciones bidireccionales, traspaso de múltiples células, etc. La tecnología Li-Fi soporta las comunicaciones móviles, permite un acceso sin problemas y puede ofrecer una velocidad de megabits por segundo (Mbps). Sin embargo, todos los receptores de las tecnologías FSO, VLC y Li-Fi, consisten en PD, que raramente existen en los dispositivos receptores actuales y el costo de comercialización para cambiar dicha estructura de los receptores es alto. Por lo tanto, el desarrollo de un sistema práctico de OWC para aprovechar los beneficios inmediatos utilizando dispositivos comerciales ya existentes sigue siendo una cuestión pendiente [75]. Esto motiva el concepto de OCC. Las diferentes características y limitaciones de OCC y VLC se presentan en la Tabla 1.

OCC es una tecnología de comunicación que utiliza sensores ópticos de imagen como receptores basados en bandas IR o visibles, y, por lo tanto, también se denomina como comunicaciones de sensores de imagen [76,77]. Comparada con otros tipos de tecnologías OWC, la OCC tiene muchas ventajas. En primer lugar, no necesita ninguna modificación en el receptor, en el que los actuales teléfonos inteligentes, cámaras digitales, cámaras de vehículos traseros y cámaras de vigilancia son todos posibles candidatos a receptores, por lo que OCC puede facilitar un gran número de aplicaciones.

|  |  |  |
| --- | --- | --- |
| Características | OCC | VLC |
| Rango de transmisión | Hasta varios kilómetros | Bajo |
| Longitud de onda | UV, IR y luz visible | Luz visible |
| SNR | Alto | Bajo |
| Receptor | Cámara | Fotodetector |
| Decodificación | Alta complejidad | Baja complejidad |
| Velocidad de datos | Menor que VLC (en kbps) | 11.67 kbps – 96 Mbps |
| Protocolo | IEEE.802.15.7r1 | IEEE.802.15.7 |
| Implementación MIMO | Fácil | Difícil |

Tabla 1. Comparación de OCC y VLC.

En segundo lugar, los receptores del OCC se basan en millones de píxeles, y estos proporcionan grandes grados de libertad (DoF) para transmitir datos y manejar el acceso de un número masivo de usuarios. Tercero, los sensores de imágenes de las cámaras actuales pueden normalmente tratar con tres colores que permiten la transmisión sobre el dominio de color. Debido a estas ventajas en cuanto a costo, popularidad y capacidad de transporte de información, OCC ha atraído mucho interés en áreas como IoT (del inglés Internet of Things), la localización en interiores, la captura de movimiento y los sistemas de transporte inteligente (STI). Sin embargo, OCC tiene sus limitaciones, que incluyen, entre otras, una baja tasa de datos debido a la baja tasa de muestreo en el receptor, el efecto de desenfoque, la inestabilidad de la velocidad de cuadro y el bloqueo aleatorio. La comunidad de investigadores de OCC ha propuesto diferentes soluciones para superar las limitaciones ya mencionadas. Por lo tanto, el tema central de este apartado es presentar las diversas limitaciones de OCC y sus soluciones.



Figura 10. Categorías de OCC.

OCC puede dividirse en tres categorías, como se ilustra Figura 10. La primera categoría es la de un solo LED como transmisor y la cámara como receptor (LED2C), la segunda categoría consiste en las matrices del LED como transmisor y la cámara como receptor (LEDA2C), que también se denomina MIMO visual [78]. La tercera consiste en una pantalla como transmisor y una cámara como receptor (S2C) que es la que utilizaremos en este trabajo [79].



Figura 11. Estructura del receptor de la cámara óptica, procedimiento de comunicación y aplicaciones.

Basándonos en la clasificación anterior, los posibles candidatos para un transmisor en los sistemas OCC son los LEDS, las pantallas de televisión, las pantallas de monitor, semáforos y las luces de los coches (Figura 11). La luz de todas estas diferentes fuentes pasa por el canal inalámbrico y llega a la lente del receptor. Después, la luz pasa a través del obturador que controla la exposición del sensor de la cámara. Existen dos tipos diferentes de obturador, el obturador global y el obturador rodante. El obturador global controla los sensores de la imagen simultáneamente, mientras que el obturador rodante controla los sensores de la imagen para mostrarla fila por fila [80]. Se genera una imagen de la fuente capturada en el plano imagen de la cámara, que puede procesarse adicionalmente utilizando técnicas de procesamiento para extraer los datos requeridos para diversas aplicaciones. En los sensores de imagen, la luz se transforma en una señal electrónica y se cuantifica para obtener una imagen y se comprime a un formato de calidad específico. Mediante el procesamiento de imágenes y los métodos de extracción de datos, la información transportada en el dominio color/espacio/temporal de la luz, puede decodificarse para su localización y navegación. La variación del canal de luz puede utilizarse para el reconocimiento del movimiento. Aunque la velocidad de los datos es muy limitada debido a la reducida velocidad de fotogramas, pueden transmitir alguna información sobre localización a baja velocidad sobre el reconocimiento de la ubicación, y también puede realizar una comunicación segura cerca del campo.

Por otro lado, los enlaces OCC necesitarían de varias etapas de procesamiento de imágenes para asegurar una recepción correcta, como la detección y el seguimiento de la fuente de luz (la región de interés, ROI), la ecualización de la potencia óptica a lo largo de la ROI y el cálculo del umbral para la demodulación.

Un canal basado en cámara es análogo a un canal RF MIMO, donde cada píxel de la cámara actúa como una antena receptora y los elementos emisores de luz (pantallas) como las antenas de transmisión. En este tipo de canales la calidad de la señal en cada elemento de la antena receptora es una función de la pérdida de trayecto en el canal, el desvanecimiento multitrayecto y la interferencia de otras antenas de transmisión, también llamada interferencia cocanal [109]. Un canal de estas características tiene un desvanecimiento insignificante, pero experimenta una pérdida de trayectoria en la energía de luz e interferencia de otros elementos emisores de luz, que se manifiestan como distorsiones visuales en la salida de la cámara, es decir, en la imagen. Estas distorsiones son principalmente un derivado del proceso de formación de imágenes de la cámara y pueden modelarse (de forma determinista) utilizando la teoría clásica de formación de imágenes en cámaras.

La calidad de la señal en el receptor también se ve influenciada por el ruido del canal. El ruido en los sistemas basados en cámaras se manifiesta como una señal eléctrica ficticia en forma de corriente en cada píxel de la cámara. La corriente de ruido se genera debido a los fotones de la iluminación ambiental y de los circuitos de imagen del transmisor y receptor [110]. La corriente de ruido en un píxel suele considerarse independiente de la señal cuando la iluminación ambiental es suficientemente alta en comparación con la señal de transmisión; por ejemplo, en salas de oficinas o al aire libre [111]. A la salida de la cámara, la corriente de ruido en cada píxel es una cantidad cuantificada y se manifiesta como fluctuaciones en la intensidad (valor digital de la salida del sensor) de ese píxel. Por consiguiente, la energía de ruido acumulada en cada píxel puede ser cuantificada usando el valor medio de la varianza en la intensidad del píxel.

## **Trabajos previos en sistemas Screen-to-Camera**

En un mundo de los dispositivos inteligentes son cada vez más numerosos, equipados con pantallas y cámaras. El hecho de permitir que las pantallas y las cámaras sean capaz de comunicarse ha atraído un interés cada vez mayor. La idea es simple: la información se codifica en un marco visual que se muestra en una pantalla, y cualquier dispositivo equipado con una cámara puede apuntar a la pantalla y obtener inmediatamente la información. Al funcionar en la banda del espectro de la luz visible, la comunicación entre la pantalla y la cámara está libre de interferencias electromagnéticas, lo que ofrece una prometedora alternativa de comunicación fuera de la banda para la adquisición de información de corto alcance. Asimismo, las comunicaciones de luz visible (VLC) son un campo de investigación muy desarrollado. La comunicación entre pantalla y cámara también aprovecha el canal de luz visible. Naturalmente, las técnicas de VLC pueden ser utilizadas también en este campo en desarrollo. La mayoría de los trabajos de comunicación Screen-to-camera basados en VLC hacen hincapié en la capacidad e invisibilidad. VRCodes [115] y HiLight [116] son los primeros trabajos de este campo, que utilizan el cambio cromático y de luminancia entre frames, respectivamente, para modular los datos incrustados. Por otro lado, TextureCode [117] optimiza la selección del área de incrustación para reducir eficazmente la obstrucción. Uber-in-Ligth [118] aprovecha el parpadeo (flicker) cromático para modular los datos y propone un método mejorado para la demodulación basado en MUSIC para mejorar la precisión de la transmisión. Las comunicaciones Screen-to-camera basadas en tecnología VLC requieren un esquema de extracción de datos diferente (extracción basada en vídeo) a los códigos de barras (extracción basada en imágenes). Del mismo modo, recientes investigaciones han dado lugar a innovadores diseños de códigos de barras que aumentan la velocidad de los datos [119,120] o mejoran la fiabilidad de la transmisión [121-124]. Estos esfuerzos son muy interesantes; sin embargo, comúnmente requieren la visualización de imágenes codificadas visibles, que interfieren con el contenido de la pantalla y crean una experiencia de visualización desagradable.

