

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

ZIENTZIA  
ETA TEKNOLOGIA  
FAKULTATEA  
FACULTAD  
DE CIENCIA  
Y TECNOLOGÍA



Trabajo fin de grado  
Grado en física

# Aplicación de una red neuronal para la clasificación de materiales a través de imágenes multispectrales

Autor: Oscar Rojo Naveira  
Director: Javier Echanobe

Leioan, 2023eko ekainaren 22a / Leioa, 22 de junio de 2023

# 1. Objeto y desarrollo

Este trabajo se enmarca en el contexto de la industria automovilística y su afán por desarrollar el coche autónomo. Más concretamente, en la clasificación de objetos de la vía pública a través de algoritmos de *Machine Learning* partiendo de imágenes multispectrales. Dichas imágenes se obtienen a través de una cámara multispectral de bajo coste.

El objetivo fundamental de este trabajo es comprobar la eficacia de las cámaras multispectrales para extraer una firma espectral que nos permita distinguir materiales de la vía pública. Una buena pregunta podría ser ¿Que novedad o avance supone esto? Si bien es cierto que el reconocimiento de materiales es un tema muy avanzado hoy en día, incluso sin el uso de imágenes multispectrales, para alcanzar esos resultados las redes neuronales que se emplean cuentan con un gran número de entradas (en general todos los píxeles de la imagen) incrementando el costo computacional. Atendiendo a este hecho y partiendo de la premisa de que se trata de una aplicación en tiempo real, dotar a un sistema de la capacidad de reconocer materiales píxel a píxel con una cantidad relativamente pequeña de variables de entrada (25 bandas del sensor) supondría una reducción importante en el consumo, costo y tamaño del equipo empleado.

Para tratar de entender mejor la meta de este trabajo tenemos que contextualizar enumerando la larga serie de condicionantes que imposibilitan o complican inmensamente el trabajo de extraer firmas espectrales. Los hay de dos tipos, irremediables y evitables. Los irremediables son básicamente las condiciones meteorológicas y de iluminación cambiantes y el hecho de que las clases no son homogéneas (hay distintos materiales dentro de una misma clase etiquetada como un mismo material). Por otro lado, los evitables tienen que ver con las características del sensor y la información que proporciona. Estas últimas son las que abordaremos y trataremos de corregir aplicando técnicas algebraicas y de cálculo computacional así como de *Data mining* que se irán detallando a lo largo de la memoria.

Los objetivos concretos de este trabajo son los siguientes: buscar técnicas de preprocesamiento para la aplicación en concreto, entre las cuales se encuentra la ortogonalización de Gram-Schmidt para la obtención de una matriz de corrección espectral, calcular la separabilidad que se utilizará como medida de eficacia y analizar la clasificación con redes neuronales.

El trabajo se estructura en primer lugar con una pequeña introducción al *Machine Learning*, *Data Mining* y sus aplicaciones dentro del sector automovilístico, concretamente en la conducción autónoma. Por otro lado, también se proporciona una información meramente cualitativa sobre cámaras multispectrales, espectrómetros e imágenes multispectrales. Posteriormente, se detallan los fundamentos teóricos que acompañan este trabajo. A continuación, se describe la base de datos que se va a utilizar, los experimentos que se han llevado a cabo y se analizan detenidamente los resultados obtenidos. Finalmente, concluimos el trabajo y se instan diferentes ramas de investigación que podrían ser interesantes para futuros trabajos.

## 2. Introducción

En esta sección se definen a grandes rasgos los conceptos de espectrómetro, cámaras e imágenes hiperespectrales y multiespectrales, así como, los términos *Machine Learning* y *Data Mining*. Además, se hace un pequeño comentario sobre los vehículos inteligentes con el enfoque de este trabajo.

### 2.1. Vehículos inteligentes y sistemas de visión

Existen multitud de proyectos en este ámbito tanto en el campo científico como en el industrial. Casi todas las empresas de automóviles tienen distintos niveles de automatización en sus diferentes segmentos. El nivel de automatización de la conducción en un automóvil se puede esquematizar a través de una tabla en la que se incluyen 5 fases que van desde la conducción sin asistencia a la conducción totalmente autónoma. Actualmente a pesar de contar con unos niveles totalmente definidos, no existe en el mercado ningún coche autónomo. Sin embargo, hay muchos esfuerzos y proyectos en esa dirección, lo cual refleja los intereses socioeconómicos de multitud de empresas. Es importante mencionar este aspecto porque los intereses y necesidades de este mercado son en los que me apoyo para el desarrollo de este trabajo.

Hoy en día la mayoría de los sistemas de detección de objetos en la vía pública se basan en imágenes RGB. Sin embargo, como el contexto en el que nos encontramos, sistemas de conducción autónoma ADAS (Automated Driver Assistance Systems), demanda una gran fiabilidad, la información extra que proporcionan las cámaras multiespectrales e hiperespectrales puede ser de ayuda para mejorar la robustez de los algoritmos. Es por ello que esta nueva tecnología ha empezado a llamar la atención de los desarrolladores dedicados a ADAS [1].

Según el estado del arte, en este campo se han realizado estudios tanto con NIR (Near Infrared) como con FIR (Far-Infrared). Los realizados con FIR se centran sobre todo en la detección de peatones, ya que la información en su mayoría proviene de la temperatura. No obstante, esta información no resulta de gran utilidad ya que en verano las diferencias de temperatura con el entorno son ínfimas y en invierno los abrigos las enmascaran. Adicionalmente, se han realizado experimentos combinando RGB, NIR, MIR, (Medium Infrared) y FIR utilizando los mismos algoritmos que únicamente con el visible aumentando la distinción de objetos a costa de aumentar notoriamente el costo de procesamiento. Es por ello que recientemente, los investigadores han centrado sus estudios en la firme idea de que la información de las cámaras multiespectrales proporciona firmas espectrales de los distintos materiales. Este hecho hace posible distinguir materiales píxel a píxel utilizando únicamente como variables las bandas del sensor, reduciendo así el costo de procesamiento y mejorando la calidad de la información. Razonamiento en el que se basa este trabajo y que constituye la hipótesis principal del mismo. Por otro lado, un estudio dedicado a la detección de peatones utilizando una cámara hiperespectral de 225 bandas en el rango de  $400nm-2500nm$ , concluyó que la información contemplada en el rango  $400nm-1000nm$  ofrecía resultados similares a la completitud del conjunto de datos proporcionado por el sensor. Este dato es de especial relevancia ya que nuestra cámara y la mayoría de las cámaras multiespectrales del mercado trabajan en esa zona del espectro, visible + VNIR (Very Near Infrared)[1].

## 2.2. Espectrómetros y cámaras e imágenes hiperespectrales y multiespectrales

Se denomina espectrómetro a todo aparato que detecta, mide y analiza el campo de frecuencias para el movimiento ondulatorio del espectro electromagnético de la radiación incidente. Generalmente miden intensidad luminosa utilizando como variable independiente la longitud de onda. Los espectrómetros tienen multitud de aplicaciones. Entre las más típicas se encuentran la determinación de la composición química de cierta muestra o de las propiedades de la luz de una parte del espectro a través del espectrómetro de masas y el espectrómetro óptico respectivamente.

Las cámaras multiespectrales e hiperespectrales están englobadas dentro de la definición de espectrómetro. Para fijar el punto de partida tenemos que señalar que la cámara empleada en el desarrollo de este proyecto es multiespectral. Los motivos principales de esta decisión son meramente de costo, tanto económico como computacional. Con objeto de distinguir cualitativamente las propiedades de nuestro sensor frente a las de uno hiperespectral vamos a comentar los puntos en los que difieren. Es importante aclarar que muchas de las cámaras multiespectrales e hiperespectrales, sobre todo las de bajo coste (como la utilizada en este trabajo) están basadas en un sistema de filtros y no en un espectrómetro puro como tal. Por ese motivo aparecen efectos indeseados que se tratarán de solucionar en este trabajo.

Una de las principales características son el número de bandas proporcionadas por la cámara. Mientras que las cámaras hiperespectrales cuentan con cientos e incluso miles, las multiespectrales solo tienen unas pocas decenas. Además, la resolución de la información espectral, como es de esperar es mucho mejor en las cámaras hiperespectrales debido al uso de bandas más estrechas. A causa de estas características la información de las cámaras multiespectrales se ve reducida y podríamos incluso decir que está peor definida [2].

En contraposición, también hay puntos a favor de las cámaras multiespectrales. Entre ellos, los más destacables se refieren a la menor complejidad, costo de procesamiento de los datos y costo económico. Si bien es cierto que la información proporcionada por las cámaras hiperespectrales es más amplia y a causa de ello permite obtener mejor las firmas espectrales, también es verdad que trabajar con ella es mucho más complicado, sobre todo en cuanto a la eliminación de redundancias. Reincidiendo sobre este punto, otra desventaja del exceso de información es el costo de procesamiento (mayor número de variables de entrada) y por tanto aumento del consumo y costo de procesamiento del equipo empleado. No olvidemos que uno de nuestros objetivos es reducir precisamente ese costo.

En cuanto al formato de las imágenes, éstas adoptan una forma de matriz tridimensional o cubo. Dos de las dimensiones del cubo hacen referencia a la resolución o número de píxeles  $n = n_1 \times n_2$  y la tercera al número de bandas  $n_b$ . Es decir, cada píxel está definido en el espacio espectral  $x \in \mathbb{R}^{n_b}$  y la representación espacial de cada imagen por banda es una matriz  $X_i \in \mathbb{R}^{n_1 \times n_2}$  [3]. Esto se muestra esquemáticamente en la figura 1.

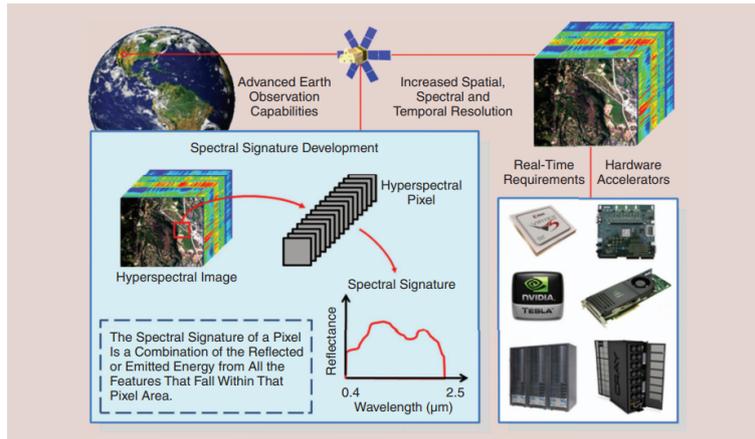


Figura 1: Formato de una imagen hiperespectral y ciertos elementos de hardware utilizados para acelerar el procesamiento de los datos [3].

### 2.3. *Machine learning y Data mining*

Con el término *Machine Learning* se hace referencia a una rama de la inteligencia artificial cuya intención es dotar a las computadoras de la capacidad de aprender a realizar una determinada tarea, llevar a cabo predicciones y/o tomar decisiones de manera automática estableciendo patrones extraídos de los conjuntos de datos que se les proporcionan [4][5]. Los algoritmos de *Machine Learning* se pueden agrupar en dos clases. Por un lado, están los algoritmos supervisados que se caracterizan por la necesidad de contar tanto con los valores de entrada como de salida para el entrenamiento y por otro, los no supervisados, a los que no se les suministran los valores de salida. El campo de aplicación del *Machine Learning* es muy amplio, se utiliza en ámbitos tan diversos como la medicina, los videojuegos, la robótica, por nombrar algunos de ellos, y por supuesto el sector automovilístico que es el que nos concierne. Algunos ejemplos típicos de algoritmos son las máquinas de soporte de vectores SVM, los árboles de decisión y las redes neuronales.

Por otra parte, lo que actualmente se conoce por *Data Mining* (Minería de datos) es un término que se enmarca en el campo de la estadística y la computación y que esta directamente relacionado con el *Machine Learning*. La mayoría de esfuerzos en esta disciplina se centran en convertir datos en activos de valor. El *Data Mining* es un proceso técnico, automático o semiautomático, que procesa y analiza grandes bases de datos para darles significado [6]. La minería de datos se estructura en pasos sucesivos que van desde realizar un análisis, limpieza y transformación de datos hasta implementar un modelo de predicción o clasificación y evaluar sus resultados [7][8][9].

### 2.4. *Machine Learning en el ámbito del automóvil*

Concretamente el *Machine Learning* y el *Data Mining* se utilizan ampliamente en el sector automovilístico. Constituyen uno de los pilares de la conducción autónoma, junto al uso de sensores y actuadores que se encargan de la recopilación de la información.

Uno de los ejemplos de aplicación de estas técnicas en el ámbito del automóvil proviene de uno de los gigantes del sector. La empresa Tesla cuenta con redes neuronales capaces de categorizar píxeles, detectar obstáculos y estimar profundidades o distancias respecto a cada cámara. Su Dataset se construye a partir de la información que proviene de su

flota de vehículos. Presumen de contar con una red de autopilotaje que involucra 48 redes neuronales, las cuales han necesitado 70.000 GPUs y multitud de horas para ser entrenadas. Evidentemente, no realiza una tarea tan simple como la que nosotros queremos llevar a cabo, sino que en cada paso efectúa 1000 predicciones [10].

Otros ejemplos de algoritmos de Machine Learning en el ámbito del automóvil inteligente los podemos encontrar en los siguientes sistemas: En [11] se utilizan máquinas de soporte de vectores (SVM) para detectar estados de somnolencia en el conductor a través de señales de la actividad cerebral y de ritmos cardíacos. En [12], los autores proponen redes neuronales profundas (*deep learning*) para detectar las líneas de carril y avisar al conductor en caso de traspasarlas. Un algoritmo de *clustering* se emplea en [13] para analizar cómo afecta la manera de conducir al consumo del vehículo. Un sistema para detectar peatones y obstáculos basado en una red neuronal recurrente (RNN) se propone en [14]. Como último ejemplo, un sistema basado en Random Forest se utiliza en [15] para evaluar los posibles riesgos de colisión de vehículos próximos en la vía.