Los sistemas mencionados han tenido como objetivo el de conseguir una comunicación discreta entre la pantalla y la cámara, que permita a la pantalla cumplir simultáneamente un doble papel: el de mostrar el contenido y la comunicación. En última instancia, imaginamos un sistema de comunicación entre la pantalla y la cámara que transmita y reciba datos en tiempo real, al tiempo que garantiza que la comunicación se produzca de forma discreta, independientemente del contenido que se muestre en la pantalla, ya sea una imagen, una película, un videoclip, una página web, una interfaz de un juego o cualquier otra aplicación. A medida que el usuario interactúa con la pantalla y cambia el contenido, la comunicación se mantiene. Por lo tanto, la comunicación se realiza totalmente como una funcionalidad adicional para la pantalla, sin poner ninguna limitación a la funcionalidad original de esta (mostrar el contenido). Como se mencionó anteriormente, no estamos solos en el trabajo hacia esta visión, ya que recientes investigaciones han realizados valiosos progresos en el diseño de comunicaciones discretas entre pantalla y cámara [126**-**129]. Sin embargo, sigue existiendo una brecha fundamental en la capacidad de realización, principalmente porque todos los diseños existentes requieren modificaciones directas de los valores de color de los píxeles (RGB). Esta metodología no puede permitir una comunicación discreta en tiempo real sobre un contenido arbitrario en la pantalla que pueda generarse sobre la marcha con las propias interacciones del usuario. La razón es doble. En primer lugar, la modificación de los valores RGB de los píxeles en el tiempo tiene que apoyarse en la Unidad de Procesamiento Gráfico (GPU) aprovechando directamente las bibliotecas de esta. Sin embargo, el sistema operativo no suele permitir que una aplicación de terceros prevea o modifique el contenido de la pantalla de otras aplicaciones o interfaces del sistema (por ejemplo, la pantalla de inicio de un teléfono inteligente o una tablet). Por lo tanto, para lograr la comunicación en tiempo real, los diseños existentes se limitan al contenido de la pantalla dentro de una sola aplicación o archivos independientes (por ejemplo, imagen o vídeo). En segundo lugar, si bien el procesador principal (CPU) puede preconfigurar y modificar los valores RGB de los píxeles, la modificación de estos valores en la CPU conlleva un retraso considerable de cientos de milisegundos, lo que hace que estos diseños sean incapaces de soportar la comunicación en tiempo real sobre contenidos dinámicos como el vídeo. De esta premisa surgen los sistemas mencionados con anterioridad.

## **Idea básica del modelo de estudio**

Para superar esta barrera mencionada en el apartado anterior, se propone un nuevo paradigma de diseño para la comunicación discreta entre pantalla y cámara. El sistema que se pretende tratar se basa en la brecha de capacidad entre el ojo humano y los dispositivos electrónicos. Se sabe que el sistema de visión humana tiene un límite físico, mientras que las pantallas y cámaras modernas han excedido nuestros ojos en términos de resolución temporal, es decir, velocidad. Específicamente, el sistema de visión humana tiene una resolución temporal de hasta 40-50Hz más allá de la cual no podemos capturar objetos que se mueven más rápido. En contraste, las pantallas modernas, especialmente aquellas con capacidad 3D, admiten 120FPS o una frecuencia de actualización más alta. Las cámaras de los teléfonos inteligentes (smartphone) pueden tomar imágenes de alta resolución a altas velocidades de cuadro. Por ejemplo, el Samsung Galaxy S5 ofrece una resolución de 16 megapíxeles y una tasa de captura de 120FPS, mientras que el iPhone 6 admite una resolución de 8 megapíxeles y una tasa de captura de 240FPS. Además, la capacidad del dispositivo avanza a un ritmo mucho más rápido que el sistema de visión humana.

Existen dos desafíos cuando se diseña la comunicación visible en modo dual de fotograma completo usando este sistema. Primero, el requisito de no afectar el canal primario de *Screen-to-eye* que implica una restricción rígida para la comunicación en modo dual. Además, el sistema de visión humana es sensible a los parpadeos. Por lo tanto, necesitamos encontrar un esquema de multiplexación de cuadros apropiado para combinar un flujo de datos con un vídeo arbitrario. El objetivo es permitir que el video presentado se perciba sin distorsión de color, artefactos y parpadeos cuando se muestra en una pantalla de alta frecuencia.

En segundo lugar, el canal secundario *Screen-to-camera* podría experimentar una posible interferencia del contenido de video original en el canal primario. Necesitamos diseñar un esquema de codificación efectivo de trama de datos. Debe ser resistente tanto a los cambios en el contenido de video como a las limitaciones conocidas de comunicación entre la pantalla y la cámara (por ejemplo, falta de coincidencia de velocidad de fotogramas, efecto de obturador, mala calidad de captura). Este sistema aprovecha la discrepancia de capacidad y las características distintivas del sistema de visión humana y los dispositivos (pantalla y cámara), por ejemplo, las pantallas pueden mostrar el contenido más rápido de lo que perciben los ojos humanos; las cámaras tienen obturador, pero los ojos humanos no, etc. El vídeo y los datos transportados por los cuadros compuestos operan en diferentes escalas de tiempo. El contenido del vídeo se percibe a un ritmo lento debido a los límites físicos del ojo humano, mientras que los cuadros de datos se muestran a una velocidad rápida, que sólo puede ser capturada y decodificado por la cámara.

En el diseño de nuestro proyecto, superamos el primer desafío aprovechando la propiedad de fusión de parpadeo del sistema de visión humana, es decir, la exhibición de un efecto de filtro paso bajo al ver escenas que cambian más rápido que la frecuencia de parpadeo crítica, y la superior resolución temporal de las pantallas. De este modo, ideamos un concepto de frame complementario, incrustando un frame de datos en un par de frames de vídeo multiplexados. Un fotograma multiplexado tendrá artefactos obvios, pero cuando se reproduce lo suficientemente rápido, los artefactos de los frames complementarios se anulan cada uno. Se propone además una técnica de suavizado de frames para incrustar gradualmente los frames de datos en el contenido de vídeo. Para combatir el segundo desafío, diseñamos una estructura jerárquica del marco de datos y también esquemas especiales de codificación. En particular, el esquema de codificación utiliza los estados de encendido y apagado de un patrón artificial de tablero de ajedrez para señalar un -1/1 bit, donde el valor -1 actúa como nuestro 0 lógico.

En la comunicación entre pantalla y cámara, la información se modulará en función de la intensidad de la luz de los píxeles del transmisor (pantalla) que se reciben, y se decodificarán a partir de la intensidad de los píxeles de la imagen capturada por la cámara en el receptor. La intensidad del píxel capturado por la cámara es una cantidad digital Q1 que es proporcional a la cantidad de corriente de fotones generada en el píxel a partir de la energía lumínica acumulada en su área (cuanto más pequeña sea el área, menor será la intensidad de luz que acumule). Cuando el píxel de la pantalla que emite la luz está en el foco del objetivo de la cámara, todos los rayos de luz del píxel de la pantalla se enfocan en un píxel de la cámara y, por lo tanto, no se produce ninguna pérdida de energía en el píxel. Por otro lado, si el píxel de la pantalla es perturbado (en posición y/u orientación) por el enfoque de la cámara o incurre en una pérdida de energía debido al tamaño finito de la apertura del objetivo de la cámara, no todos los rayos de luz convergerán en el píxel de la cámara, lo que dará lugar a una reducción de la energía acumulada y, por tanto, a un menor valor de intensidad del píxel. La pérdida de la intensidad de la luz recibida en un píxel de la cámara da lugar a la deformación visual del tamaño o la forma del píxel de la pantalla visualizada; un efecto que se denomina distorsión de la perspectiva. Para controlar la transparencia de los píxeles se utilizó el canal alfa, ya que regula el cambio de transparencia de los píxeles para que sólo sea perceptible para las cámaras, pero no para los ojos humanos. Además, al controlar la translucidez de cada píxel independientemente, permitimos que haya múltiples elementos transmisores simultáneos en la pantalla. Esto crea un canal de comunicación MIMO entre las pantallas y las cámaras, que puede aumentar la velocidad de los datos y mejorar la fiabilidad de la transmisión.

En resumen, nuestro sistema tiene dos canales de comunicación visual concurrentes, el canal primario Screen-to-eye para los humanos y el canal secundario Screen-to-Camera para los dispositivos. Ambos canales se originan en la misma pantalla, coexistiendo y contrarrestando el uno al otro. Se ha implementado este sistema y realizado un estudio preliminar. Los experimentos realizados confirman que dicho sistema permite una comunicación visible de modo dual, tanto a los humanos como a los dispositivos, simultáneamente.

## **Justificación y objetivos del modelo**

Hoy en día, las cámaras actuales se utilizan con frecuencia para leer códigos QR, que pueden considerarse a su vez como una forma de comunicación visual en la que la cámara actúa como un receptor. El uso generalizado de estos códigos motiva la construcción de nuevas aplicaciones de comunicación con cámaras, donde las pantallas podrían ser moduladas para enviar dichos códigos variables en el tiempo para ser decodificados por la cámara posteriormente.

Los patrones visuales espaciales (por ejemplo, los códigos de respuesta rápida QR anteriormente mencionados) no son del todo consumibles, si es que lo son, por los seres humanos que se sienten más cómodos con los contenidos de fácil acceso, como textos, imágenes y vídeos. Cuando una pantalla se convierte en la misma fuente, tanto para el contenido primario favorable para las personas como para la información lateral óptima para las cámaras, ambos compiten por la pantalla. Reconociendo esa controversia y el hecho de que la entrega de contenido consumible por los seres humanos es más bien un propósito primario, en las prácticas actuales se ha recurrido a algún tipo de compromiso, ya sea espacial o temporal.

En el caso del código QR, el código tomaría una pequeña área y residiría en una esquina de toda la pantalla (Figura 12a), u ocuparía toda la pantalla a su vez después de los contenidos para humanos (Figura 12b). Sin embargo, la anterior comprobación espacial no sólo limita su capacidad de transporte de información, sino que también incurre en un trabajo adicional para capturar adecuadamente los códigos (por ejemplo, estar lo suficientemente cerca de los códigos QR).



Figura 12. Las prácticas actuales de comprometer las comunicaciones de Screen-to-eye y de Screen-to-Camera. a) Un pequeño código QR en la esquina (en el caso de un partido de hockey sobre hielo); b) Dos frames consecutivos, uno para cada canal.

Sin embargo, el anterior compromiso espacial no sólo limita la capacidad de transporte de información, sino que también implica un trabajo adicional para capturar adecuadamente los códigos, como bien se dijo en el apartado anterior. Aunque pequeño, el aspecto del código QR se sigue considerando una distracción, lo que perjudica las experiencias de los usuarios. Este último compromiso temporal produce, en efecto, un mayor tamaño de pantalla y hace que sea menos exigente en cuanto a la captura y también que sea posible transportar más bits. Sin embargo, esa práctica sólo es viable en los escenarios y en los dispositivos interactivos. En lo que cierne a los objetivos, tenemos dos concretos para nuestro caso de estudio. En primer lugar, el flujo de datos con información a través del canal *Screen-to-camera* no debería afectar a la percepción del usuario del contenido entregado a través del canal primario. En segundo lugar, tratamos de lograr una alta tasa de datos en el canal secundario de *Screen-to-camera*, lo que es vital para permitir nuevos escenarios de aplicación para la comunicación visual de datos en el futuro. Para ello, tenemos que gestionar eficazmente las interferencias del canal primario de *Screen-to-eye*. Hay que tener en cuenta que el contenido visual del canal primario inevitablemente interferirá con el canal secundario. Los elementos de la pantalla y la cámara poseen una gran cantidad de píxeles que pueden aprovecharse para enviar un elevado volumen de datos a través de códigos en 2D que varían con el tiempo (secuencia de imágenes). Por ejemplo, un usuario podría apuntar con una cámara a una pantalla de un ordenador de escritorio o a la pantalla de un smartphone que muestra el código variable en el tiempo para descargar un archivo. Investigaciones recientes han explorado más a fondo esta dirección diseñando prototipos en los que los códigos 2D variables en el tiempo pueden ser transmitidos desde los monitores [106,107] y pantallas de teléfonos inteligentes [108] a una cámara como receptor. Aunque estos trabajos han diseñado y medido la capacidad de información para una solución puntual específica, todavía no está claro cuánto se puede mejorar en estas situaciones o si hay algún límite en el rendimiento de dichos sistemas.

Por último, nuestro sistema deberá:

* **Ser discreto para los usuarios**. La comunicación de datos entre la cámara y la pantalla deberá estar completamente oculta para los usuarios y no requerirá mostrar ninguna imagen codificada visible en la pantalla. Por lo tanto, la comunicación de datos no interferirá con ningún contenido que el usuario esté viendo en la pantalla, y se podrá incorporar fácilmente a cualquier pantalla existente (por ejemplo, la de los teléfonos inteligentes, ordenadores portátiles, …) sin sacrificar su funcionalidad original, es decir, la visualización del contenido.
* **Soportar cualquier escena**. La comunicación de datos puede producirse sin importar el contenido de la pantalla, ya sea una imagen, videoclip, una película, una escena de juego, una ventana de aplicación de varias capas, o la pantalla de inicio de un teléfono inteligente o una tableta. Además, la comunicación continuará a medida que el contenido de la pantalla cambia sobre la marcha (por ejemplo, el usuario navega por una página web, juega y ve una película). Para que sea compatible con todos los tipos de contenidos, la comunicación debe ser independiente del contenido de la pantalla, que puede generarse sobre la marcha y variar con el tiempo a medida que el usuario interactúa con la pantalla.
* **Operar con dispositivos inteligentes de venta al público**. El sistema puede transmitir y recibir datos utilizando los dispositivos inteligentes existentes (por ejemplo, teléfonos inteligentes o tabletas). Estos dispositivos tienen una potencia de cálculo limitada, y sus cámaras no son tan sofisticadas como las cámaras réflex de una sola lente (SLR) de alta gama. Por lo tanto, los diseños de codificación y decodificación tienen que ser ligeros y robustos, de modo que los dispositivos inteligentes como transmisores y receptores puedan procesar los datos a la vez que soportan el contenido dinámico como el vídeo.

## **Limitaciones y soluciones del modelo**

A la hora de diseñar nuestro modelo deberemos de tener en cuenta los siguientes objetivos para realizar correctamente un sistema de comunicación S2C.

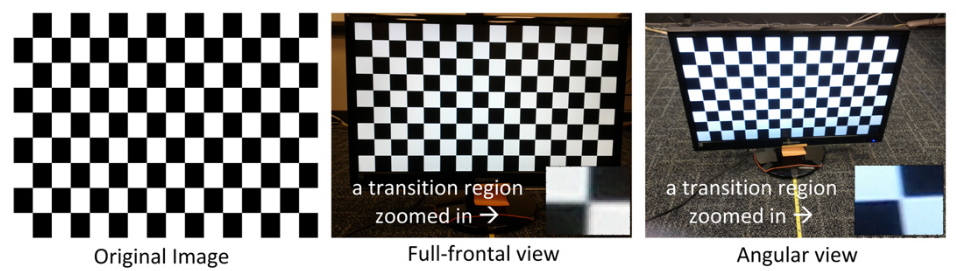
* **Ser discreto para los usuarios**. La comunicación de datos de S2C está completamente oculta a los usuarios y no requiere mostrar ninguna imagen codificada visible en la pantalla. Por lo tanto, la comunicación de datos no interfiere con ningún contenido que el usuario esté viendo en la pantalla, y se puede incorporar fácilmente a cualquier pantalla existente (por ejemplo, pantallas de teléfonos inteligentes, de ordenadores) sin sacrificar su funcionalidad original (es decir, la visualización del contenido).
* **Soporte de cualquier situación**. La comunicación de datos puede ocurrir independientemente del contenido de la pantalla, ya sea una imagen, un videoclip, una película, una escena de juego, una aplicación, o la pantalla de inicio de un teléfono inteligente. Además, la comunicación continúa a medida que el contenido de la pantalla, la comunicación debe ser independiente del contenido de la pantalla, que puede generarse sobre la marcha y variar con el tiempo a medida que el usuario interactúa con la pantalla.
* **Procesamiento de datos dinámicos en tiempo real**. El sistema puede transmitir y recibir datos sobre la marcha, en lugar de preprocesar una imagen o un archivo de vídeo para incrustar los datos. Esto es necesario cuando los datos llegan sobre la marcha o cuando el contenido de la pantalla no se conoce de antemano (por ejemplo, una escena de videojuegos. El reto es que cuando la pantalla muestra vídeo (por ejemplo, películas, videoclips), la frecuencia de fotogramas es de al menos 24 fotogramas por segundo (FPS). Por lo tanto, el sistema tiene que terminar de codificar los datos en un cuadro en un plazo de 42 ms antes de que la pantalla muestre el siguiente cuadro. Cuando el vídeo se reproduce a 60 FPS o 120 FPS, la codificación tiene que terminar dentro de 16 ms o 8 ms, por ejemplo.
* **Operar con dispositivos inteligentes**. El sistema S2C puede transmitir y recibir datos en tiempo real utilizando los dispositivos inteligentes existentes (por ejemplo, teléfonos inteligentes, tabletas). Estos dispositivos tienen una potencia de cálculo limitada, y sus cámaras no son tan sofisticadas como las cámaras réflex de una sola lente (SLR) de alta gama. Por ello, los diseños de codificación y descodificación tienen que ser ligeros y robustos, de modo que los dispositivos inteligentes como transmisores y receptores puedan procesar los datos a tiempo y al mismo tiempo soportar contenido dinámico como un video.
* **Distorsiones**. Las distorsiones que dependen de la perspectiva de la cámara se producen debido a la naturaleza del mecanismo de obtención de las imágenes en la cámara y se manifiesta como una deformación del tamaño y la forma del objeto capturado (píxel de la pantalla que emite luz) en la imagen, lo que da lugar a una compresión o ampliación visual de la proyección del objeto en la imagen. Cuando la pantalla está a una distancia de desenfoque del objetivo de la cámara (o en un ángulo oblicuo), estas distorsiones se hacen evidentes provocando una interferencia entre los píxeles adyacentes de la imagen de la cámara, lo que se denomina como interferencia entre píxeles o IPI. El efecto combinado del ruido de fondo y el IPI degradará la calidad de la señal recibida, y, por lo tanto, reducirá la capacidad de información en los canales basados en OCC. Por ejemplo, consideremos que los bloques de píxeles de una pantalla son iluminados por un patrón de tablero de ajedrez y capturados por una cámara como se muestra en la Figura 13. Podemos observar que las distorsiones de la perspectiva hacen que los píxeles de la pantalla se deformen en tamaño cuando la pantalla no está en el foco de la cámara, y en forma cuando no está alineada frontalmente.

Figura 13. Ilustración de la distorsión de la perspectiva en el canal screen-camera. Los píxeles de la pantalla visualizados son borrosos y se reduce el tamaño en la vista frontal completa y también en forma en la vista angular.

* + **Escalado de la perspectiva**. Si el píxel de la pantalla estuviera en el foco, y suponiendo que la pantalla y la cámara tienen la misma resolución, su imagen en la cámara debería ocupar la misma área que un píxel. Pero en realidad, los rayos de luz del píxel pueden no terminar exactamente en los límites del píxel de la cámara y existirán algunas zonas que acumulen interferencias. Estas áreas de desalineación y la geometría del píxel de la pantalla dependerán de la perspectiva y se debe a la distorsión debida a la escala de la perspectiva del área del píxel.
  + **Lens-blur**. Podemos observar en la Figura 13 que los píxeles de la pantalla se visualizan borrosos, especialmente en las regiones de transición entre los bloques blancos y negros. Este efecto de desenfoque se atribuye a la lente de la cámara y se denomina como desenfoque de la lente (lens-blur). Este efecto se modela típicamente utilizando la función de dispersión puntal (PSF) [131], que representa la respuesta de un sistema de imágenes a una fuente puntual. En el canal screen-camera esto se traduce como la distorsión de los píxeles en las regiones de transición entre los píxeles más brillantes (alta intensidad) y los más oscuros (baja intensidad), y conduce a la interferencia (IPI) entre los píxeles vecinos, tal y como se puede observar en la Figura 13. Dado que el área y la máxima energía que puede ser muestreada en cada píxel de la cámara es finita, la IPI conduce a una reducción efectiva de la energía de la señal por píxel.