### 3. Fundamentos teóricos

#### 3.1. Redes neuronales

Las redes neuronales son una técnica popular contemplada dentro del *Machine Learning*. Esta idea surgió por analogía a las redes neuronales humanas formadas por neuronas interconectadas mediante axones y dendritas en las que las regiones de interconexión se denominan sinapsis [4]. Las redes neuronales artificiales se basan en este modo de operación y tratan de imitar el método de aprendizaje del cerebro humano. Esta semejanza se ilustra en la figura 2.

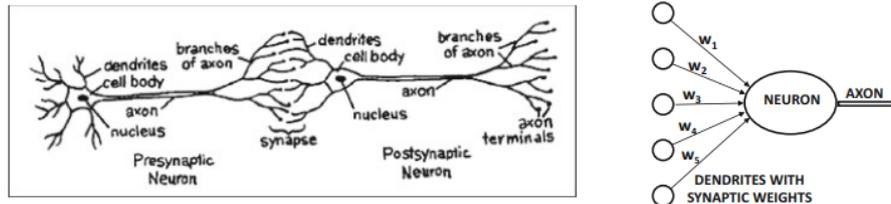


Figura 2: Imagen ilustrativa de la similitud entre una neurona biológica y una neurona artificial [4].

Las redes neuronales están formadas por neuronas artificiales interconectadas entre sí y agrupadas en capas, cuyas variables de entrada  $x_i$  son las salidas de las neuronas de las capas adyacentes ponderadas por unos valores que llamamos pesos  $w_i$ . Los pesos en lenguaje coloquial se podría decir que representan las fuerzas de conexión entre neuronas de distintas capas. Además, cuentan con una entrada adicional denominada *bias*, a menudo referida como sesgo, no ponderado  $b$ . Esta entrada adicional da cuenta de la predisposición de la neurona a ser activada o no. De forma matemática podemos expresar la salida  $y$  de una neurona de la siguiente forma:

$$y = f\left(\sum_{i=1}^N w_i x_i + b\right) \tag{1}$$

donde  $N$  es el número de neuronas conectadas a la neurona que estamos considerando y  $f$  es una función fuertemente no lineal denominada función de activación. Hay varias funciones de activación entre las cuales predomina la función sigmoide. En la figura 3 se muestran las más típicas [4].

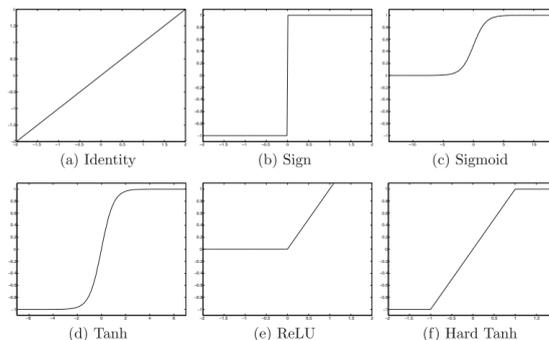


Figura 3: Funciones de activación más típicas [4].

Para esquematizar la propagación de la información y el comportamiento de la neurona la figura 4 resulta muy representativa, donde  $y$  es el valor de salida de la neurona,  $x$  las entradas,  $w$  los distintos pesos y  $b$  el *bias*.

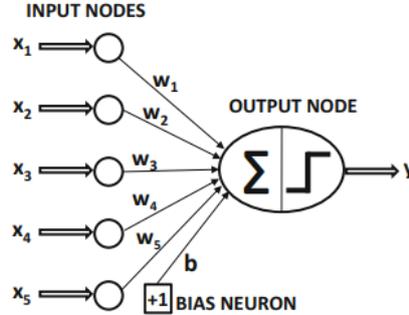


Figura 4: Esquema de funcionamiento de una neurona artificial [4].

Hasta ahora hemos descrito el funcionamiento de la unidad de procesamiento en este tipo de arquitecturas, sin embargo, para entender totalmente en qué consiste una red neuronal es necesario aumentar el nivel de abstracción y así visualizar la disposición de éstas en la estructura global. Estas neuronas como comentábamos se organizan por capas. A la primera capa se le denomina capa de entrada, a las intermedias capas ocultas y a la de salida capa de salida. Las redes neuronales de varias capas ocultas que se alimentan de atrás adelante son comúnmente conocidas como redes *feedforward* [4]. Esto se muestra en detalle en la figura 5 donde los círculos representan neuronas. En este esquema global de la red neuronal *feedforward* la salida de una neurona  $j$  perteneciente a la capa  $k$  viene dada por:

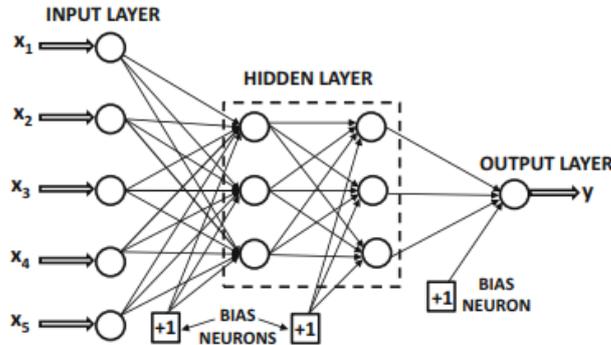


Figura 5: Arquitectura de una red neuronal con dos capas ocultas [4].

$$y_j^k = f(b_j^k + \sum_{i=1}^N w_{ij}^{k-1} y_i^{k-1}) \quad (2)$$

Las redes neuronales aprenden en base a la prueba y error. Para realizar esta tarea se actualizan los pesos en sucesivas iteraciones según la regla del descenso del gradiente que pretende minimizar el error entre la salida deseada y la obtenida en cada iteración. Este error vendrá dado por una función denominada función error. Para minimizarla, la variable respecto de la cual se deriva cambia su valor en contra de la pendiente de

la función de tal manera que si la derivada es positiva la variable adquiere valores más pequeños y si es negativa al contrario [4].

$$w_{i_{n+1}} = w_{i_n} - \eta \left. \frac{\partial f_{error}}{\partial w_i} \right|_{w_i=w_{i_n}} \quad (3)$$

Normalmente se utiliza el error cuadrático medio como medida del error cometido. La función error depende de todos los pesos y bias de la red neuronal de ahí el origen de la derivada parcial. Evidentemente asumiendo este método de aprendizaje damos por hecho que hablamos sobre aprendizaje supervisado ya que necesitamos conocer la salida esperada. El parámetro  $\eta$  se refiere a la tasa de aprendizaje, es decir, el tamaño de los pasos en la actualización de los pesos. Es importante escoger adecuadamente este parámetro para no entrar en bucles sin salida por un valor excesivamente grande o una convergencia muy lenta debido a valores muy pequeños.

Existe un algoritmo muy eficaz que se encarga de la actualización de los pesos mediante el cálculo de las derivadas parciales denominado *Error backpropagation*. Como su propio nombre indica se trata de una propagación de errores hacia atrás en la red neuronal. Para que este algoritmo sea aplicable la función de activación ha de ser derivable [5]. Este algoritmo se beneficia de las propiedades derivativas de las funciones de activación como puede ser la sigmoide cuya derivada esta caracterizada por:

$$f'(x) = f(x)(1 - f(x)) \quad (4)$$

En esta sección no se describe todo el desarrollo de las ecuaciones de actualización de los pesos debido a que el proceso es bastante extenso y carece de relevancia para este trabajo.

Cabe considerar por otra parte, los problemas derivados de un mal entrenamiento de una red neuronal. Uno podría pensar que a mayor entrenamiento mejor será el modelo ya que se ajusta mejor a los datos de entrenamiento. Sin embargo, la realidad no es así, tenemos que tener en cuenta el problema del *Overfitting*. Este hace referencia a que un modelo que se ajuste a la perfección a unos datos de entrenamiento no garantiza que realice buenas predicciones para datos desconocidos o de test. En otras palabras, siempre hay un desacuerdo (mayor o menor) entre los resultados arrojados por la red para los datos de test y entrenamiento [4].

Para evaluar el comportamiento de la red entrenada vamos a utilizar dos medidas de eficiencia.

- **Matriz de confusión:** Su propio nombre nos da pistas de a que se refiere. Este tipo de matrices da una medida de cuán efectivas están siendo las predicciones. En un problema de clasificación de  $s$  clases, la matriz de confusión adoptará un tamaño de  $s \times s$ . Los elementos de la diagonal representan los aciertos (en número absoluto o en términos porcentuales), es decir, los elementos etiquetados en cierta clase que se han identificado como tales. El resto de elementos corresponden a predicciones erróneas. Un ejemplo ilustrativo de una matriz de confusión se muestra en la figura 6.
- **Tasa de acierto:** La tasa de acierto es otro indicador de la validez del algoritmo. Es el ratio entre los aciertos obtenidos y el número de muestras procesadas.

Output Class	1	2	3	4	
1	6839 20.2%	58 0.2%	73 0.2%	265 0.8%	94.5% 5.5%
2	6 0.0%	4301 12.7%	76 0.2%	580 1.7%	86.7% 13.3%
3	8 0.0%	229 0.7%	12494 36.9%	164 0.5%	96.9% 3.1%
4	10 0.0%	434 1.3%	49 0.1%	8279 24.4%	94.4% 5.6%
	99.7% 0.3%	85.6% 14.4%	98.4% 1.6%	89.1% 10.9%	94.2% 5.8%
	1	2	3	4	
	Target Class				

Figura 6: Ejemplo de matriz de confusión para un problema de 4 clases.

### 3.2. Reflectancia

Los datos que vamos a procesar para la identificación de objetos en la vía pública corresponden a valores de reflectancia obtenidos con nuestra cámara multispectral. Analicemos en qué consiste esta magnitud física. La reflectancia se define como el ratio entre la radiancia  $\rho$  que refleja un material y la que incide sobre él. La radiancia es una magnitud física cuyas unidades son  $Wsr^{-1}m^{-2}$ . Se entiende por radiancia la energía emitida por una superficie por unidad de tiempo, unidad de ángulo sólido y unidad de área.

El sensor con el que cuenta la cámara inicialmente nos proporciona una imagen *raw* (en bruto) que nos da una medida de las radiancias de los distintos materiales que se encuentran en su campo de visión. Con el fin de deshacernos de dependencias en la respuesta del sensor con sus características la empresa que nos facilita el sensor, *Imec*, invita a modificar la información de la imagen, traduciendo la radiancia en reflectancia. Eliminando en teoría la influencia de la eficiencia cuántica así como la de transmisión y de la iluminación [16]. Sin embargo, a lo largo del trabajo veremos cómo esto no es del todo cierto, más bien se podría considerar una idealización.

El punto de partida para realizar la traducción anterior es obtener la respuesta maximal de una superficie blanca uniforme desde el punto de vista espacial y espectral para todos los píxeles de una imagen. Esta imagen será nuestra referencia para escalar los valores de radiancia de las imágenes *raw* [16]. El escalado viene dado por:

$$v = \frac{\rho}{\rho_{ref}} \quad (5)$$

donde  $\rho_{ref}$  es la radiancia de la superficie blanca y  $v$  es la reflectancia cuyos valores estarán dentro del rango  $[0,1]$ .

En multitud de ocasiones la referencia tiene valores de radiancia muy elevados respecto a la imagen *raw*. En respuesta a este hecho, solo se utiliza una fracción del rango

dinámico del sensor. Además para maximizarlo se asume que la radiancia de cada pixel varía linealmente con el tiempo de integración  $\tau$  del sensor [16].

$$v = \frac{\rho\tau_{ref}}{\rho_{ref}\tau} \quad (6)$$

Finalmente, se elimina el ruido estático con un par de imágenes negras con tiempos de integración  $\tau$  y  $\tau_{ref}$  obtenidas tapando el objetivo de la cámara [16].

$$v = \frac{\rho - \rho_{0,\tau}}{\rho_{ref} - \rho_{0,\tau_{ref}}} \frac{\tau_{ref}}{\tau} \quad (7)$$

Es importante no olvidar que hemos asumido que la radiancia varía linealmente con el tiempo de integración. Para verificar dicha linealidad se toman distintas imágenes de un mismo entorno con distintos tiempos de integración. Con el fin de alcanzar la linealidad en toda imagen es necesario obtener las imágenes en negro correspondientes [16].

### 3.3. Sensor multiespectral

El sensor de la cámara utilizada es un sensor CMV2K-SM5x5-NIR de tecnología CMOS cuyas características explicamos a continuación. Previamente, repasamos el concepto del espectro electromagnético.

#### 3.3.1. Espectro electromagnético

El espectro electromagnético es el conjunto que abarca todos los tipos de radiación electromagnética clasificados en función de la longitud de onda que comprenden. Las ondas emitidas tanto por un microondas como las emitidas por el sol en forma de luz ultravioleta son dos tipos de radiación electromagnética que pertenecen a dicho espectro.

En este trabajo nos interesa situar el rango del espectro visible y el infrarrojo cercano ya que es la zona de trabajo de nuestro sensor ( $400nm-1000nm$ ). Debido a nuestra naturaleza, los humanos, solo somos capaces de visualizar una parte ínfima del espectro ( $400nm-700nm$ ) considerada el espectro visible. Por otro lado, está el infrarrojo que abarca las longitudes de onda comprendidas entre los  $700nm$  hasta  $1mm$ . El espectro electromagnético se muestra en la figura 7.