Para resumir, los diseños de comunicación S2C discretos existentes requieren modificaciones directas de los valores de color de los píxeles de contenido (RGB) y no transmiten datos dinámicos sobre un contenido dinámico arbitrario, en el que los datos o el contenido de la pantalla llegan sobre la marcha. Además, al crear una capa opaca con valores RGB modificados, estos diseños pueden anular la optimización de la representación existente. Para mitigar el problema, los diseños actuales tendrán que renunciar a la comunicación en pantalla completa y sacrificar la velocidad de los datos o la fiabilidad de la transmisión. Además, la integración de la comunicación con el contenido de la pantalla limita en gran medida sus escenarios aplicables.

El tamaño físico del píxel de la pantalla puede no coincidir exactamente con el del sensor de imagen de la cámara. Los píxeles de la pantalla y del sensor de imagen de la cámara pueden no tener el mismo tamaño exacto, e incluso si es así, pueden no ser exactamente rectángulos similares y pueden persistir algunas micro-desviaciones que pueden causar desalineaciones en los píxeles de la imagen. Esto puede causar que un píxel de la pantalla de la imagen no se alinee con un píxel de la cámara, incluso si el píxel de la pantalla estuviera en el foco de la cámara. Tales desalineaciones causarán una desviación en el factor de distorsión para cada píxel a medida que la perspectiva cambia. Sin embargo, se puede suponer que esas desviaciones son insignificantes si se comparan con las distorsiones debidas a la escala de la perspectiva. Al considerar un factor de distorsión promedio sobre la imagen de la cámara, tales micro-desviaciones serán casi insignificantes.

Como solución se propone una comunicación desvinculada usando el canal alfa. Al operar en este canal el sistema ya no necesita modificar directamente los valores RGB de los píxeles, logrado el mismo efecto. Dado que los valores alfa son mezclados por la GPU, la codificación es casi instantánea, lo que es crítico para soportar la comunicación en tiempo real sobre cualquier contenido dinámico. En segundo lugar, al realizar la comunicación en una capa de imagen separada, el sistema hace que la comunicación discreta esté disponible universalmente y sea verdaderamente paralela al contenido que se reproduce en la pantalla, sin importar si se trata de un contenido estático, dinámico o de varias capas, e independientemente de la velocidad de fotogramas a la que se reproduce y de la resolución. Los usuarios podrán usar la pantalla tal como está, mientras que la comunicación entre la pantalla y la cámara se produce detrás de la escena en tiempo real, de forma discreta.

Finalmente, para superar estas limitaciones, se propuso un nuevo paradigma de diseño para una comunicación S2C discreta. Motivados por el hecho de que la estrecha integración de la comunicación y el contenido de la pantalla es problemática, planteamos desacoplar la comunicación S2C de las capas de las imágenes del contenido de la pantalla. Además, diseñamos una codificación de datos ligera que aprovecha la potencia de la GPU y no requiera modificaciones directas de los valores RGB del contenido. Y lo que es más importante, dado que la mezcla de alfa es una característica que ofrece la GPU, los valores alfa pueden modificarse llamando a una función del sistema sin necesidad de modificaciones a nivel del sistema operativo.

# **CAPÍTULO 5: DISEÑO DEL SISTEMA DE TRANSMISIÓN**

## **La herramienta MATLAB**

MATLAB (“Matrix Laboratory”) es un programa para realizar cálculos numéricos con vectores y matrices [69]. Como caso particular puede también trabajar con números escalares, con cadenas de caracteres y con otras estructuras de información más complejas. Como caso particular puede también trabajar con número escalares, con cadenas de caracteres y con otras estructuras de información más complejas. MATLAB tiene también un lenguaje de programación propio para desarrollar aplicaciones técnicas, fácil de utilizar y que aumenta significativamente la productividad de los programadores respecto a otros entornos de desarrollo.

MATLAB es un gran programa de cálculo técnico y científico. Para ciertas operaciones es muy rápido, cuando puede ejecutar sus funciones en código nativo con los tamaños más adecuados para aprovechar sus capacidades de vectorización. En otras aplicaciones resulta bastante más lento que el código equivalente desarrollado en C/C++. En la versión 6.5, MATLAB incorporó un acelerador JIT (Just In Time), que mejora significativamente la velocidad de ejecución de los ficheros **\*.m** en ciertas circunstancias, por ejemplo cuando no se hacen llamadas a otros ficheros **\*.m**, no se utilizan estructuras y clases.

Para el trabajo con imágenes, MATLAB, contiene funciones de bibliotecas, que permiten leer, modificar y crear imágenes en formato RGB, ya que cada una de estas se interpreta como una matriz tridimensional, es decir, tres matrices bidimensionales representantes de los colores rojo, verde y azul respectivamente. Por lo que un píxel de una imagen a color se pudiera interpretar como un número de 24 bits (8 bits para cada color), siendo el valor 0 la ausencia de color y el valor 255, la máxima intensidad del mismo, o presencia total del color en cuestión.

Todo esto hace que MATLAB sea la herramienta adecuada para este método a implementar, tanto por el desarrollo de las funciones, que especialmente en el proceso codificador de la información, necesitan de una ejecución rápida que incluye lectura, recorrido y análisis del fichero gráfico portador del mensaje, incluyendo una posterior modificación imperceptible de este; así como por la requerida optimización del trabajo con los bits de información, a nivel de píxel en todo el dominio espacial de la imagen.

## **Descripción del algoritmo propuesto**

El objetivo de nuestro proyecto como se ha mencionado en apartados anteriores es el de construir una comunicación simultánea entre los usuarios y los dispositivos. Se establece un canal de datos secundario encima del canal de vídeo primario, sin que ello produzca una interferencia en la experiencia de visualización del usuario, pero sin dejar de poder transmitir la información. En nuestro escenario de uso, una cámara captura los datos de la pantalla, mientras el usuario está viendo un programa de televisión sin una degradación notable de la calidad.

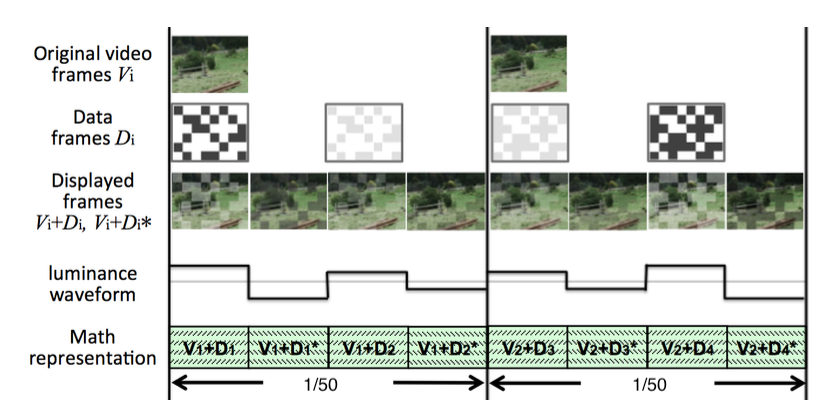


Figura 14. Diseño de nuestro sistema. Usamos una pantalla de 60 Hz y vídeos de 50 FPS. Vi y Di representan el i-ésimo frame del vídeo original y de datos.

La idea clave de nuestra propuesta es explotar la brecha de percepción entre la visión humana y los modernos sistemas de cámaras. La Figura 14ilustra el diseño general. Asegura dos objetivos de diseño en las comunicaciones screen-camera y screen-eye. Primero, nuestro sistema protege la experiencia de visualización de vídeo de la pantalla para los usuarios, mientras que incrusta los bits de datos. Segundo, permite la comunicación de datos bajo las restricciones impuestas por el canal de vídeo primario.

Por consiguiente, el algoritmo propuesto requiere de dos partes esenciales: el codificador esteganográfico, que se encarga de la introducción del mensaje dentro de la imagen utilizada como cubierta y que realiza extensos análisis y comparaciones para la toma de decisiones de sus procesos y por lo tanto, necesita un mayor tiempo de ejecución; y el decodificador esteganográfico, mecanismo inverso, pero mucho más sencillo y rápido, pues solamente requiere de una o más llaves, enviadas a través de otras vías, para el descifrado de la información útil.

En la descripción siguiente se utiliza el término de píxel para señalar un valor entre 0 y 255, como si se tratase de una imagen a escala de grises, para una mejor compresión del algoritmo, pero hay que tener en cuenta que el programa trabaja con imágenes a color RGB, que son matrices de 3 dimensiones, por lo que realmente el píxel está formado por tres valores entre 0 y 255 (24 bits) representantes del nivel de luminancia de cada color.

* **Concepto de frame complementario**. Siguiendo la propiedad de filtro paso bajo del sistema visual humano, si colocamos dos frames de vídeo seguidos, uno detrás del otro, siendo uno de ellos el inverso del otro, percibiremos un cuadro de color promedio uniforme. Esta observación origina los frames complementarios, concepto central de nuestro sistema. Primero, definimos los píxeles complementarios. Si dos píxeles **p** y **p\*** se complementan entre sí con respecto al nivel de luminancia **v**, si sus valores de píxeles suman **2v**, es decir, **vp + vp\* = 2v**. En consecuencia, dos frames **P** y **P\*** serán cuadros complementarios con respecto al nivel de luminancia **v**, si todos sus píxeles son complementarios respecto a este. Es fácil ver que, después de un filtrado paso bajo, dos frames complementarios producen frames con un nivel de luminosidad medio **v**. Los cuadros complementarios permiten así agrupar adecuadamente los frames de vídeo y los frames de datos (por ejemplo, V ± D) sin producir distorsiones de color y cambios notables en el nivel de luminancia del vídeo.
* **Fusión de parpadeo.** Se presenta una nueva forma de incrustar datos en los frames de vídeo. Gracias al CAPÍTULO 2: EFECTOS DE LA PERCEPCIÓN VISUAL HUMANA aprendimos que el sistema visual humano es propenso a parpadeos y a la interferencia visible. Aplicando el concepto explicado con anterioridad, generamos dos frames de datos más vídeo (**Vi+Di**, **Vi+Di\***) con respecto al frame de vídeo original **Vi**. Así mismo, obtenemos cada frame de datos **Di**, calculamos su frame complementario **Di\***, y luego integramos ambos en el frame del vídeo original. Nuestro estudio muestra que este esquema es efectivo para lograr un efecto insignificante en las experiencias de visualización del usuario.
* Suavizado de bloques.
* **Canal Alpha**. Es una técnica estandarizada ampliamente utilizada en los gráficos por ordenador. El canal alpha se ha utilizado para formar una imagen compuesta con transparencia parcial o total [103]. Mientras que el elemento de la imagen almacena los valores de intensidad de color en los diodos rojos, verdes y azules (RGB) para cada píxel de la imagen en primer plano [104], el canal alfa almacena un valor (α) entre 0 y 1 para indicar la translucidez del píxel. Un valor de α de 0 significa que el píxel de la imagen es totalmente transparente, y 1 (valor por defecto) significa que el píxel es totalmente opaco. Dado que las pantallas LED y OLED de hoy en día utilizan un negro mate como imagen de fondo, un valor α de 0 significa que el píxel se atenúa a negro, y un valor de α de 1 significa que se muestra la intensidad de color original del píxel. Por lo tanto, disminuir el valor de α de un píxel esencialmente oscurece el píxel. Como ejemplo, la Figura 14(a)muestra la pantalla de un Samsung Note 3, donde se establece uniformemente los valores de α de la parte derecha de la pantalla en 0,5, lo que resulta en una apariencia más oscura.

Este efecto de oscurecimiento se percibe cuando la intensidad del color (suma de los valores de los canales RGB) cambia por la cámara (receptor). Para examinar la intensidad de color resultante percibida por las cámaras del mercado, se configuró un teléfono Samsung Note 3 como transmisor y un teléfono Samsung S5 a 15 cm de distancia como receptor [105]. Se ajustó los valores de α de todos los píxeles de la pantalla del Note 3, y se calculó la intensidad del color en cada píxel usando el marco capturado por la cámara del S5. Como se muestra en la Figura 14(b), el aumento del valor de α lleva a un aumento lineal de la intensidad del color percibida por la cámara del receptor. Esto se debe a que el valor de α se utiliza para multiplicar los valores de color RGB. Lo más importante es que, incluso para el cambio de valor de α de 0,01 (α = 0,99), que no es perceptible por el ojo humano (Figura 14(a)), la cámara de S5 es capaz de detectar el cambio de intensidad del color. Los resultados demuestran el potencial de los dispositivos inteligentes equipados con cámaras para decodificar la información incorporada en los cambios de translucidez de los píxeles.

Figura 15. El impacto de cambiar los valores de los píxeles de α. (a) Muestra una imagen con diferentes valores de α: el valor predeterminado 1 (la izquierda), 0,99 (el centro) y 0,5 (la derecha), utilizando un Samsung Note 3; (b) Muestra la intensidad de color del cuadro capturado por el Samsung S5 al variar los valores de α en la pantalla del transmisor.

Motivado por las observaciones anteriores, nuestro sistema codifica la información en la transparencia de los píxeles (α) y la cambia en la pantalla del transmisor. Para decodificar la información, los dispositivos inteligentes equipados con cámaras examinan los cambios de la intensidad de color percibida en los frames capturados. Para hacer que los cambios de luminosidad sean más precisos, el transmisor tiene que ser capaz de leer la información en la pantalla. Para hacer que dichos cambios sean imperceptibles para el ojo humano, se debe de cambiar el valor de α en 0,01 (α = 0,99) para atenuar un píxel, y la translucidez cambia a medida que se actualiza la pantalla. Las pantallas LED o OLED de los smartphones actuales admiten una frecuencia de actualización de 60 Hz, que es suficiente para evitar el efecto de parpadeo directo cuando un usuario ve una imagen estática. Además, podemos dividir la pantalla en cuadrículas, y cambiar los valores de α de todos los píxeles de cada cuadrícula de forma independiente. Por lo tanto, cada cuadrícula funciona como un elemento transmisor independiente, resultando en un canal MIMO entre la pantalla y las cámaras. En general, operando en el canal alfa, nuestro sistema proporciona un canal de comunicación discreto para los usuarios con baja codificación. Además, al dividir la pantalla en pequeñas cuadrículas, se reduce el impacto de la sincronización vertical del frame, porque cada cuadrícula opera en forma dependiente y no todas las cuadrículas contienen píxeles de cuadro no sincronizados.

* **Transmisión**. El sistema de codificación tiene como cometido tomar el flujo de datos de entrada, distribuirlos en bloques y combinar dichos bloques con los diferentes frames de una señal de vídeo, para de esta manera, integrar la transmisión de la información en la emisión de imágenes emitidas por una pantalla. El objetivo es que la señal de datos quede embebida en las imágenes emitidas de una forma inapreciable por el espectador, como se ha comentado en los apartados anteriores de este trabajo. El conjunto de bloques en los que se ha desarrollado el codificador se detalla a continuación:
  + **Encoder.** Se encarga de codificar los datos entrantes en códigos, que denominaremos códigos expandidos). Para ello se divide el flujo binario en cadenas de N bits (en nuestro caso 4 bits, aunque se harán pruebas con otros valores) y las expande, siguiendo estrategias de Espectro Ensanchado por Secuencia Directa (DSSS de sus siglas en inglés, Direct Sequency Spread Spectrum). El concepto de expansión hace referencia a que para cada bit de la cadena se generan M bits. Cada código, por tanto, representa un bit, y se sitúa en una celda de la imagen. De ahí que el código expandido tenga una longitud M de 16 bits. Por tanto, si se codifican 4 bits, se tienen 4 celdas independientes en la imagen, y si cada celda cuenta con 16 ranuras, la longitud del código final será de . Estos códigos de 4 bits (que corresponden a una celda) siguen dos configuraciones: el código seleccionado para el bit 1 será el que corresponde espacialmente a la configuración del tablero de ajedrez (Figura 15(a)), mientras que para el bit 0, el que corresponde a la conformación en cruz (Figura 15(b)).

****

Figura 16. Códigos expandidos utilizados para la transmisión. (a) Código para el 1, (b) Código para el 0.

* + **Masker**. Este apartado se encarga de adaptar los códigos expandidos a la imagen en sí. Dicho con otras palabras, este bloque se encarga de generar una máscara 2D que tiene las mismas dimensiones que la imagen. Para la generación de estas máscaras el bloque grid, que representa un grid de colocación de las celdas en la imagen (cada celda será la representación 2D de un bit del código correspondiente). Esta máscara deberá pasar por una serie de rutinas de filtrado que permitan reducir la variabilidad espacial y temporal del código, esto es, que permitan suavizar los bordes y las transiciones de las celdas, de tal forma que se minimice la capacidad de detección de los códigos por el ojo humano.
  + **Blender**. Se encarga de mezclar la imagen con la máscara generada por el Masker. En este caso, se utiliza una mezcla aditiva. Este tipo de mezcla se explica como la combinación de determinadas cantidades de luz (rojo, verde y azul), con objeto de crear nuevas composiciones de colores. Si se mezclan las tres fuentes de luz en su máxima intensidad, el ojo humano percibirá el color blanco como resultado. La mezcla de los mismos colores con menor intensidad se percibirá como un gris neutro y si se apagan las tres fuentes se logra el negro. Por otra parte, si sólo una de las tres fuentes de luz está apagada y las otras dos emiten con su intensidad máxima, se pueden obtener mezclas como el amarillo (rojo y verde), el cian (azul y verde) o el magenta (rojo y azul). Las distintas combinaciones de dos o tres colores primarios de fuentes luminosas en sus diferentes intensidades permiten reproducir en cualquier pantalla la mayoría de los colores. Por tanto, la mezcla aditiva se utiliza en los monitores de ordenadores, televisores, tabletas, etc. La pantalla de un monitor está compuesta por un cierto número de píxeles, y cada píxel contiene tres pequeñas fuentes luminosas RGB. La mezcla de los colores de estas tres fuentes luminosas le da al píxel su color específico. Por ende, la imagen generada será la imagen original más la máscara superpuesta, la cual está multiplicada por un cierto factor, el factor alpha, α. Este factor, tal y como se detalló anteriormente, es un parámetro de crucial importancia, puesto que juegan un papel determinante tanto para que los códigos puedan ser detectados por la cámara, como para que éstos no sean apreciables por el ojo humano. La discusión del valor de este factor se detallará en los siguientes apartados.
* **Recepción**. El proceso de recepción se puede dividir en dos etapas, descubrimiento y adquisición. El receptor utilizará la longitud del código del transmisor para diferenciar las diferentes tramas de frames de información. En la etapa de descubrimiento, el receptor correlaciona la información que tiene del transmisor como una plantilla y si excede de un valor umbral de detección impuesto, se considera que la fuente de información no habrá sido detectada con éxito. Por otro lado, en la etapa de adquisición, el receptor correlaciona la plantilla con la ROI. Cuando se encuentra una plantilla, se sincroniza. Esto reduce la carga de cálculo necesaria para el preprocesamiento de toda la imagen. Finalmente, se realiza el proceso de binarización y el ensamblaje de datos. A continuación, se exponen algunas consideraciones que deben de tenerse en cuenta a la hora de implementar la recepción de la señal:
  + **Discusión sobre el desenfoque del movimiento**. Las aplicaciones de comunicación Screen-to-camera normalmente implican un cierto grado de movimiento, por ejemplo, cuando la cámara se sostiene en la mano, o cuando la pantalla o la cámara están situadas sobre un objeto en movimiento. El movimiento debido a las sacudidas de la mano o a los movimientos laterales puede causar un cambio drástico de enfoque entre la pantalla y la cámara. En tal caso, se pueden asumir algunas oscilaciones en los píxeles, especialmente cuando la cámara no es estable, donde los píxeles parecen interferir entre sí, causando eventualmente un efecto visual borroso en la imagen, lo que se conoce como motion-blur.