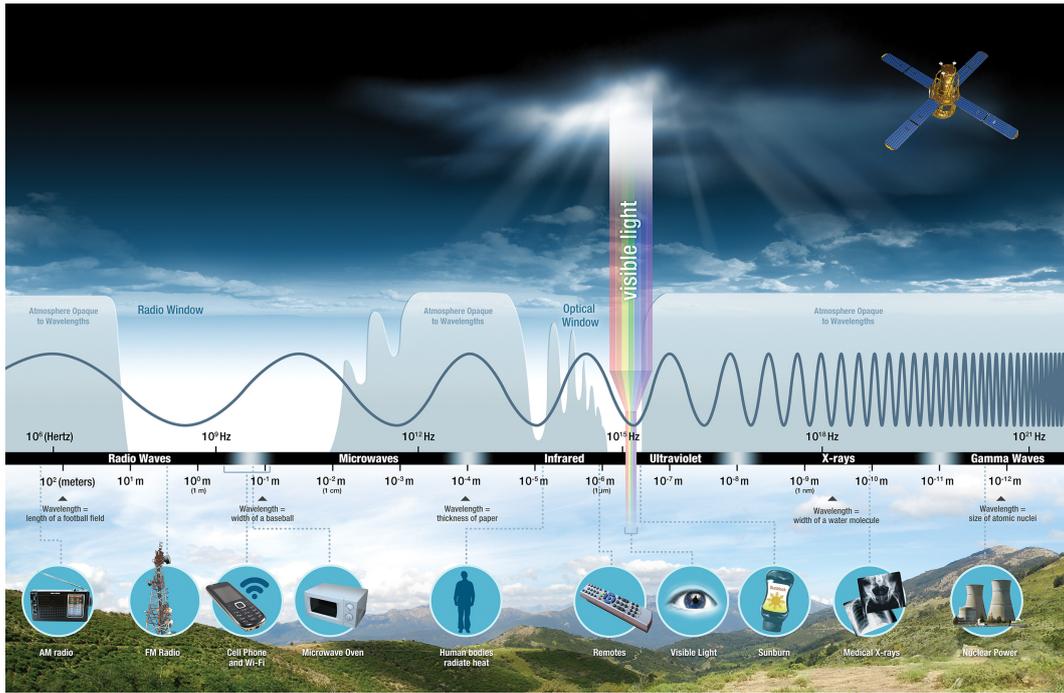


Figura 7: Espectro electromagnético completo [17].

### 3.3.2. Descripción y ordenamiento de los filtros

En general, las cámaras comerciales típicas generan las imágenes utilizando las radiación referente a longitudes de onda del visible seleccionándolas a través de filtros RGB que permiten el paso de las frecuencias correspondientes a los colores rojo, verde y azul. Sin embargo, los sensores hiperespectrales capturan imágenes que incluyen información más allá del rango visible del espectro electromagnético.

En nuestro caso, contamos con el sensor antes citado CMV2K-SM5x5-NIR de tecnología CMOS con una disposición de filtros *snapshot mosaic* como se muestra en la figura 8. Cuenta con 25 bandas y la zona activa de éstas según las especificaciones del fabricante es una pequeña parte del visible junto al infrarrojo cercano ( $600nm-975nm$ ). Sin embargo, la zona de trabajo real si no colocamos ninguno de los filtros pasa banda, que limitan su rango activo, es más amplia ( $400nm-1000nm$ ) [16].

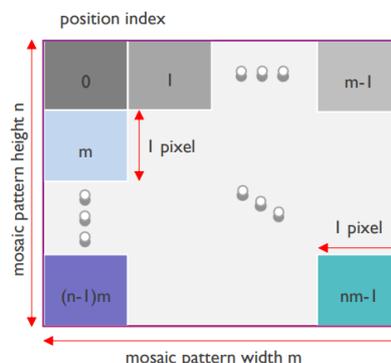


Figura 8: Organización de los filtros de Fabry-Perot en el patrón mosaico.

Las 25 bandas antes comentadas se corresponden con filtros de Fabry-Pérot y se distribuyen en un patrón mosaico de  $n \times m$  pixels (25 en el caso de nuestro sensor,  $n = 5$  y  $m = 5$ ) como se muestra en la figura 8. Este patrón se repite tantas veces como la resolución de la imagen permite (2048x1088, útiles 2045x1080 en nuestro caso) siguiendo la estructura de la figura 9. Dada la resolución efectiva de nuestro sensor es fácil deducir que los parámetros  $w$  (anchura) y  $h$  (altura) de la imagen toman los valores  $w = 2045/5 = 409$  y  $h = 1080/5 = 216$ .

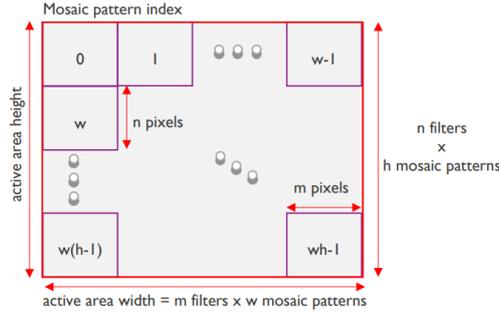


Figura 9: Organización de los patrones mosaico a lo largo de toda la imagen.

Las 25 bandas son las respuestas de los filtros de Fabry-Perot como el que de forma esquemática se muestra en la figura 10.

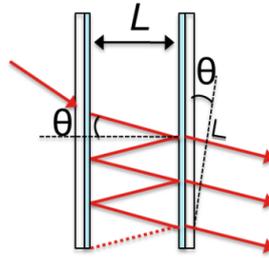


Figura 10: Filtro de Fabry-Perot [16].

Los filtros de Fabry-Perot satisfacen la ecuación:

$$k\lambda = 2nL \cos \theta \quad (8)$$

donde  $n$  es el índice de refracción del material en el interior de la cavidad,  $\lambda$  es la longitud de onda central del filtro,  $k$  el orden del armónico,  $L$  la anchura de la cavidad y  $\theta$  el ángulo respecto de la dirección normal de los espejos.

### 3.3.3. Respuesta de los filtros (bandas)

Evidentemente la función de transferencia de los filtros no es una delta de Dirac como sería en el caso ideal, es decir; un filtro óptimo transmitiría la radiación en una única determinada longitud de onda. Por el contrario, estos filtros tienen un pico pronunciado en una longitud de onda con una anchura que está caracterizada por el parámetro *full width at half maximum* (FWHM). Además, se dan efectos de segundo orden denominados *Crosstalk*, *Spectral shift* y *Spectral leaking* [16].

- **Crosstalk:** Se da cuando los electrones de un píxel que se corresponde con un filtro afectan a la señal de otro píxel referente a otro filtro dando lugar a respuestas no deseadas fuera de los picos [16].
- **Spectral leaking:** Caracteriza la respuesta de los filtros para radiación de longitudes de onda externas al rango activo. Este fenómeno se produce por la elección de los materiales en los filtros [16].
- **Spectral shift:** Los filtros de Fabry-Perot en este sensor son espejos DBR (*Distributed Bragg Reflector*), basados en las reflexiones de Fresnel en incidencia normal entre interfases de materiales con alto y bajo índice de refracción  $n_h$  y  $n_b$  respectivamente [18]. Estas reflexiones se ilustran en la figura 11. En este tipo de espejos no podemos asumir un cambio de fase de  $\lambda/2$  en cada reflexión. Como consecuencia de ello, la longitud de onda del armónico de segundo orden está un poco desplazada [16].

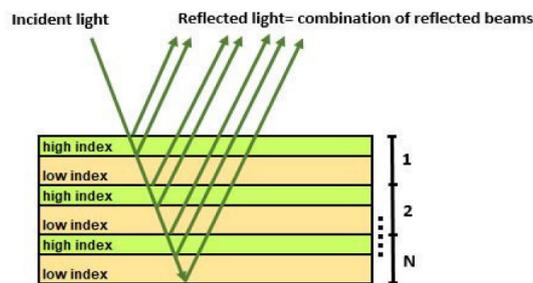


Figura 11: Reflectancia de un espejo tipo DBR [18].

Estos efectos se observan claramente en la figura 12.

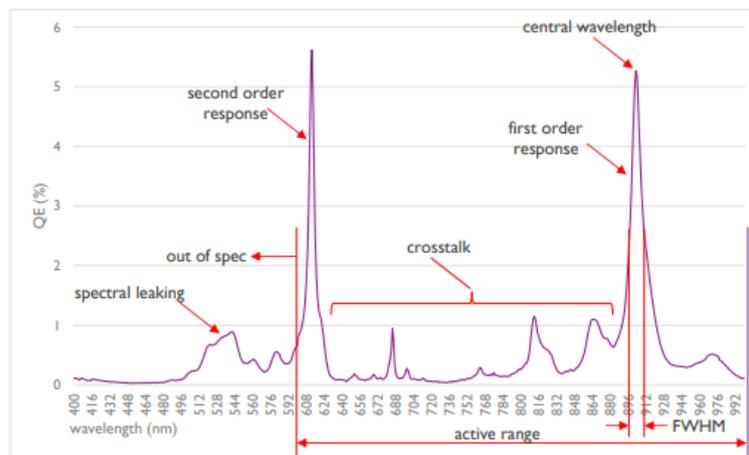


Figura 12: Efectos de segundo orden en los filtros de Fabry-Perot [16].

### 3.3.4. Modo de uso y recomendaciones del fabricante

El fabricante *Imec* nos ofrece cierta información sobre la utilización correcta de su sensor. En ella se incluyen aspectos como la iluminación, las lentes, filtros externos, corrección de reflectancia y corrección espectral en la cual se centra este trabajo.

En cuanto a la iluminación *Imec* recomienda el uso de luz halógena de ajuste frío con revestimiento de aluminio u oro, combinada con una luz LED (Light Emitting Diode) de color blanco o rojo [16]. Es necesario mencionar este aspecto porque es fácil deducir que se trata de una iluminación muy concreta que no encontraremos en la vía pública siendo así este uno de los puntos negativos que escapan de nuestro control.

Por otro lado, a pesar de ser un punto que no tratamos, la empresa también mencionar que la mejor respuesta del sensor se consigue cuando el CRA (*Chief Ray Angle*) es ortogonal al sensor. El ángulo del cono de luz entorno al CRA se controla por medio de la apertura expresada a través del número F. A menor apertura, el ángulo del cono de luz se ve más limitado y la respuesta es mejor. En contraposición, una menor apertura también es sinónimo de una menor cantidad de luz, por lo que serán necesarios mayores tiempos de integración. Sin embargo, como bien sabemos el sensor se localizará sobre un vehículo tratando de distinguir objetos de la vía en tiempo real por lo que los tiempos de integración son limitados [16].

Adicionalmente, se recomienda encarecidamente el uso de los filtros pasa banda, mencionados con anterioridad. Estos filtros se utilizan para eliminar la respuesta del sensor fuera del rango activo o *Spectral leaking* y seleccionar los picos de primer o segundo orden [16]. Un ejemplo de la respuesta con y sin filtro en condiciones controladas se ilustra en la figura 13.

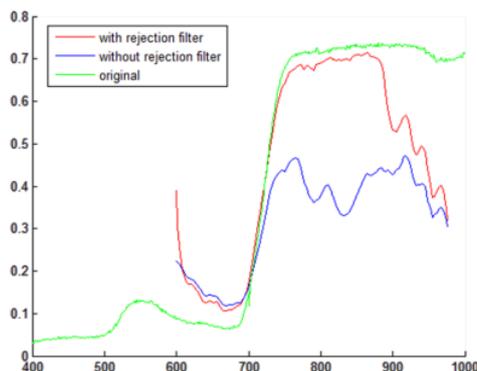


Figura 13: Respuesta del sensor con y sin filtro frente a un espectro original [16].

Como se aprecia en la figura 13 la respuesta es mejor. No obstante, no se eliminan todos los efectos de segundo orden indeseados. El *Crosstalk* sigue presente en la respuesta de los filtros. Ese efecto se trata posteriormente en el punto de corrección espectral. En el caso de nuestra cámara se cuenta con dos filtros paso banda A ( $600nm-875nm$ ) y B ( $675nm-975nm$ ). La respuesta de los filtros viene dada en la figura 14.

En cuanto al preprocesamiento de las imágenes una vez obtenidas es importante tener en cuenta el punto 3.2, en el cual explicábamos el procedimiento para convertir los datos de radiación en reflectancia. Por medio de este paso eliminamos supuestamente las dependencias con la eficiencia cuántica y de transmisión del sensor, así como con la iluminación.

Finalmente, para resolver la cuestión del *Crosstalk Imec* propone sendas matrices de corrección espectral para los casos en los que previamente se haya aplicado uno de los dos filtros A o B. Para cada uno de los casos existe una matriz de corrección [16]. Se trata de

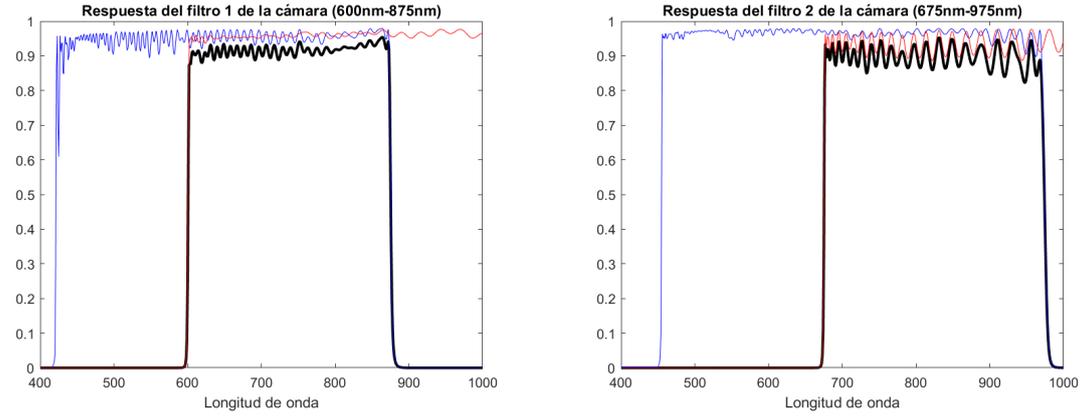


Figura 14: Respuesta de los filtros paso banda A izquierda y B derecha.

matrices de cambio de base que pasan de la base original con 25 bandas a una base de bandas virtuales con 24 componentes en el caso del filtro A y 25 en el B. La empresa es muy optimista y muestra gráficamente en la figura 15 la respuesta del sensor aplicando un filtro y la corrección espectral junto con el espectro real para el caso de la vegetación, además de la respuesta únicamente con el filtro.

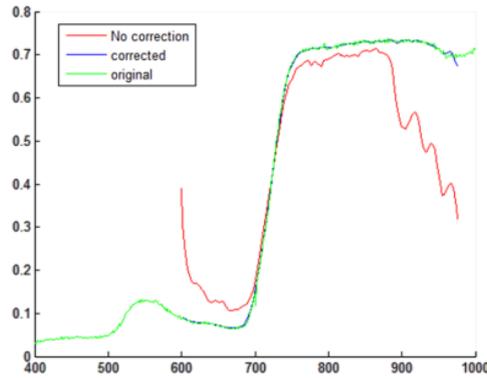


Figura 15: Hipotética respuesta del sensor en condiciones ideales con filtro y corrección espectral frente a un espectro real.

La matriz de corrección espectral representa una combinación lineal de las bandas originales en función de las nuevas o virtuales. La corrección espectral ha de ser aplicada posteriormente a haber realizado la corrección de reflectancia [16]. La fórmula que refleja todo el preprocesamiento de la imagen es:

$$v = C \frac{\rho - \rho_{0,\tau} \tau_{ref}}{\rho_{ref} - \rho_{0,\tau_{ref}} \tau} \quad (9)$$

donde  $C$  es la matriz de corrección espectral que en un caso sin redundancias sería de 25x25 en nuestro sensor. Tal y como está definida la matriz de  $Imec$  el producto se realiza como matriz por vector columna (cada vector columna contiene las 25 componentes originales de un píxel).

Posteriormente veremos que la realidad no es tan idílica, y que todos los problemas mencionados sumados a alguno externo son muy complicados de solucionar. A pesar de aplicar el filtro, la corrección de reflectancia, la corrección espectral y deducir cual es el

número  $F$  óptimo, no se consiguen resultados del todo satisfactorios. Los motivos pueden ser muy diversos y vamos a destacar los más relevantes.

Para empezar la iluminación evidentemente no la podemos controlar aunque podemos tratar de aminorarlo utilizando técnicas de preprocesamiento como puede ser un tipo concreto de normalización. Por otro lado, se comprobó experimentalmente que el sensor recibe poca radiación y el uso de un filtro externo A o B reducía aun más su transmitancia, así, en días o condiciones de poca iluminación la respuesta era ínfima. A causa de este hecho se decidió utilizar el sensor sin filtro lo que implícitamente implicaba renunciar a su vez a la corrección espectral. Motivo por el cual uno de los puntos importantes de este trabajo es encontrar una corrección espectral personalizada para el caso sin filtro. Otro punto adverso es el hecho de que las clases que se trata de clasificar denominadas *grosso modo* materiales no lo son del todo exactamente: ni son homogéneos, ni tienen la misma composición química como puede ser el tipo de metal utilizado en todas las señales de tráfico de la vía. Como podemos observar, existen infinidad de inconvenientes que dificultan notoriamente la tarea de extraer firmas o información distintiva de los materiales, mejor dicho clases.

### 3.4. Corrección espectral: Ortogonalización de Gram-Schmidt

Anteriormente comentábamos que *Imec* nos ofrece dos matrices de corrección para los casos en los que aplicamos un filtro paso banda. Sin embargo, no lo hace para el caso sin filtro. Al estar nuestro interés centrado en este último caso, a simple vista parecía una buena idea obtener una matriz de corrección espectral personalizada. Para ello nos basamos en el algoritmo de Gram-Schmidt que se explica a continuación.

En 1907 Erhard Schmidt publicó un paper sobre un algoritmo de ortogonalización que al paso de los años se convirtió en el popular método de Gram-Schmidt. Schmidt afirmaba en el paper que su método era una modificación del previamente publicado por J. P. Gram que careció de popularidad mientras que éste se extendió rápidamente entre la comunidad matemática [19].

El algoritmo de Gram-Schmidt nos permite obtener una secuencia ortogonal  $\{\mathbf{q}_n\}$  partiendo de una base linealmente independiente  $\{\mathbf{x}_n\}$  de un espacio de producto interno finito o infinito definiendo los elementos  $\{\mathbf{q}_n\}$  como [19]:

$$\mathbf{q}_1 = \mathbf{x}_1, \quad \mathbf{q}_n = \mathbf{x}_n - \sum_{k=1}^{n-1} \frac{\langle \mathbf{q}_k, \mathbf{x}_n \rangle}{\|\mathbf{q}_k\|^2} \mathbf{q}_k, \quad n \geq 2 \quad (10)$$

Si además queremos obtener una secuencia ortonormal simplemente reemplazamos  $\mathbf{q}_n$  por  $\frac{\mathbf{q}_n}{\|\mathbf{q}_k\|}$ .

Partiendo de esta teoría podemos deducir que la base original va a quedar expresada en función de la nueva a través de una matriz  $n \times n$  de coeficientes  $p_{nk}$ . Despejando de la ecuación 10:

$$\mathbf{x}_n = \mathbf{q}_n + \sum_{k=1}^n \frac{\langle \mathbf{q}_k, \mathbf{x}_n \rangle}{\|\mathbf{q}_k\|^2} \mathbf{q}_k, \quad \mathbf{x}_n = p_{nn} \mathbf{q}_n + \sum_{k=1}^{n-1} p_{nk} \mathbf{q}_k \quad (11)$$

donde se deduce inmediatamente que los coeficientes de la diagonal son la unidad y los  $p_{nk}$  externos a la diagonal son:

$$p_{nk} = \frac{\langle \mathbf{q}_k, \mathbf{x}_n \rangle}{\|\mathbf{q}_k\|^2} \quad (12)$$

Por otro lado, se infiere rápidamente que la forma de la matriz es triangular tal y como se muestra a continuación siendo  $n$  la dimensionalidad del espacio de partida.

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ p_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & 1 \end{pmatrix} \quad (13)$$

Esta es la matriz de cambio de base que expresa las coordenadas de un punto que inicialmente está en la base original en la base ortogonal.

En nuestro caso el espacio algebraico es un espacio de producto interno infinito y este se calcula a través de la expresión:

$$\langle v_i, v_j \rangle = \int_a^b v_i(x)v_j(x)dx \quad (14)$$

Las miembros  $v_i$  y  $v_j$  del espacio definido por la base  $\{v_n\}$  en el caso de nuestra aplicación son las 25 bandas del sensor y su respuesta debería estar caracterizada como un continuo en el rango activo del sensor dando lugar a:

$$\langle v_i, v_j \rangle = \int_\lambda v_i(\lambda)v_j(\lambda)d\lambda \quad (15)$$

Sin embargo, la empresa proporciona la respuesta de cada banda en el rango  $400nm-1000nm$  con saltos de  $1nm$  lo cual nos obliga a utilizar un método de interpolación lineal para calcular esas integrales. En este trabajo se ha optado por utilizar la regla del trapecio, ya que está implementada en *Matlab* y consiste en aproximar la integral por:

$$\int_\lambda v_i(\lambda)v_j(\lambda)d\lambda = \int_a^b f(\lambda)d\lambda \approx \frac{b-a}{2N} \sum_{i=1}^N (f(\lambda_i) + f(\lambda_{i+1})) \quad (16)$$

Siendo  $a$  y  $b$   $400nm$  y  $1000nm$  respectivamente y  $N$  el número de intervalos de integración que en nuestro caso son 600. La aproximación que supone el aplicar esta regla se ve claramente en la figura 16. La función de *Matlab* que se encarga de realizar esta integración es *trapz*.

La expresión general para los coeficientes que no pertenecen a la diagonal de la matriz y no nulos es:

$$p_{nk} = \frac{\langle \mathbf{q}_k, \mathbf{x}_n \rangle}{\|\mathbf{q}_k\|^2} = \frac{\int_\lambda q_n(\lambda)x_k(\lambda)d\lambda}{\int_\lambda q_n(\lambda)q_n(\lambda)d\lambda} \approx \frac{\frac{b-a}{2N} \sum_{i=1}^N (f(\lambda_i) + f(\lambda_{i+1}))}{\frac{b-a}{2N} \sum_{i=1}^N (g(\lambda_i) + g(\lambda_{i+1}))} = \frac{trapz(f)}{trapz(g)} \quad (17)$$

Es imprescindible mencionar uno de los motivos principales por los que resulta tan interesante este método. La clasificación de materiales es una aplicación en tiempo real y

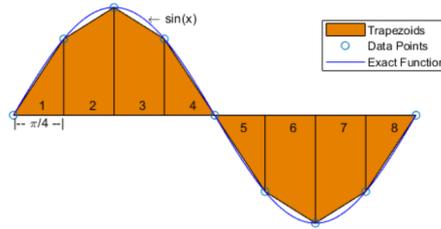


Figura 16: Integración numérica, regla del trapecio aplicada a la función  $\sin(x)$  [20] .

el equipo *Digital Electronics Design Group* (GDED) tiene ya implementada la parte de clasificación de imágenes en *Hardware*. No obstante, la parte de preprocesamiento todavía esta en *Software* y es la que más encarece el cómputo de las operaciones. A raíz de ello una vez alcanzada una solución óptima sería interesante que está fuese fácilmente implementable en *Hardware* para acelerar las operaciones y poder aumentar las *fps* (número de imágenes procesadas por segundo) al máximo. En este aspecto, según [21] la ortogonalización de Gram-Schmidt es un método muy interesante, ya que no requiere del cálculo de matrices inversas. Por otro lado, en [21] exhiben la mejora en la distinción de materiales aplicando el método de Gram-Schmidt a cierto algoritmo de clasificación no supervisado. Evidentemente, no podemos dar por hecho que en nuestro caso vaya a dar los mismos resultados ya que no se trata ni siquiera de la misma técnica de *Machine Learning*.

### 3.5. Separabilidad espectral de los materiales

Cuando mencionamos separabilidad espectral de materiales, hacemos referencia a medidas que nos permitan determinar la distancia o la capacidad que tendría cierto algoritmo predictivo de clasificar bien dos conjuntos de datos que correspondan a dos clases diferentes en un espacio de variables concreto. Adicionalmente, tienen un valor añadido, permiten eliminar redundancias reduciendo el espacio de variables aunque éste no sea el fin con el que se emplean en este trabajo [22].

La separabilidad espectral permite medir la distancia entre clases o el grado de solapamiento entre dos conjuntos de datos de distintas clases. Estas distancias se cuantifican con métricas ya definidas. Existen medidas más rigurosas que otras. Las menos rigurosas son aquellas que solo tienen en cuenta la media. Sin embargo, las hay más estrictas, en las que además se añaden otros parámetros, como podría ser la anchura o varianza de una distribución de datos [22]. El grado de solapamiento que se quiere medir se puede observar en el ejemplo de la figura 17.

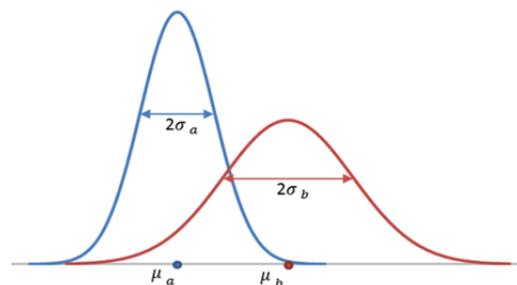


Figura 17: Gráfica de dos distribuciones normales en el espacio de una misma variable [22].

En general, las métricas que se utilizan atienden a todo el espacio de variables en su conjunto y determinan la separabilidad sobre ese espacio. Es importante mencionarlo ya que la figura 17 podría dar lugar a malinterpretaciones, y podríamos deducir erróneamente que estas distancias se calculan variable a variable.

Para el cómputo de estas distancias por lo general se precisan dos medidas de cada conjunto de datos. Por un lado, el vector de medias  $\vec{\mu} = (\mu_1, \dots, \mu_n)$  donde  $\mu_i$  es la media de los datos de una clase en la variable  $i$ . Por otro lado, se requieren las matrices de covarianza  $\Sigma$  cuyos elementos  $\sigma_{ij}$  se calculan como sigue [22]:

$$\sigma_{ij} = (X_i - \mu_i)(X_j - \mu_j) \quad \text{con} \quad i, j = 1, \dots, n \quad (18)$$

donde  $n$  es el número de variables. Si  $i = j$  la medida  $\sigma_{ii}$  es la varianza de la variable  $i$  respecto a su media y cuando  $i \neq j$  la medida  $\sigma_{ij}$  hace referencia a la covarianza de la variable  $i$  respecto a la  $j$ .

A continuación, se mencionan varias métricas. Sin embargo, en la práctica la medida que se ha utilizado ha sido la de *Jeffries-Matusita* considerada muy precisa y también la más usada.