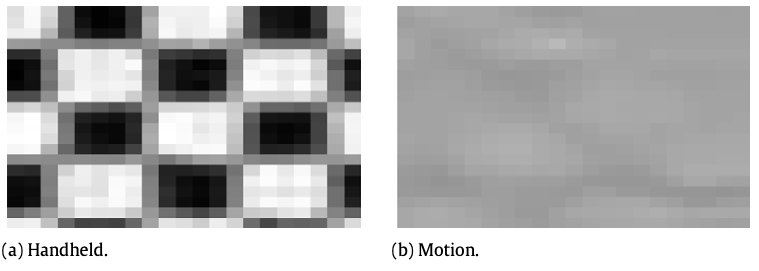
El desenfoque por movimiento surge principalmente debido al movimiento que se produce entre los frames de la cámara o entre ellos mismos. Las cámaras de los teléfonos inteligentes se sostienen normalmente con la mano y las vibraciones causadas por el movimiento de la mano pueden causar el desenfoque por movimiento, pero suelen ser mucho menores que las que se producen cuando la pantalla o la cámara están en movimiento. Las cámaras equipadas en vehículos pueden sufrir más borrosidad en comparación con los escenarios con la mano, ya que el muestreo de la cámara puede ser demasiado lento cuando se compara con la velocidad del movimiento.

Figura 17. Ilustración del desenfoque por movimiento en imágenes de una pantalla que muestra un patrón de tablero de ajedrez, tomadas por una cámara estática (a) y cuando está en movimiento (b).

Las cámaras de hoy en día están equipadas con una capacidad de compensación de movimiento muy efectiva que compensa el desenfoque de movimiento a través de un mecanismo de filtrado llamado de-blurring. Esta tecnología [112,113] es una técnica que se utiliza comúnmente para mitigar el efecto de desenfoque en la imagen aplicando un filtro que invierte el efecto de desenfoque en la imagen. La calidad de la imagen de-blurred dependerá en gran medida de la eficacia del filtro, así como de la cantidad de movimiento/oscilación inducida en los píxeles. Las imperfecciones en el proceso de de-blurring también pueden conducir a la reducción de la calidad de la señal en comparación con un escenario ideal (pantalla y cámara estáticas). Si el movimiento es rápido, entonces la cámara puede no ser capaz de exponer la totalidad de los píxeles de la pantalla, y por tanto, hacer que la energía de la señal se extienda a muchos píxeles y resulte una imagen más borrosa, como se muestra en la Figura 16 (b) a modo de ejemplo ilustrativo. Podemos observar en estas imágenes las distorsiones debidas al desenfoque por movimiento que llevan a la interferencia entre píxeles.

## **Diagrama de flujo del programa**

Con todos estos bloques, el sistema a tratar coordina la generación de los vídeos siguiendo el siguiente flujo, el cual puede ser dividido en tres fases: fase de obtención de los códigos, fase de generación de máscaras y fase de codificación de imagen.

## **Consideraciones del transmisor**

* **Determinación del factor α.** La invisibilidad es el objetivo fundamental del diseño de un sistema screen-to-camera, por lo que primero que se deberá de hacer será la evaluación de la calidad perceptiva de los contenidos incrustados en el vídeo. Para una mejor decodificación por parte del usuario, se querrá maximizar el umbral de cambio entre frames (αmax). Pero un valor de α tan grande también aumentará la posibilidad del efecto de flicker o parpadeo. Queremos encontrar los valores máximos en los que alpha en cada frame no sea capaz de producir parpadeo, pero si una buena decodificación.

Para la determinación del factor α, se han de tener en cuenta diversas consideraciones. Por un lado, es necesario determinar la mínima variación de intensidad en la imagen detectable por la cámara, de tal forma, que la cámara sea capaz de captar las variaciones en la imagen producidas por la codificación de los datos, y, por tanto, sea capaz de decodificar la información.

* **Signal-to-interference-ratio (SIR) en canales Screen-to-camera**. Este apartado se encarga de cuantificar la calidad de la señal en el transmisor usando un promedio de SIR por símbolo para determinar si el símbolo puede ser codificado o no a partir de un umbral dado.
* **Determinación del número de frames por símbolo.**

## **Consideraciones del receptor**

## **Estructura de los frames**

## **Técnicas de codificación empleadas**

## 

* Código Hadamard.
* **Codificación Diferencial**.
* Multiplexación espacial de canales tiempo/frecuencia.
* Técnicas de espectro ensanchado.

## **Creación del código**

## **Algoritmo de decisión**

## **Codificación y multiplexación**

* Codificador espacial
* Codificador temporal

## **Implementación**

# **CAPÍTULO 6: VALIDACIÓN EXPERIMENTAL Y RESULTADOS**

En esta sección describimos los experimentos que realizamos para validar nuestro modelo de canal Screen-to-camera. El motivo principal de estos experimentos fue determinar la capacidad y la calidad de un canal de estas características. Dentro de los trabajos llevados a cabo, se ha realizado una experimentación preliminar del funcionamiento del sistema, que se realizó según la configuración que se muestra en la **FIG**. En esta figura se observa que la pantalla está a una distancia de XX metros, y el móvil se encuentra en una posición central enfrentado al monitor. En este experimento se utilizaron 4 vídeos codificados y comprimidos a 50 FPS de una duración de 15 segundos, respetando el Teorema del Muestreo de Nyquist, y se grabaron con una cámara convencional del móvil Huawei Mate 10.

FOTO DEL EXPERIMENTO

El conjunto de experimentos, como se muestra en la **Fig**, consistió en un monitor con retroiluminación por LED de 13 pulgadas (Macbook Air del 2015) de resolución 1440x900 píxeles, que sirvió como transmisor de pantalla y dos cámaras de 12 y 20 MP con posibilidad de grabación a 1080p @30 fps, integrada en el Huawei Mate 10 que sirvió como receptor. La configuración de la exposición y el balance de blancos de la cámara se configuran de forma automática (configuración predeterminada en los dispositivos Android). Todos los experimentos se llevaron a cabo bajo las mismas condiciones de iluminación ambiental con las mediciones tomadas en el interior de un laboratorio equipado con iluminación fluorescente en el techo. Fijamos la cámara en un trípode para asegurar la menor cantidad de error en la medición de la distancia y el ángulo entre la pantalla y la cámara.

En la **figura** se muestra el resultado tras la diferencia de dos imágenes consecutivas. Sin embargo, pese al ruido existente, la señal, gracias a que está codificada siguiendo las estrategias de espectro ensanchado puede ser recuperada.

FOTO DE LA DIFERENCIA DE IMÁGENES

Pedimos a 5 probadores que nos den su opinión sobre la calidad perceptiva de los contenidos incrustados de UnseenCode bajo diferentes configuraciones de framerate de pantalla. El contenido original y el contenido incrustado se muestran uno al lado del otro. Los probadores deben informar si pueden observar parpadeo, y si no, si hay otras diferencias entre el contenido original y el contenido incrustado.

El resultado se da en Tab. IV. En la mayoría de los ajustes, el BER es inferior al 5%, y podemos esperar un buen rendimiento de decodificación con un código de corrección de errores adecuado. Un BER alto puede ocurrir en las siguientes condiciones:

- El contenido de la pantalla es complejo y/o dinámico. Para la imagen de la antena, el BER es más alto que el de la escena del documento lisa y monocromática. Y para una escena de video dinámica, el BER es aún mayor.

- El tiempo de exposición es demasiado largo en relación con el intervalo de cuadros, como el ajuste de 144 FPS y 1/30s. En este caso, la diferencia en el tiempo de exposición entre cada cuadro de un par de fusión no es significativa, y la imagen del código de barras está sumergida en el ruido.

- El valor de alpha es pequeño y en este caso, la amplitud de la diferencia entre frames es demasiado débil para ser capturada.

1. **Consideraciones prácticas**

Probamos 6 vídeos.

1. **Percepción del usuario**

# **CONCLUSIONES Y RECOMENDACIONES**

En este trabajo se resumen las características del diseño de la arquitectura propuesta para la codificación de los vídeos originales con cadenas aleatorias de datos de XX bits. En este documento también se resumen los desafíos a los que se ha tenido que enfrentar esta solución para poder ser viable y a las limitaciones necesarias para que pueda funcionar correctamente. Concretamente, para poder aumentar el factor alpha de los códigos incrustados ha sido necesario aplicar filtrados gaussianos a las máscaras para reducir su variabilidad paulatina entre la transmisión de códigos consecutivos. De esta forma, las variaciones en la imagen (relativas a los códigos incrustados) se hacen imperceptibles al espectador, pero a su vez pueden se detectadas por la cámara tal y como se expone en la sección anterior. Además, la elección de los códigos utilizados para cada bit de transmisión no es arbitraria, sino que ha sido cuidadosamente evaluada y pretende minimizar las variaciones espaciales para códigos consecutivos, haciendo que la transmisión pase de forma inadvertida al usuario.