### 3.5.1. Distancia Euclidea

Para calcular esta distancia es necesario únicamente utilizar los vectores de medias  $\vec{\mu}$  de las dos clases cuya separabilidad se pretende calcular [22].

$$ED = \|\mu_a - \mu_b\| = [(\mu_a - \mu_b)^T(\mu_a - \mu_b)]^{1/2} \quad (19)$$

### 3.5.2. Distancia divergencia

El punto de partida para esta medida de disparidad es la razón de verosimilitud.

$$L_{ab}(x) = p(x|a)/p(x|b) \quad (20)$$

siendo  $p(x|a)$  es la probabilidad de que la clase  $a$  contenga  $x$ .  $L_{ab}(x)$  da cuenta del grado de solapamiento entre las clases  $a$  y  $b$  [22]. Las probabilidades en la razón de verosimilitud se ilustran en la figura 18.

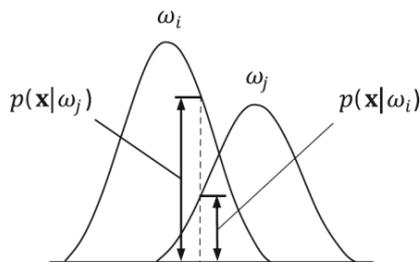


Figura 18: Probabilidades en la razón de verosimilitud [22].

Para clases totalmente diferentes se da entonces que  $L_{ab} = 0, \infty \quad \forall x$ . Si definimos  $L'_{ab}$  como 21:

$$L'_{ab} = \ln[p(x|a)] - \ln[p(x|b)] \quad (21)$$

la distancia de divergencia es la suma del valor esperado de las razones de verosimilitud logarítmicas condicionadas a cada clase definida según la ecuación 22.

$$d_{ab} = E[L'_{ab}|a] + E[L'_{ba}|b] \quad (22)$$

Para conjuntos de datos o distribuciones de probabilidad continuas:

$$E[L'_{ab}|a] = \int_x L'_{ab} p(x|a) dx \quad (23)$$

$$d_{ab} = \int_x (p(x|a) - p(x|b)) \ln \frac{p(x|a)}{p(x|b)} dx \quad (24)$$

Para cualquier tipo de distribuciones las distancias son siempre positivas o cero. En relación con esto, en caso de que las variables  $x_i$  sean independientes la probabilidad  $p(\mathbf{x}|a) = \prod_{i=1}^n p(x_i|a)$  de donde se deduce que  $d_{ab}(\mathbf{x}) = \sum_{i=1}^n d_{ab}(x_i)$ . Atendiendo a que la divergencia es siempre positiva en caso de aumentar el número de variables  $d_{ab}(x_1, \dots, x_{n+1}) \geq d_{ab}(x_1, \dots, x_n)$ .

En distribuciones normales entonces la métrica de divergencia se puede definir a través de la ecuación 25.

$$d_{ab} = \frac{1}{2} \text{tr}[(\Sigma_a - \Sigma_b)(\Sigma_a - \Sigma_b)] + \frac{1}{2} \text{tr}[(\Sigma_a - \Sigma_b)(\mu_a - \mu_b)(\mu_a - \mu_b)^T] \quad (25)$$

El principal inconveniente de esta métrica es su inexistente cota superior, lo cual imposibilita establecerla como un índice de separabilidad para cualquier conjunto de datos. Esto sucede porque a medida que alcanzamos el máximo de precisión en la clasificación la distancia de divergencia sigue aumentando de forma cuadrática. Es probable por tanto que esta medida de lugar a interpretaciones erróneas ya que un aumento considerable en  $d_{ab}$  no tiene porque suponer una mejoría considerable en la predicción [22].

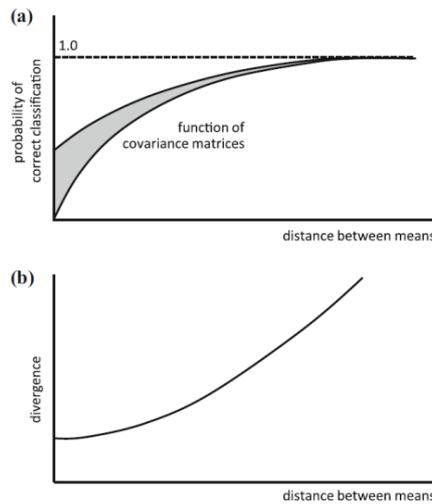


Figura 19: Distancia divergencia vs precisión en la clasificación [22].

### 3.5.3. Distancia de Jeffries-Matusita

Con el objetivo de obtener una distancia que en la completitud de su rango proporcione una medida representativa de la capacidad de distinción de un algoritmo predictivo ante dos conjuntos de datos se define la distancia de *Jeffries-Matusita* [22]. A diferencia de la distancia de divergencia, ésta cuenta con una cota superior.

$$JM_{ab} = \int_x (\sqrt{p(\mathbf{x}|a)} - \sqrt{p(\mathbf{x}|b)})^2 d\mathbf{x} \quad (26)$$

Para distribuciones normales (consideraremos que nuestros datos siguen distribuciones normales) podemos aproximar la expresión por:

$$JM_{ab} = 2(1 - e^{-B_{ab}}) \quad (27)$$

En la que  $B_{ab}$  hace referencia a la distancia de Bhattacharyya.

$$B_{ab} = \frac{1}{8} Mh_{ab} + \frac{1}{2} \ln \left[ \frac{(\Sigma_a + \Sigma_b)/2}{(|\Sigma_a| |\Sigma_b|)^{1/2}} \right] \quad (28)$$

Siendo  $Mh_{ab}$  a su vez la distancia de Mahalanobis.

$$Mh_{ab} = [(\mu_a - \mu_b)^T \left( \frac{\Sigma_a + \Sigma_b}{2} \right)^{-1} (\mu_a - \mu_b)] \quad (29)$$

Cuando dos distribuciones son considerablemente separables la distancia de *Jeffries-Matusita* tiende a 2 asintóticamente. En ese caso se entiende que un algoritmo de clasificación acertará eventualmente casi en el 100% de los casos.

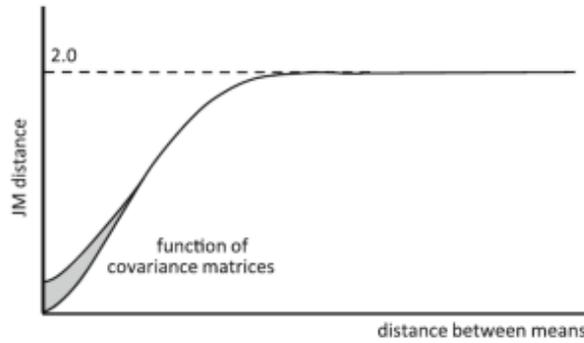


Figura 20: Distancia de *Jeffries-Matusita* en función de la distancia entre las medias [22].

## 4. Resultados experimentales

### 4.1. Base de datos

Nuestros experimentos se apoyan en una base de datos dedicada al desarrollo de sistemas ADAS a través de imágenes multispectrales. Estas imágenes han sido obtenidas por el grupo GDED de la UPV/EHU a partir de una cámara Photonfocus MV1-D2048x1088-HS02-96-G2 que incorpora el sensor de 25 bandas de *Imec*. Para hacerse con estas imágenes se colocó la cámara en el exterior de un vehículo y fue conectada a un ordenador portátil mediante un cable Ethernet como se muestra en la figura 21. Esta base de datos contiene cientos de imágenes que el grupo GDED ha puesto accesibles para la comunidad científica a través de una página web [23].



Figura 21: Montaje de la cámara en la parte frontal del vehículo (superior) y ordenador portátil grabando en el asiento del copiloto (inferior)[1].

Debido a que la cámara no ajusta automáticamente el tiempo de exposición a los cambios de iluminación, se escogieron  $10ms$  para los días soleados y  $20ms$  para los nublados. La información se codificaba en secuencias de 12 bits y se fijó una frecuencia de adquisición máxima de  $11fps$  para evitar consumir demasiada memoria [1].

La cámara multispectral en primera instancia nos proporciona imágenes en 2D que después de varias fases de preprocesamiento se transforman en cubos 3D. En primer lugar, se recortan las imágenes y se aplica la corrección de reflectancia comentada en el apartado 3.2. A continuación, opcionalmente se aplica a algunas imágenes un filtro de medianas de tamaño en píxeles  $3 \times 3$ . El filtro de medianas es una técnica de eliminación de ruido en imágenes. Posteriormente, se eliminan los patrones mosaico (ver apartado 3.3.2) de la imagen (*demosaicing*) a través de un proceso formado por una extracción y alineamiento de bandas (traslación al centro) basada en interpolación bilineal [1]. Este conjunto de imágenes lo denominamos *Base de datos-vehículo*.

El principal objetivo de esta base de datos era proporcionar diversidad de imágenes en cuanto a tipo de carreteras, condiciones de iluminación y climatología. Cada imagen en la base de datos está asociada a 4 documentos:

- El archivo binario en bruto (raw) de la cámara (.bin).
- El cubo de 25 bandas obtenido a partir de la imagen raw en una matriz de *Matlab* con extensión (.mat).

- Una imagen PNG (Portable Networks Graphics) que contiene una imagen RGB falsa (.png) para visualizar el contenido.
- Otra imagen (.png) que contiene las etiquetas de la clase a la que pertenece cada píxel de la imagen en cuestión.
- Asimismo archivos (.tar) que contiene una secuencias de video de 20 segundos de las cuales se han extraido las imágenes.

Adicionalmente, contamos con unas imágenes que no se encuentran dentro de esa base de datos y que han sido tomadas fuera del vehículo en condiciones de iluminación controladas. Es una colección de 20 imágenes (10 = número de clases definidas) en las que cada una de ellas contiene en su inmensa mayoría un material o clase. Estas imágenes adicionales conforman una base de datos aparte que nos sirve de referencia y que denominamos *Base de datos-materiales*. Cabe destacar que estas imágenes no cuentan con la aplicación del filtro de medianas. Dentro de esta reducida base de datos se distinguen dos subconjuntos: las imágenes tomadas sin filtro paso banda y las tomadas con filtro y corrección espectral de *Imec* (preprocesamiento hipotéticamente perfecto).

Para nuestros experimentos se han utilizado las siguientes imágenes:

1. Para unas pruebas preliminares un conjunto de imágenes de la *base de datos vehículo* (40 imágenes) a las que se les había aplicado un filtro de medianas.
2. Posteriormente, una vez ya familiarizados con el problema, las imágenes de la *base de datos materiales* para las pruebas de separabilidad y clasificación de dos materiales.
3. Finalmente, otro subconjunto de imágenes de la *base de datos vehículo* (100 imágenes) para la clasificación de un número mayor de materiales.

En cada base de datos se distinguen 10 clases entre las cuales se encuentran: asfalto, marcas, vegetación, metal pintado, cielo, cemento, peatón, agua, metal sin pintar y vidrio.

## 4.2. Metodología

El propósito de los experimentos es analizar varias técnicas de preprocesamiento dirigidas a optimizar la clasificación de materiales. Los esfuerzos se centran principalmente en conseguir una que se adapte lo mejor posible al caso sin filtro paso banda. El motivo es que la inmensa mayoría de la base de datos está construida en estas condiciones, puesto que reduce la transmitancia del sensor que ya de por sí es baja. No obstante, realizaremos una comparativa de los resultados obtenidos en los casos con y sin filtro y con y sin corrección espectral para determinar la mejor opción.

La metodología que se emplea en los experimentos es bastante intuitiva. A pesar de no atender al orden cronológico exacto, uno de los primeros pasos es calcular la matriz de cambio de base a la base ortogonal (Custom Spectral Correction) para el caso sin filtro paso banda. Una vez finalizada dicha tarea, el resto de experimentos se pueden distinguir en dos casos.

1. Tomando imágenes de la *base de datos-materiales* aplicamos las técnicas de preprocesamiento para posteriormente calcular la separabilidad y centrarnos en las que

se obtengan mejores resultados. Finalmente, éstas serán empleadas en la clasificación de materiales a través de la red neuronal y se contrastarán los resultados. De entre las diversas técnicas de preprocesamiento empleadas, las más satisfactorias (utilizando como medida de bondad la distancia de *Jeffries-Matusita*) son la *Matrix normalization*, *Band normalization* y *Normal distribution* que se explicarán más adelante. A su vez se utilizará la matriz de cambio de base.

2. Posteriormente, se realizan simulaciones más realistas. Tomando las imágenes de la *Base de datos-vehículo* aplicaremos las técnicas de preprocesamiento antes mencionadas, así como una adicional, *Clasificación por incrementos*. El objetivo será clasificar un distinto número de materiales en cada simulación; en particular 4, 6, 7 y 9 materiales.

### 4.3. *Setup experimental*

Los experimentos anteriormente mencionados se realizan sobre una red neuronal cuyos parámetros definimos a continuación. Se utilizan dos topologías una para el caso de la clasificación de dos materiales y otra más compleja en uno de los experimentos con mayor número de materiales a clasificar. En cuanto a la estructura y funcionamiento los parámetros más relevantes son:

- **Learning rate  $\alpha$ :** Es un parámetro que se refiere al tamaño de los pasos en la actualización de los pesos en cada iteración  $\alpha = 1.0$ .
- **Número de capas ocultas:** Capas intermedias, sin considerar la capa de entrada y salida  $N_1 = 1$  y  $N_2 = 2$ .
- **Número de neuronas de las capas ocultas:** El número de neuronas en las capas intermedias es para la primera topología  $n_{11} = 150$  y para la segunda  $n_{21} = 200$  y  $n_{22} = 500$ .
- **Método de entrenamiento:** Manera en la que se calculan las derivadas o se propaga el error hacia atrás *Scaled Conjugate Gradient Backpropagation*.
- **Función error:** Representa la expresión que va a determinar la magnitud del error cometido entre las predicciones y los valores reales de las muestras. Hemos escogido el más típico, el error cuadrático medio o *Mean Square Error*.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (30)$$

- **Máximas iteraciones sin mejora:** Número de iteraciones que se permiten para que el error de validación disminuya, en caso contrario se detiene la simulación  $max_1 = 50$  y  $max_2 = 80$ .
- **Límite de iteraciones:** Cada iteración o época se define como un ciclo completo sobre el conjunto de datos de entrenamiento  $n_{e1} = 200$  y  $n_{e2} = 20000$ .
- **Tandas:** Los pesos de la red neuronal se inicializan aleatoriamente y es posible que a través del descenso del gradiente en vez de llegar a un mínimo global de la función error no estancuemos en uno local. Para ello no se realiza un solo entrenamiento

sino varias tandas y obtenemos la tasa media. Se fija este parámetro en  $n_{t1} = 10$  y  $n_{t2} = 8$  puesto que las simulaciones en el segundo caso son mucho más costosas.