# **BIBLIOGRAFÍA REFERENCIADA**

1. Cisco Corporation, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018, 2014.
2. Parth H. Pathak, Xiaotao Feng, Pengfei Hu, y Prasant Mohapatra. (2015). “Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges”. IEEE Communications surveys & Tutorials, Vol. 17, No. 4, Fourth Quarte, 2047-2077.
3. The Institute of Electrical and Electronics Engineers, Inc. (4 de marzo de 2009). IEEE 802.15 WPANTM Task Group 7 (TG7) Visible Light Communication.   
   Recurso digital:< <http://www.ieee802.org/15/pub/TG7.html>>.
4. Md. Shahjalal, Moh. Khalid Hasan, Mostada Zaman Chowdhury, y Yeong Min Jang, (2019). “Future Optical Camera Communication Based Applications and Opportunities for 5G and Beyond”. Department of Electronics Engineering, Kookmin University, Scoul, Korea. IEEE ICAIIC 2019.
5. Tianxing Li, Chuankai An, Xinran Xiao, Andrew T. Campbell, and Xia Zhou. 2015. Real-Time Screen-Camera Communication Behind Any Scene. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (MobiSys '15). ACM, New York, NY, USA, 197-211.
6. Porter, T., and Duff, T. Compositing digital images. In ACM SIGGRAPH *Computer Graphics* (1948).
7. M. Yuen, H. Wu, A survey of hybrid MC/DPCM/DCT video coding distortions, Signal Process. 70 (3) (1998) 247–278.
8. S. D. Perli, N. Ahmed, and D. Katabi. Pixnet: interference-free wireless links using lcd-camera pairs. In *MobiCom*, 2010.
9. T. Hao, R. Zhou, and G. Xing. Cobra: color barcode streaming for smartphone systems. In *MobiSys*, 2012.
10. Q. P. Wenjun Hu, Hao Gu. Lightsync: Unsynchronized visual communication over screen- camera links. In *MobiCom*, 2013.
11. A. Wang, S. Ma, C. Hu, J. Huai, C. Peng, and G. Shen. Enhancing reliability to boost the throughput over screen-camera links. In *MobiCom*, 2014.
12. Cisco Visual Networking Index: Forecast and Methodology, 2013–2018.   
     Recurso digital: <<http://tinyurl.com/mev32z8>>.
13. D. A. Atchison, G. Smith, and G. Smith. *Optics of the human eye*. Butterworth-Heinemann Oxford, UK:, 2000.
14. R. D. Valois and K. D. Valois. Spatial Vision. Oxford University Press, 1988.
15. D. Kelly. Flicker. In Visual psychophysics, pages 273–302, 1972.
16. E. Simonson and J. Brožek. Flicker fusion frequency: background and applications. Physiological reviews, 1952.
17. G. Brindley, J. Du Croz, and W. Rushton. The flicker fusion frequency of the blue-sensitive mechanism of colour vision. The Journal of physiology, 183(2):497–500, 1966.
18. D. G. Green. Sinusoidal flicker characteristics of the color-sensitive mechanisms of the eye. *Vision research*, 9(5):591–601, 1969.
19. W. A. Hershberger and J. S. Jordan. The phantom array: a perisaccadic illusion of visual direction. The Psychological Record, 48(1):2, 2012.
20. J. Roberts and A. Wilkins. Flicker can be perceived during saccades at frequencies in excess of 1 khz. *Lighting Research and Technology*, 45(1):124–132, 2013.
21. I. Vogels and I. Hernando. Effect of eye movements on perception of temporally modulated light. Recurso digital: <<http://2012.experiencinglight.nl/doc/28.pdf>>.
22. Kandel RE, Schwarts J, Jessell TM. Principles of neural science. 3 ed. New York: Elsevier Science Publishing; 1991. Chapter 28. Phototransduction and information processing in the retina.
23. Lamme AFV, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci.
24. Supèr H. Figure-ground activity in V1 and guidance of saccadic eye movements. J. Physiol.
25. Supèr H. Working memory in the primary visual cortex. Arch. Neurol.
26. Perez D, Travieso D, Magano A, Lopez A, Marquez MT. Precepción visual. Tema 1.
27. Campbell FW, Green DG. Optical and retinal factors affecting visual resolution. J Physiol.
28. Artal P, Navarro R. Monochromatic modulation transfer function of the human eye for different pupil diameters: an analytical expression. J Opt Soc Am A.
29. Navarro R, Artal P, Williams DR. Modulation transfer of the human eye as a function of retinal eccentricity. J Opt Soc A.
30. Williams DR, Artal P, McMahon MJ, Navarro R, Brainard DH. Off- axis optical quality and retinal sampling in the human eye. Vision Res.
31. Wang YZ, Thibos LN, Bradley A. Effects of refractive error on detection acuity and resolution acuity in peripheral vision. Invest Ophthalmol Vis Sci.
32. B. Breitmeyer, H. Ogmen, Visual Masking: Time Slices Through Conscious and Unconscious Vision, Oxford University Press, New York, NY, USA, 2006.
33. A.C. Bovik, Automatic prediction of perceptual image and video quality, Proc. IEEE 101.
34. Z. Wang, A.C. Bovik, Reduced- and no-reference image quality assessment, IEEE Signal Process. Mag.
35. G.E. Legge, J.M. Foley, Contrast masking in human vision, J. Opt. Soc. Amer.
36. Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process.
37. S.J. Daly, Visible differences predictor: An algorithm for the assessment of image fidelity, in: Proc. SPIE Human Vis. Visual Process. and Digital Display III.
38. D.J. Heeger, Normalization of cell responses in cat striate cortex, Visual Neurosci.
39. P.C. Teo, D.J. Heeger, Perceptual image distortion, in: Proc. SPIE Human Vision, Visual Process., and Digital Display V, vol. 2179.
40. G. Sperling, Temporal and spatial visual masking. I. Masking by impulse flashes, J. Opt. Soc. Amer.
41. B.E. Rogowitz, Spatial/temporal interactions: Backward and forward metacontrast masking with sine-wave gratings, Vis. Res.
42. F. Hermens, G. Luksys, W. Gerstner, M. Herzog, U. Ernst, Modeling spatial and temporal aspects of visual backward masking, Psychol.
43. D.J. Simons, R.A. Rensink, Change blindness: Past, present, and future, Trends Cogn.
44. D.M. Levi, Crowding—an essential bottleneck for object recognition.
45. J.W. Suchow, G.A. Alvarez, Motion silences awareness of visual change, Curr.
46. L.K. Choi, A.C. Bovik, L.K. Cormack, Spatiotemporal flicker detector model of motion silencing, Perception.
47. L.K. Choi, A.C. Bovik, L.K. Cormack, The effect of eccentricity and spatiotemporal energy on motion silencing, J. Vis.
48. A.J. Seyler, Z. Budrikis, Detail perception after scene changes in television image presentations, IEEE Trans.
49. A.N. Netravali, B. Prasada, Adaptive quantization of picture signals using spatial masking.
50. B.G. Haskell, F.W. Mounts, J.C. Candy, Interframe coding of videotelephone pictures.
51. A. Puri, R. Aravind, Motion-compensated video with adaptive perceptual quantization.
52. B. Girod, The information theoretical significance of spatial and temporal masking in video signals, in: Proc. SPIE Human Vis. Visual Process. and Digital Display.
53. D. Johnston, S.C. Knauer, K.N. Matthews, A.N. Netravali, E.D. Petajan, R.J. Safranek, P.H. Westerink, Adaptive Non-Linear Quantizer, 1992.
54. C.H. Chou, C.W. Chen, A perceptually optimized 3-D subband codec for video communication over wireless channels, IEEE Trans. Circuits Syst. Video Technol. 6 (2) (1996) 143–156.
55. Z. Chen, C. Guillemot, Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model, IEEE Trans. Circuits Syst. Video Technol. 20 (6) (2010) 806–819.
56. J.Lubin, D.Fibush,SarnoffJNDvisionmodel,in:T1A1.5WorkingGroupDocument #97-612, ANSI T1 Standards Committee, 1997.
57. X. Fan, W. Gao, Y. Lu, D. Zhao, Flickering reduction in all intra frame coding, in: Proc. JVT-E070, Joint Video Team of ISO/IEC MPEG & ITU-T VCEG Meeting, 2002.
58. E. Gelasca, T. Ebrahimi, On evaluating video object segmentation quality: A perceptually driven objective metric, IEEE J. Sel. Top. Signal Process. 3 (2) (2009) 319–335.
59. L.K. Choi, L.K. Cormack, A.C. Bovik, On the visibility of flicker distortions in naturalistic videos, in: Proc. 5th Int. Workshop Quality of Multimedia Experience, QoMEX, 2013. pp. 164–169.
60. L.K. Choi, L.K. Cormack, A.C. Bovik, Eccentricity effect of motion silencing on naturalistic videos, in: Proc. IEEE 3*𝑟𝑑* Global Conf. Sig. and Inf. Process., GlobalSIP, 2015, pp. 1190 1194.
61. R. Blake, R. Sekuler, Perception, fifth ed., McGraw-Hill, New York, NY, USA, 2006.
62. M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant, N.C. Rust, Do we know what the early visual system does?, J. Neurosci. 25 (46) (2005) 10577–10597.
63. E.P. Simoncelli, D.J. Heeger, A model of neuronal responses in visual area MT, Vis. Res. 38 (5) (1998) 743–761.
64. A.B. Watson, A.J. Ahumada, Model of human visual-motion sensing, J. Opt. Soc. Amer. A 2 (2) (1985) 322–342.
65. D. Master, "Introducción a la esteganografía". 2004.
66. R. G. Cárdenas, "Esteganografía lingüística". 2007.
67. K.M.Sullivan, "Image steganalysis: hunting & escaping". 2005.   
    Recurso digital:< <http://vision.ece.ucsb.edu/publications/05ThesisSullivan.pdf>>.
68. F. Diatel, "Esteganografía y marcas de agua". 2002.
69. J. García, J. I. Rodríguez & J. Vidal, "Aprenda Matlab 7.0 como si estuviera en primero". Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid. Diciembre 2005.
70. Ericsson, Ericsson mobility report, Tech. rep., Ericsson (2018).
71. S. Wu, H. Wang, C. H. Youn, Visible light communications for 5Gwireless networking systems: From fixed to mobile communications,IEEE Netw. 28 (6) (2014) 41–45.
72. P. H. Pathak, X. Feng, P. Hu, P. Mohapatra, Visible light communi-cation, networking, and sensing: A survey, potential and challenges,IEEE Commun. Surveys Tuts. 17 (4) (2015) 2047–2077.