Asimismo también hay que tener presente como se realiza la división de los datos.

- **Porcentaje de datos de validación:** Es la cantidad de datos que se emplea para controlar el *Overfitting*  $v = 20\%$ .
- **Porcentaje de datos de entrenamiento:** Es la cantidad de datos empleados para actualizar los pesos a través del método de entrenamiento escogido  $e = 30\%$ .
- **Porcentaje de datos de test:** Son los datos que se utilizan para ver como se ajusta la red a datos desconocidos. Para estos datos no se actualizan los pesos  $t = 50\%$ .
- **Modo de división:** Para escoger el uso de cada dato, si test, validación o entrenamiento se emplea la función *Sample* de *Matlab* que los escoge aleatoriamente.

#### 4.4. Ortogonalización (*spectral correction*)

Aplicando el algoritmo de Gram-Schmidt para la base de producto interno infinito formada por las 25 bandas del sensor y utilizando la regla del trapecio para aproximar las integrales numericamente, obtenemos tanto la matriz de cambio de base, como las bandas virtuales en las que se expresaría las nuevas coordenadas o componentes de cada píxel. El procedimiento empleado para el cálculo de la nueva matriz y base se detalla en la sección 3.4. En la base ortogonal, el producto escalar entre los elementos es nulo.

El objetivo de la ortogonalización de las bandas del sensor es obtener una información más nítida en la que las componentes no se hallen tan entremezcladas y deshacernos de efectos de segundo orden como el *Crosstalk*. Sin embargo, la reducción de efectos de segundo orden no es el epicentro de la cuestión, sería mas bien una consecuencia casual.

El aporte porcentual de las 25 bandas del sensor para luz ortocolimada en las longitudes de ondas comprendidas entre  $400nm-1000nm$  producidos en saltos de  $1nm$  por un monocromador se esquematizan en la figura 22. Esta información se considera el modelo principal del sensor en el que se incluyen todos los efectos de primer y segundo orden.

Las 25 bandas virtuales obtenidas tras la ortogonalización se muestran en la figura 23. Es importante destacar que se trata de bandas virtuales que en pocos casos podrían representar la respuesta de un filtro real. Esto se debe a la aparición de aportes porcentuales negativos que darían lugar a reflectancias negativas, algo físicamente imposible.

#### 4.5. Separabilidad espectral

En esta sección nos centramos en realizar un estudio preliminar a la ejecución de la red neuronal para descartar opciones que a priori no resulten prometedoras. El motivo del estudio es ahorrar tiempo de cómputo y proporcionar una explicación más teórica y razonada del progreso que suponen ciertas técnicas. Para ello analizaremos la separabilidad entre las clases por medio de la distancia de *Jeffries-Matusita*

El objetivo principal es encontrar un preprocesamiento capaz de deshacerse de la influencia de la iluminación. Se emplearán técnicas de normalización para deshacernos

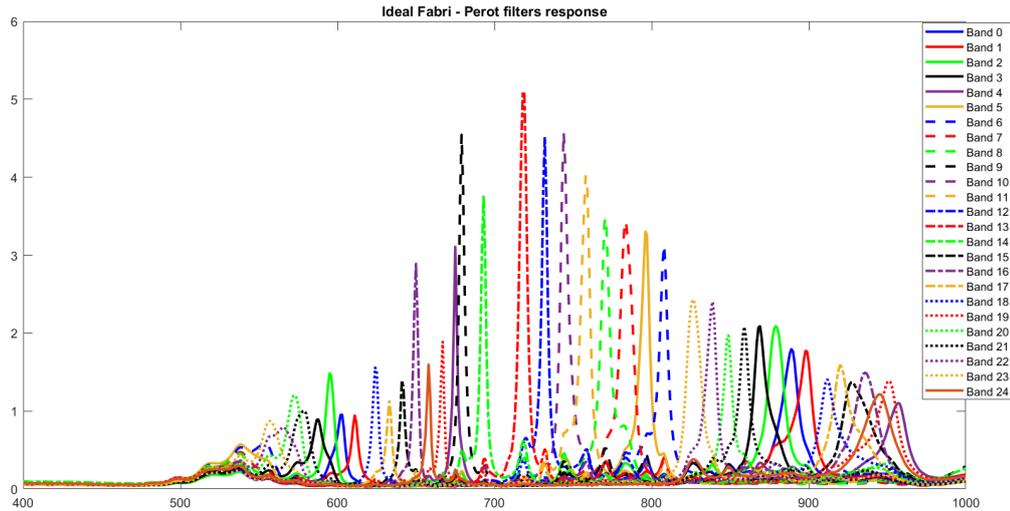


Figura 22: Aporte porcentual de los 25 filtros de Fabry-Perot para luz ortocolimada.

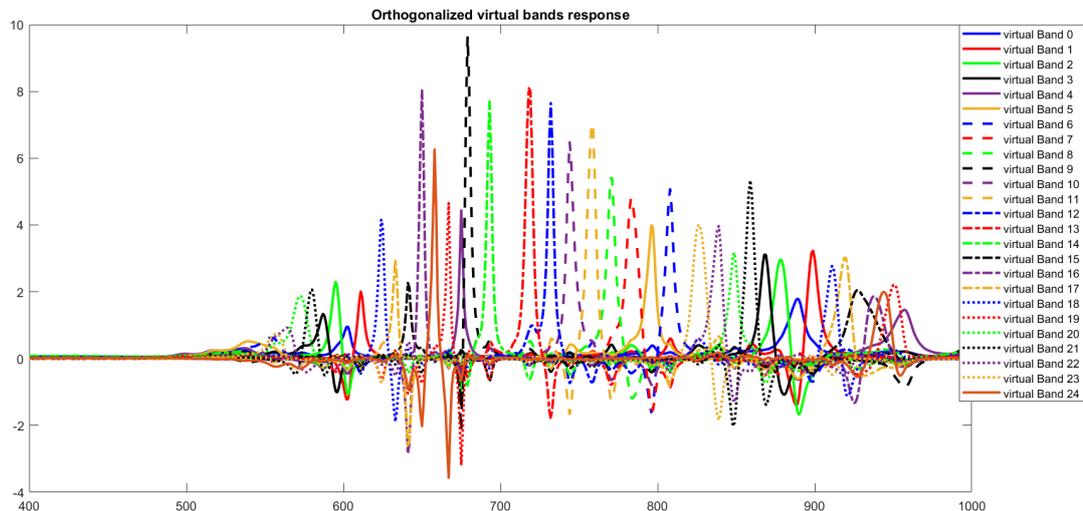


Figura 23: Aporte porcentual de las 25 bandas virtuales aplicando el método de Gram-Schmidt.

del aumento de proporcionalidad a mayor intensidad lumínica y del *offset* o componente común de la iluminación.

#### 4.5.1. Influencia de la iluminación

La separabilidad de los materiales esta fuertemente influenciada por la iluminación de la vía. En caso de que las imágenes se obtengan bajo distintas condiciones de iluminación los conjuntos de datos de las distintas clases aparentan estar más solapados, es decir, la información espectral aparece menos distinguible. En principio esto no debería ser así ya que según *Imec* transformando la irradiancia en reflectancia se elimina la dependencia con la iluminación. Sin embargo, estas afirmaciones se corroboran con los datos de la Tabla 1 y Tabla 2 en las que se muestran las 45 combinaciones posibles entre pares de materiales.

Tabla 1: Distancias de *Jeffries-Matusita* para imágenes de la *Base de datos-vehículo*. La media de las 45 combinaciones es  $JM_{dl} = 1,8338$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.8524	1.9293	1.6392	1.9939	1.6432	1.9051	1.999	1.7305	1.5551
Marcas	1.8524	0	1.8514	1.6605	1.9876	1.5773	1.9658	1.9985	1.7906	1.8985
Vegetación	1.9293	1.8514	0	1.6795	1.9996	1.716	1.923	1.9976	1.7831	1.9445
Metal pintado	1.6392	1.6605	1.6795	0	1.9954	1.3259	1.8298	1.9961	1.3741	1.3477
Cielo	1.9939	1.9876	1.9996	1.9954	0	1.9954	1.9997	2	1.9982	1.9942
Cemento	1.6432	1.5773	1.716	1.3259	1.9954	0	1.9061	1.998	1.6624	1.6321
Peaton	1.9051	1.9658	1.923	1.8298	1.9997	1.9061	0	1.9954	1.9067	1.8183
Agua	1.999	1.9985	1.9976	1.9961	2	1.998	1.9954	0	1.9965	1.9981
Metal sin pintar	1.7305	1.7906	1.7831	1.3741	1.9982	1.6624	1.9067	1.9965	0	1.7276
Vidrio	1.5551	1.8985	1.9445	1.3477	1.9942	1.6321	1.8183	1.9981	1.7276	0

Tabla 2: Distancias de *Jeffries-Matusita* para imágenes de la *Base de datos-materiales*. La media de las 45 combinaciones es  $JM_{ml} = 1,9424$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peatón	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	1.999	1.9736	2	1.9697	1.9088	1.8638	1.9849	1.8047
Marcas	1.9997	0	1.9567	1.9576	2	1.9702	1.9962	1.9872	1.8051	2
Vegetación	1.999	1.9567	0	1.9639	2	1.957	1.9016	1.9976	1.9684	2
Metal pintado	1.9736	1.9576	1.9639	0	2	1.8215	1.8632	1.954	1.7951	1.9974
Cielo	2	2	2	2	0	2	2	2	1.9103	2
Cemento	1.9697	1.9702	1.957	1.8215	2	0	1.8273	1.602	1.9174	1.9893
Peaton	1.9088	1.9962	1.9016	1.8632	2	1.8273	0	1.9835	1.9766	1.9915
Agua	1.8638	1.9872	1.9976	1.954	2	1.602	1.9835	0	1.9708	1.842
Metal sin pintar	1.9849	1.8051	1.9684	1.7951	1.9103	1.9174	1.9766	1.9708	0	1.9988
Vidrio	1.8047	2	2	1.9974	2	1.9893	1.9915	1.842	1.9988	0

Es evidente que en las imágenes obtenidas en las mismas condiciones de iluminación se refleja una notoria mejoría en la separabilidad. Intuimos que el principal motivo podría ser que se utiliza una imagen en blanco y otra en negro para todas las iluminaciones a la hora de realizar la corrección de reflectancia.

#### 4.5.2. Eliminación de la componente común de la iluminación

Como vimos en la sección 4.5.1 la iluminación juega un papel crucial en la separabilidad de los materiales. Para tratar de eliminar esa dependencia se emplean aquí distintas técnicas de preprocesamiento y se recalcula la separabilidad espectral. Hemos realizado numerosas pruebas lo que imposibilita incluir todas ellas en la memoria por el límite de extensión. Se incluyen únicamente las que han proporcionado los resultados más satisfactorios. Los ensayos se han realizado sobre imágenes de la *Base de datos-materiales*.

En primer lugar, destacar que las pruebas se han realizado tanto aplicando la matriz de corrección espectral obtenida a través del método de Gram-Schmidt como sin ella. Los resultados más interesantes se han conseguido con los siguientes métodos de preprocesamiento:

- **Band Normalization:** Llamamos *Band Normalization* al proceso de trasladar los datos de cada una de las bandas al rango  $[0,1]$ . Esto se consigue aplicando la fórmula 31 con  $i = 1, \dots, 88344$ , refiriéndose a los  $216 \times 409$  datos de cada banda que se corresponden con el número de patrones mosaico en una imagen, y  $j = 1, \dots, 25$  número de bandas.

$$X(i, j) = \frac{X(i, j) - \min_i(X(i, j))}{\max_i(X(i, j)) - \min_i(X(i, j))} \quad (31)$$

- **Matrix Normalization:** A diferencia del *Band Normalization* que realiza las normalizaciones banda por banda, en *Matrix Normalization* directamente se escoge el máximo y el mínimo considerando todos los datos de la matriz.

$$X = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (32)$$

- **Normal Distribution:** Totalmente al margen de las dos anteriores también hemos probado a asemejar los datos de cada imagen a distribuciones normales considerando toda la matriz. La ecuación 33 refleja esta transformación.

$$X = \frac{X - \mu(X)}{\sigma(X)} \quad (33)$$

Donde  $\mu(X)$  y  $\sigma(X)$  son la media y desviación estandar de una imagen.

A continuación, se exponen las Tablas 3, 4, 5, 6, 7 y 8 que contienen las distancias de *Jeffries-Matusita* para los 3 preprocesamientos anteriormente descritos con y sin la corrección espectral. Es importante tener en cuenta que partimos de una separabilidad inicial media de  $JM_{ml} = 1,9424$ .

Tabla 3: Distancias de *Jeffries-Matusita* para el uso de *Band Normalization* sin corrección espectral aportando una media  $JM_{ml-BN} = 1.9533$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peatón	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	1.9867	1.9347	1.9946	1.9886	1.9964	1.9946	1.7876	1.7749
Marcas	1.9997	0	1.7739	2	1.9668	1.9678	2	2	1.9947	1.9977
Vegetación	1.9867	1.7739	0	1.9993	1.9085	1.9296	2	1.9999	1.9597	1.9812
Metal pintado	1.9347	2	1.9993	0	1.9998	1.9999	1.8506	1.9396	1.8259	1.9299
Cielo	1.9946	1.9668	1.9085	1.9998	0	1.9769	2	1.9998	1.9912	1.9855
Cemento	1.9886	1.9678	1.9296	1.9999	1.9769	0	2	1.9999	1.9945	1.9908
Peaton	1.9964	2	2	1.8506	2	2	0	1.9972	1.9849	1.9931
Agua	1.9946	2	1.9999	1.9396	1.9998	1.9999	1.9972	0	1.9291	1.9904
Metal sin pintar	1.7876	1.9947	1.9597	1.8259	1.9912	1.9945	1.9849	1.9291	0	1.5846
Vidrio	1.7749	1.9977	1.9812	1.9299	1.9855	1.9908	1.9931	1.9904	1.5846	0

Tabla 4: Distancias de *Jeffries-Matusita* para el uso de *Band Normalization* con corrección espectral previa aportando una media  $JM_{ml-SC+BN} = 1.9826$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	1.9958	1.9886	1.9987	1.9997	1.9996	1.9988	1.9289	1.9701
Marcas	1.9997	0	1.9677	2	1.9966	1.9983	2	2	1.9996	1.9999
Vegetación	1.9958	1.9677	0	1.9997	1.9593	1.9999	2	2	1.9914	1.9971
Metal pintado	1.9886	2	1.9997	0	1.9999	2	1.954	1.9945	1.8208	1.9624
Cielo	1.9987	1.9966	1.9593	1.9999	0	1.9997	2	2	1.9935	1.9902
Cemento	1.9997	1.9983	1.9999	2	1.9997	0	2	2	1.9998	1.9999
Peaton	1.9996	2	2	1.954	2	2	0	1.9996	1.9904	1.996
Agua	1.9988	2	2	1.9945	2	2	1.9996	0	1.9844	1.997
Metal sin pintar	1.9289	1.9996	1.9914	1.8208	1.9935	1.9998	1.9904	1.9844	0	1.7472
Vidrio	1.9701	1.9999	1.9971	1.9624	1.9902	1.9999	1.996	1.997	1.7472	0

Se deduce directamente que los mejores resultados de separabilidad espectral dados por la distancia de *Jeffries-Matusita* son los proporcionados por la *Normal Distribution* junto con la corrección espectral de Gram-Schmidt. Esto puede estar relacionado con que ese preprocesamiento asemeja los datos a una distribución normal y tal y como calculamos la separabilidad implícitamente asumimos que contamos con este tipo de distribuciones.

Tabla 5: Distancias de *Jeffries-Matusita* para el uso de *Matrix Normalization* sin corrección espectral previa aportando una media  $JM_{ml-MN} = 1.922$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	1.9939	1.8226	1.9989	1.9457	1.9421	1.7701	1.7276	1.5636
Marcas	1.9997	0	1.9815	1.9997	1.9113	1.933	2	1.9996	1.9933	1.9989
Vegetación	1.9939	1.9815	0	1.9958	1.9963	1.8883	1.9984	1.9992	1.9893	1.9988
Metal pintado	1.8226	1.9997	1.9958	0	1.9995	1.9816	1.8811	1.9566	1.7951	1.8983
Cielo	1.9989	1.9113	1.9963	1.9995	0	1.9742	2	1.9994	1.9763	1.9983
Cemento	1.9457	1.933	1.8883	1.9816	1.9742	0	1.9986	1.8732	1.9248	1.866
Peaton	1.9421	2	1.9984	1.8811	2	1.9986	0	1.9911	1.9744	1.9529
Agua	1.7701	1.9996	1.9992	1.9566	1.9994	1.8732	1.9911	0	1.9294	1.2618
Metal sin pintar	1.7276	1.9933	1.9893	1.7951	1.9763	1.9248	1.9744	1.9294	0	1.8097
Vidrio	1.5636	1.9989	1.9988	1.8983	1.9983	1.866	1.9529	1.2618	1.8097	0

Tabla 6: Distancias de *Jeffries-Matusita* para el uso de *Matrix Normalization* con corrección espectral previa aportando una media  $JM_{ml-SC+MN} = 1.9846$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	2	1.9532	2	2	1.997	1.9958	1.843	1.9249
Marcas	1.9997	0	1.9989	2	2	2	2	1.9999	1.9989	1.999
Vegetación	2	1.9989	0	1.9998	2	1.996	2	2	1.999	2
Metal pintado	1.9532	2	1.9998	0	2	1.9998	1.944	1.9847	1.8368	1.9826
Cielo	2	2	2	2	0	2	2	2	1.9999	2
Cemento	2	2	1.996	1.9998	2	0	2	2	1.9983	2
Peaton	1.997	2	2	1.944	2	2	0	1.9992	1.9909	2
Agua	1.9958	1.9999	2	1.9847	2	2	1.9992	0	1.9649	2
Metal sin pintar	1.843	1.9989	1.999	1.8368	1.9999	1.9983	1.9909	1.9649	0	1.8998
Vidrio	1.9249	1.999	2	1.9826	2	2	2	2	1.8998	0

Tabla 7: Distancias de *Jeffries-Matusita* con la denominada *Normal Distribution* sin corrección espectral previa aportando una media  $JM_{ml-DN} = 1.9717$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	1.9998	1.9865	2	2	2	2	1.9972	1.9778
Marcas	1.9997	0	1.9966	1.9551	1.9985	1.9999	2	1.9992	1.794	1.9992
Vegetación	1.9998	1.9966	0	1.9742	1.9993	1.9971	1.9991	1.998	1.9796	2
Metal pintado	1.9865	1.9551	1.9742	0	1.9986	1.9868	1.9746	1.9286	1.7169	1.9968
Cielo	2	1.9985	1.9993	1.9986	0	1.7125	2	1.9951	1.9804	2
Cemento	2	1.9999	1.9971	1.9868	1.7125	0	2	1.9605	1.9222	2
Peaton	2	2	1.9991	1.9746	2	2	0	1.9982	1.9982	2
Agua	2	1.9992	1.998	1.9286	1.9951	1.9605	1.9982	0	1.9083	2
Metal sin pintar	1.9972	1.794	1.9796	1.7169	1.9804	1.9222	1.9982	1.9083	0	1.9989
Vidrio	1.9778	1.9992	2	1.9968	2	2	2	2	1.9989	0

Tabla 8: Distancias de *Jeffries-Matusita* con la denominada *Normal Distribution* con corrección espectral previa aportando una media  $JM_{ml-DN+SC} = 1.9956$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	1.9997	2	1.9745	2	2	2	2	1.9602	2
Marcas	1.9997	0	2	1.9999	2	2	2	2	1.994	2
Vegetación	2	2	0	2	2	2	2	2	2	2
Metal pintado	1.9745	1.9999	2	0	2	2	1.9996	1.9999	1.8978	1.9948
Cielo	2	2	2	2	0	2	2	2	2	2
Cemento	2	2	2	2	2	0	2	2	2	2
Peaton	2	2	2	1.9996	2	2	0	2	2	2
Agua	2	2	2	1.9999	2	2	2	0	1.9987	2
Metal sin pintar	1.9602	1.994	2	1.8978	2	2	2	1.9987	0	1.9847
Vidrio	2	2	2	1.9948	2	2	2	2	1.9847	0

Además, se observa una mejoría latente en todos los casos cuando se aplica la ortogonalización, lo cual indica que aparentemente es un método interesante. Por otro lado, el único caso en el que se observa una pequeña reducción en la separabilidad es en el de la *Matrix Normalization* sin corrección espectral.

## 4.6. Redes neuronales

En esta sección pasamos del cálculo estadístico teórico que sencillamente es un indicativo del posible rendimiento de un algoritmo de clasificación a las simulaciones reales que nos verifican el comportamiento de la red en los distintos casos. Evidentemente, no tiene porque cumplirse estrictamente la relación entre la separabilidad y los resultados de la red, ya que influyen más factores.

### 4.6.1. Correlación del índice de separabilidad con la tasa de acierto

Hasta el momento hemos partido de la premisa de que el índice de *Jeffries-Matusita* está directamente relacionado con las predicciones de la red neuronal. Para aportar validez a esta afirmación se ha ejecutado la red neuronal clasificando pares de clases utilizando las imágenes de la *Base de datos-materiales*. Como se puede observar en la Tabla 9 los resultados obtenidos corroboran la existencia de cierta relación entre las distancias estadísticas de *Jeffries-Matusita* y los resultados que arroja la red neuronal.

Tabla 9: Tasas de acierto de la red neuronal para la clasificación entre pares de materiales de la *Base de datos-materiales*.

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	99.68	99.863	98.443	100	99.293	99.264	97.651	96.959	96.071
Marcas	99.68	0	99.086	94.701	100	98.555	99.032	96.706	95.527	99.791
Vegetación	99.863	99.086	0	99.458	100	99.303	97.41	99.874	99.465	99.978
Metal pintado	98.443	94.701	99.458	0	99.668	99.424	99.163	99.509	96.115	99.638
Cielo	100	100	100	99.668	0	100	100	99.997	98.24	100
Cemento	99.293	98.555	99.303	99.424	100	0	98.311	94.175	96.358	99.641
Peaton	99.264	99.032	97.41	99.163	100	98.311	0	99.853	97.408	99.964
Agua	97.651	96.706	99.874	99.509	99.997	94.175	99.853	0	97.428	96.03
Metal sin pintar	96.959	95.527	99.465	96.115	98.24	96.358	97.408	97.428	0	98.561
Vidrio	96.071	99.791	99.978	99.638	100	99.641	99.964	96.03	98.561	0

A pesar de que la correlación no sea perfecta se observa fácilmente la relación directa entre la efectividad de la red y la separabilidad espectral de los materiales. Por ejemplo, atendiendo a las distancias de la columna de la clase *Cielo* en la Tabla 2 vemos que son prácticamente 2 o muy cercanas y eso se plasma en forma de una tasa de predicción casi perfecta en la Tabla 9. A menor distancia estadística peor es teóricamente la respuesta de la red. Obviamente no se cumple a la perfección, se observa claramente que la peor distancia *Metal sin pintar vs Metal pintado* no es la que peor se clasifica siendo esta *Marcas vs Metal pintado*. Sin embargo, *Metal sin pintar vs Metal pintado* se corresponde con una tasa relativamente baja. Ello puede ser debido a que los datos siguen una distribución normal para el cálculo de las distancias, hecho que hemos asumido pero no podemos afirmar rotundamente.

La tasa de acierto media que resume los resultados de la Tabla 9 es  $Tasa_{media} = 98.5687$ . De donde se deduce que para la clasificación de dos materiales el modelo funciona casi a la perfección sin incluir ningún tratamiento adicional. No obstante, ya veremos que cuando incrementamos el número de materiales a clasificar además de trabajar en condiciones no tan controladas de iluminación los resultados son más discretos.

### 4.6.2. Clasificación con diferentes métodos de preprocesamiento

En esta sección se comprueba la utilidad en la clasificación de los métodos de preprocesamiento para la eliminación de la componente común de la iluminación comentados en

la sección 4.5.2. Adicionalmente, también se estudia el rendimiento que otorga la utilización de uno de los filtros de *Ivec* con su correspondiente corrección espectral. La tasa de partida en las imágenes que utilizamos correspondientes a la *Base de datos-materiales* sin preprocesamiento es  $Tasa_{media} = 98.5687$ . Debido al límite de extensión de la memoria solo se muestran los resultados para el mejor caso entre la utilización o no de la matriz de corrección espectral. Sin embargo, la media de las tasas la damos para los dos casos. Los resultados obtenidos se muestran en las Tablas 10, 11, 12 y 13.