73. Z. Ghassemlooy, P. Luo, S. Zvanovec, Optical camera communications,in: Optical Wireless Communications, Springer, 2016, pp. 547–568.
74. M. Z. Chowdhury, M. T. Hossan, A. Islam, Y. M. Jang, A comparativesurvey of optical wireless technologies: Architectures and applications,IEEE Access 6 (2018) 9819–9840.
75. J. Hao, Y. Yang, J. Luo, CeilingCast: Energy efficient and location-bound broadcast through LED-camera communication, in: Proc. Info.Commun. Conf., (INFOCOM), IEEE, 2016, pp. 1–9.
76. R. Boubezari, H. Le Minh, Z. Ghassemlooy, A. Bouridane, Smartphonecamera based visible light communication, J. Lightw. Technol. 34 (17)(2016) 4121–4127.
77. T. Nguyen, A. Islam, T. Yamazato, Y. M. Jang, Technical issues onIEEE 802.15. 7m image sensor communication standardization, IEEECommun. Mag. 56 (2) (2018) 213–218.
78. A. Ashok, M. Gruteser, N. Mandayam, J. Silva, M. Varga, K. Dana,Challenge: Mobile optical networks through visual MIMO, in: Proc. ofMobiCom/MobiHoc, ACM, 2010, pp. 105–112.
79. E. Wengrowski, W. Yuan, K. J. Dana, A. Ashok, M. Gruteser, N. Man-dayam, Optimal radiometric calibration for camera-display communi-cation, in: IEEE Winter Conf. on App. of Computer Vision, 2016, pp.1–10.
80. N.-T. Le, Y. M. Jang, MIMO architecture for optical camera commu-nications, The J. of the Korean Institute of Commun. Sci. 42 (1) (2017)8–13.
81. Y.Jiang,K.Zhou,andS.He,“Humanvisualcortexrespondstoinvisible chromatic flicker,” Nature Neuroscience, vol. 10, no. 5, p. 657, 2007.
82. A.Eisen-Enosh,N.Farah,Z.Burgansky-Eliash,U.Polat,andY.Mandel, “Evaluation of critical flicker-fusion frequency measurement methods for the investigation of visual temporal resolution,” Scientific Reports, vol. 7, no. 1, p. 15621, 2017.
83. M. Menozzi, F. Lang, U. Naepflin, C. Zeller, and H. Krueger, “CRT versus LCD: Effects of refresh rate, display technology and background luminance in visual performance,” Displays, vol. 22, no. 3, pp. 79–85, 2001.
84. N. T. Le, M. A. Hossain, and Y. M. Jang, “A survey of design and implementation for optical camera communication,” Signal Process. Image Commun. 53, 95–109 (2017).
85. Z. Ghassemlooy, P. Luo, and S. Zvanovec, “Optical Camera Communications,” in Optical Wireless Communications: An Emerging Technology, M. Uysal, C. Capsoni, Z. Ghassemlooy, A. Boucouvalas, and E. Udvary, eds. (Springer, 2016), pp. 547–568.
86. Mannos J., Sakrison D.: The effects of a visual fidelity criterion of the encoding of images. IEEE Transactions on Information Theory 20, 4 (1974), 525-536.
87. Glassner A.: Principles of digital image synthesis. Morgan Kaufmann Pub, 1995.
88. Webb P.: Bioastronautics data book. Scientific and Technical Information Division, National Aeronautics and Space Administration, 1964.
89. Campbell F., Robson J.: Application of Fourier analysis to the visibility of gratings. The Journal of Physiology 197,3 (1968), 551.
90. Enroth-Cugell C., Robson J.: The contrast sensitivity of retinal ganglion cells of the cat. The Journal of Physiology 187,3 (1996), 517.
91. Livingstone M.: Art, illusion and the visual system. Scientific American 258, 1 (1988), 78.
92. Michelson A.: A 1927 Studies in Optics. Scientific American (1927).
93. Li R., Polat U., Makous W., Bavelier D.: Enhancing the contrast sensitivity function through action video game training. Nature Neuroscience 12,5 (2009), 549-551.
94. Murphy B.: Pattern thresholds for moving and stationary gratings during smooth eye movement. Vision Research 18,5 (1978), 521 – 530.
95. Tyler C.: Analysis of visual modulation sensitivity. II Peripheral retina and the role of photoreceptor dimensions. Journal of the Optical Society of America A 2,3 (1985), 393 - 398.
96. Nakayama K.: Properties of early motion processing: Implications for the sensing of egomotion. The Perception and Control of Self Motion (1990), 69 – 80.
97. Dally S.: Engineering observations from spatiovelocity and spatiotemporal visual models. Vision Models and Applications to Image and Video Processing (1998), 179 – 200.
98. Reddy M.: Perceptually modulated level of detail for virtual environments. University of Edinburgh (1997).
99. I.J. Cox et al., Digital Watermarking and Steganography, 2nd ed., Burlington: Morgan Kaufmann, 2008.
100. P. Sallee, “Model-Based Steganography,” Digit. Watermarking, LNCS, Seoul, Rep. of Korea, vol. 2939, Oct. 20–22, 2004, pp. 154–167.
101. J. Fridrich, “Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes,” Proc. IH, LNCS, Toronto, Canada, vol. 3200, May 23–25, 2005, pp. 67–81.
102. R. Bohme and A. Westfeld, “Breaking Cauchy Model-Based JPEG Steganography with First Order Statistics,” Proc. ESORICS, LNCS, vol. 3193, 2004, pp. 125–140.
103. PORTER, T., AND DUFF, T. Compositing digital images. In *ACM Siggraph Computer Graphics* (1984).
104. TAN, K. W., ET AL. FOCUS: a usable & effective approach to OLED display power management. In UbiComp (2013).
105. Li, T., An, C., Campbell, A., Zhou, X., 2014. HiLight.
106. S.D. Perli, N. Ahmed, D. Katabi, PixNet: Interference-free wireless links using LCD-camera pairs, in: Proceedings of MobiCom’10, ACM, New York, NY, USA, 2010, pp. 137–148.
107. S. Hranilovic, F. Kschischang, A pixelated-MIMO wireless optical communication system, IEEE J. Sel. Top. Quantum Electron. 12 (4) (2006) 859–874.
108. T. Hao, R. Zhou, G. Xing, COBRA: color barcode streaming for smartphone systems, in: Proceedings of MobiSys’12, ACM, New York, NY, USA, 2012, pp. 85–98.
109. A. Goldsmith, Wireless Communications. Cambridge, 2005.
110. T.S. Lomheim, G.C. Holst, CMOS/CCD Sensors and Camera Systems, second ed., The International Society for Optical Engineering (SPIE), 2011.
111. A.Tang, J.Kahn, K.-P.Ho, Wireless infrared communication links using multi-beam transmitters and imaging receivers, in:ICC96, Conference Record, Converging Technologies for Tomorrow’s Applications, vol. 1, June 1996, pp. 180–186.
112. S.K. Nayar, M. Ben-Ezra, Motion-based motion deblurring, IEEE Trans. Pattern Anal. Mach. Intell. 26 (6) (2004) 689–698.
113. Jia Chen, Lu Yuan, Chi-Keung Tang, Long Quan, Robust dual motion deblurring, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008, IEEE Conference on, June 2008, pp. 1–8.
114. E. Reinhard, E. A. Khan, A. O. Akyuz, and G. Johnson, Color imaging: fundamentals and applications. AK Peters/CRC Press, 2008.
115. G. Woo, A. Lippman, and R. Raskar, “Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter,” in Proc. IEEE Int. Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2012, pp. 59–64.
116. T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou, “Real-time screen- camera communication behind any scene,” in *Proc. of the 13th Annual Int. Conf. on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 2015, pp. 197–211.
117. V. Nguyen, Y. Tang, A. Ashok, M. Gruteser, K. Dana, W. Hu, E. Wen- growski, and N. Mandayam, “High-rate flicker-free screen-camera com- munication with spatially adaptive embedding,” in Proc. IEEE Int. Conf. on Computer Communications (INFOCOM). IEEE, 2016.
118. M. Izz, Z. Li, H. Liu, Y. Chen, and F. Li, “Uber-in-light: Unobtrusive visible light communication leveraging complementary color channel,” in *Proc. IEEE Int. Conf. on Computer Communications (INFOCOM)*. IEEE, 2016.
119. HAO, T., ZHOU, R., AND XING, G. COBRA: Color barcode streaming for smartphone systems. In Proc. of MobiSys (2012).
120. PERLI, S. D., AHMED, N., AND KATABI, D. PixNet: Interference-free wireless links using LCD-camera pairs. In *Proc. of MobiCom* (2010).
121. HU, W., GU, H., AND PU, Q. LightSync: Unsynchronized visual communication over screen-camera links. In Proc. of MobiCom (2013).
122. HU, W., MAO, J., HUANG, Z., XUE, Y., SHE, J., BIAN, K., AND SHEN, G. Strata: Layered coding for scalable visual communication. In Proc. of MobiCom (2014).
123. RAJAGOPAL, N., LAZIK, P., AND ROWE, A. Visual light landmarks for mobile devices. In *Proc. of IPSN* (2014).
124. WANG, A., MA, S., HU, C., HUAI, J., PENG, C., AND SHEN, G. Enhancing reliability to boost the throughput over screen-camera links. In *Proc. of MobiCom* (2014).
125. CARVALHO, R., CHU, C.-H., AND CHEN, L.-J. IVC: Imperceptible video communication. In Proc. of HotMobile (poster) (2014).
126. WANG, A., PENG, C., ZHANG, O., SHEN, G., AND ZENG, B. InFrame: Multiflexing full-frame visible communication channel for humans and devices. In *Proc. of HotNets* (2014).
127. WOO, G., LIPPMAN, A., AND RASKAR, R. VRCodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In Proc. of ISMAR (2012).
128. YUAN, W., DANA, K., VARGA, M., ASHOK, A., GRUTESER, M., AND MANDAYAM, N. Computer vision methods for visual mimo optical systems. *Proceedings of the IEEE International Workshop on Projector-Camera Systems (held with CVPR)* (2011), 37–43.
129. YUAN, W., ET AL. Dynamic and invisible messaging for visual mimo. In *IEEE Workshop on Applications of Computer Vision (WACV)* (2012).
130. L. M. Marvel & Charles T. Retter, "A methodology for data hiding using images", IEEE 1998. Recurso digital: <http://www.argreenhouse.com/society/TacCom/papers98/30_07i.pdf>.
131. B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, USA, 1986.

# **ANEXOS**