Tabla 10: Tasas de acierto de la red neuronal para las imágenes de la *Base de datos-materiales* aplicando la *Band Normalization* sin corrección espectral. La tasa media para el caso sin corrección es  $tasa_{BN} = 98.4047$  y en caso de realizar el cambio de base es  $tasa_{SC+BN} = 97.8945$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
<b>Asfalto</b>	0	99.669	99.858	98.477	100	99.265	92.889	97.666	96.955	96.137
<b>Marcas</b>	99.669	0	99.06	94.599	99.999	98.528	98.982	96.718	95.445	99.776
<b>Vegetación</b>	99.858	99.06	0	99.413	99.999	99.319	97.325	99.873	99.489	99.98
<b>Metal pintado</b>	98.477	94.599	99.413	0	99.611	99.377	99.18	99.463	96.236	98.615
<b>Cielo</b>	100	99.999	99.999	99.611	0	100	100	99.995	98.268	100
<b>Cemento</b>	99.265	98.528	99.319	99.377	100	0	98.254	94.062	95.969	99.651
<b>Peaton</b>	92.889	98.982	97.325	99.18	100	98.254	0	99.867	97.395	99.976
<b>Agua</b>	97.666	96.718	99.873	99.463	99.995	94.062	99.867	0	97.401	96.036
<b>Metal sin pintar</b>	96.955	95.445	99.489	96.236	98.268	95.969	97.395	97.401	0	99.432
<b>Vidrio</b>	96.137	99.776	99.98	98.615	100	99.651	99.976	96.036	99.432	0

Tabla 11: Tasas de acierto de la red neuronal para las imágenes de la *Base de datos-materiales* aplicando la *Matrix Normalization* sin corrección espectral. La tasa media para el caso sin corrección  $tasa_{MN} = 98.5460$  y para el caso con corrección de  $tasa_{SC+MN} = 98.0039$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
<b>Asfalto</b>	0	99.685	99.859	98.506	100	99.31	99.25	97.622	96.854	95.941
<b>Marcas</b>	99.685	0	99.018	94.448	100	98.593	99.068	96.628	95.212	99.786
<b>Vegetación</b>	99.859	99.018	0	99.334	100	99.284	97.314	99.891	99.441	99.983
<b>Metal pintado</b>	98.506	94.448	99.334	0	99.645	99.392	99.177	99.521	96.293	99.653
<b>Cielo</b>	100	100	100	99.645	0	100	100	99.995	98.236	100
<b>Cemento</b>	99.31	98.593	99.284	99.392	100	0	98.278	93.96	95.155	99.65
<b>Peaton</b>	99.25	99.068	97.314	99.177	100	98.278	0	99.86	97.34	99.965
<b>Agua</b>	97.622	96.628	99.891	99.521	99.995	93.96	99.86	0	97.43	96.115
<b>Metal sin pintar</b>	96.854	95.212	99.441	96.293	98.236	95.155	97.34	97.43	0	99.873
<b>Vidrio</b>	95.941	99.786	99.983	99.653	100	99.65	99.965	96.115	99.873	0

Tabla 12: Tasas de acierto de la red neuronal para las imágenes de la *Base de datos-materiales* aplicando la *Normal Distribution* sin corrección espectral. La tasa media sin corrección espectral es de  $tasa_{DN} = 98.4602$  y con corrección  $tasa_{SC+DN} = 98.0137$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
<b>Asfalto</b>	0	99.664	99.863	98.482	100	99.248	99.229	97.681	96.896	96.025
<b>Marcas</b>	99.664	0	99.059	94.605	99.999	98.489	99.029	96.709	95.499	99.771
<b>Vegetación</b>	99.863	99.059	0	99.393	100	99.312	97.442	99.879	99.44	99.983
<b>Metal pintado</b>	98.482	94.605	99.393	0	99.633	99.41	99.128	99.484	96.118	99.648
<b>Cielo</b>	100	99.999	100	99.633	0	100	100	99.993	98.349	100
<b>Cemento</b>	99.248	98.489	99.312	99.41	100	0	98.291	94.086	95.906	99.65
<b>Peaton</b>	99.229	99.029	97.442	99.128	100	98.291	0	99.84	97.248	99.972
<b>Agua</b>	97.681	96.709	99.879	99.484	99.993	94.086	99.84	0	97.444	90.932
<b>Metal sin pintar</b>	96.896	95.499	99.44	96.118	98.349	95.906	97.248	97.444	0	99.878
<b>Vidrio</b>	96.025	99.771	99.983	99.648	100	99.65	99.972	90.932	99.878	0

La conclusión inmediata es que los resultados son muy similares, a pesar de la utilización o no de las técnicas de preprocesamiento. En contra de lo esperado, los peores resultados de clasificación son los que provienen de las recomendaciones del fabricante. Tratando

Tabla 13: Tasas de acierto de la red neuronal para las imágenes de la *Base de datos-materiales* aplicando el filtro pasa banda y la corrección espectral de *Imec*. La tasa media obtenida es  $tasa_{Imec} = 97.6456$ .

	Asfalto	Marcas	Vegetación	Metal pintado	Cielo	Cemento	Peaton	Agua	Metal sin pintar	Vidrio
Asfalto	0	99.173	99.935	97.283	100	93.852	99.786	91.508	96.586	93.834
Marcas	99.173	0	99.698	88.694	100	95.867	98.611	92.977	90.764	99.642
Vegetación	99.935	99.698	0	99.653	99.998	99.716	97.733	99.84	99.616	99.976
Metal pintado	97.283	88.694	99.653	0	98.765	98.408	99.263	99.637	97.3	99.84
Cielo	100	100	99.998	98.765	0	100	99.999	99.999	98.05	100
Cemento	93.852	95.867	99.716	98.408	100	0	98.144	88.649	94.802	99.065
Peaton	99.786	98.611	97.733	99.263	99.999	98.144	0	99.192	98.1	99.934
Agua	91.508	92.977	99.84	99.637	99.999	88.649	99.192	0	94.124	96.894
Metal sin pintar	96.586	90.764	99.616	97.3	98.05	94.802	98.1	94.124	0	99.145
Vidrio	93.834	99.642	99.976	99.84	100	99.065	99.934	96.894	99.145	0

de dar una explicación razonable, considero necesario realizar las siguientes aclaraciones. Por un lado, es importante mencionar que las tasas de clasificación son prácticamente perfectas, y pequeñas variaciones en torno a ellas no implican un peor desempeño en simulaciones más complejas. Por otro lado, hay que tener en cuenta que las imágenes utilizadas se obtuvieron en condiciones de iluminación más o menos controladas. Por lo tanto ese *offset* o componente común de la iluminación y la mayor proporcionalidad a mayor intensidad de la que pretendíamos deshacernos con las técnicas de preprocesamiento no se manifiestan en las imágenes de esta base de datos. Deduciendo así, que simplemente el mantenernos en el rango de clasificación de los resultados correspondientes a no tratar las imágenes se puede considerar un buen resultado.

Por último, se ha visto que la correlación con el índice de *Jeffries-Matusita* que en el experimento inicial de clasificación era tan buena, aparece aquí un poco distorsionada. Los motivos pueden ser debidos tanto a la increíble tasa en la clasificación, como a que la normalización se realiza en una misma variable que contiene las dos imágenes (dos materiales) en lugar de cada una por separado. Sin embargo, esto no debería ser un problema ya que en la vía pública una misma imagen contendrá distintas clases o materiales. Otro motivo aparente podría ser el aumento de la separabilidad intraclase, refiriendonos al aumento en la dificultad de asociar un material a su clase. Esto se debe a que los índices de la diagonal nulos, no lo serían en caso de comparar dos imágenes diferentes de un mismo material. Por tanto, podría ocurrir que esos elementos de la diagonal aumentasen hacia valores de separabilidad más elevados una vez aplicadas las técnicas de preprocesamiento.

#### 4.6.3. Clasificación de un mayor número de materiales

A la vista de los resultados y viendo que no se puede tomar una decisión en cuanto al mejor método contando con información tan limitada, aumentamos la complejidad del problema. A continuación, realizamos tareas de clasificación de diferente número de materiales con imágenes de la *Base de datos-vehículo*. Los resultados obtenidos se muestran en la tabla 14. Se corresponden con la clasificación de 4, 6, 7 y 9 materiales para cada método. Hemos prescindido de la clase *agua* por la falta de muestras en las imágenes.

En vista de los resultados obtenidos se deduce que la corrección espectral en estos casos no resulta útil, sino que penaliza la clasificación sistemáticamente. Por otro lado, a pesar de que en la clasificación de pares de materiales no hemos obtenido mejoras, en estos experimentos, las 3 normalizaciones aumentan la precisión de la red en alguno de los casos. Concretamente, *Band Normalization* en 3 de los casos (4 materiales, 6 materiales y

Tabla 14: Tasas de acierto de la red neuronal para la clasificación de distinto número de materiales. Se muestran todas las normalizaciones consideradas.

	9 materiales	7 materiales	6 materiales	4 materiales
<i>Nada</i>	59,704	74,8202	78,8674	86,479
<i>SC</i>	58,2867	72,2412	77,0547	85,527
<i>BN</i>	59,906	71,7666	81,6558	87,9656
<i>SC+BN</i>	52,3509	70,3925	76,4316	82,1726
<i>MN</i>	56,5933	73,2472	78,4589	88,9706
<i>SC+MN</i>	54,2909	70,5594	74,6742	85,6856
<i>DN</i>	59,6698	74,061	80,6184	88,8292
<i>SC+DN</i>	54,5112	71,1233	78,0453	85,298

9 materiales), *Matrix Normalization* en uno (4 materiales) y *Normal distribution (DN)* en dos (6 materiales y 4 materiales). No obstante, estas mejorías son escasas por lo que sería interesante considerar algún método alternativo, como el que se presenta a continuación.

#### 4.6.4. Clasificación por incrementos (derivada)

En la sección anterior hemos visto cómo los distintos preprocesamientos no solventan el problema de la iluminación de manera evidente. Es por ello que se propone un método alternativo. Este es la derivada o la toma de los incrementos entre las distintas bandas de un mismo píxel. Si denominamos  $b_i$  a cada uno de los valores referentes a la banda  $i = 1, \dots, 25$  de un píxel, matemáticamente las nuevas variables de entrada serán las  $p_i$  dadas por:

$$p_i = \frac{b_i - b_{i+1}}{\sum_{i=1}^{25} b_i} \quad (34)$$

Debido a la mejoría en la clasificación dada por las técnicas de preprocesamiento, éstas se utilizan junto a la derivada (D) obteniendo así la tabla 15.

Tabla 15: Tasas de acierto de la red neuronal para la clasificación de distinto número de materiales considerando la derivada y las distintas normalizaciones. La *Band Normalization* solo se aplica posteriormente a la derivada mientras que las otras dos se aplican tanto previa como posteriormente.

	9 materiales	7 materiales	6 materiales	4 materiales
<i>D</i>	64,7302	73,0645	81,0868	89,1371
<i>D+SC</i>	59,06	73,7452	80,9579	86,2083
<i>D+BN</i>	64,4095	74,8148	83,1089	89,2311
<i>SC+D+BN</i>	61,2646	71,8663	78,1184	85,8814
<i>D+MN</i>	65,5965	76,4407	83,2774	89,0866
<i>SC+D+MN</i>	60,5393	75,7847	79,8132	87,9964
<i>D+DN</i>	66,1656	75,796	80,2084	87,169
<i>SC+D+DN</i>	61,9168	73,4736	81,0158	88,0682
<i>MN+D</i>	66,1709	76,7325	80,1358	89,1008
<i>DN+D</i>	64,554	75,8795	81,8137	89,5461

Como se observa en la Tabla 15 hay una mejoría evidente si se aplica la derivada, tanto en cuanto a la mejor clasificación de los materiales, sobre todo en los casos más complejos (los de real interés), como a la reducción en una variable del espacio de entrada reduciendo así la complejidad del modelo. Los mejores resultados son los obtenidos con la *Matrix Normalization* previa a la derivada para el caso más complejo (9 materiales).

Un posible motivo es el hecho de que dicha técnica se asemeja bastante a la corrección de reflectancia que propone *Imec*. La corrección de reflectancia resta una imagen en negro tanto en el numerador como en el denominador para eliminar el ruido estático que se podría considerar el mínimo de irradiancia (asemajandolo al mínimo de toda la matriz). Por otro lado, la imagen en blanco es un máximo de radiancia que se podría considerar como el máximo de la imagen. Por lo que atendiendo a ese hecho y teniendo en cuenta que la corrección de reflectancia realizada no es del todo rigurosa (no sacamos una foto en blanco y en negro en cada caso para cada iluminación) es lícito considerar que esta mejora es debida a una especie de segunda corrección de reflectancia. La mejoría en este caso es bastante apreciable, en el caso más interesante (9 materiales) concretamente es del  $m = \frac{66.1709 - 59.704}{59.704} 100 = \%10.8316$ .

Como comparación adicional, se muestran en la figura 24 las matrices de confusión del caso sin preprocesamiento, con corrección espectral, y el mejor caso (*Matrix Normalization* + derivada). En la esquina inferior derecha se aprecian las tasas medias de  $\%60.1$  y  $\%68.1$  respectivamente. Las matrices de confusión son las resultantes de la mejor de las 10 tandas.

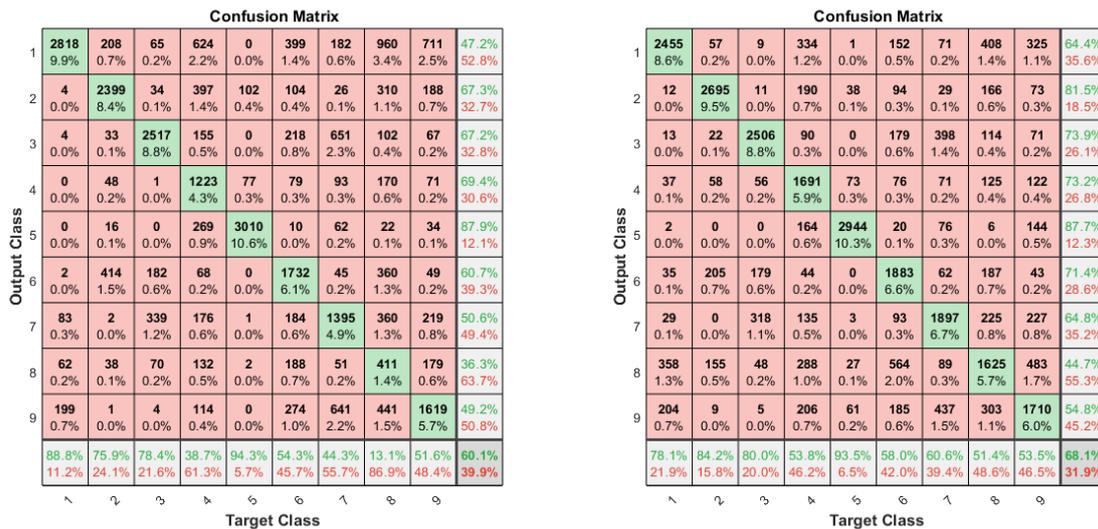


Figura 24: Matrices de confusión para la clasificación de 9 materiales. Sin preprocesamiento y con corrección espectral (izquierda), y con *Matrix Normalization* + derivada (derecha).

Una vez que se ha comprobado que la derivada aporta una mejoría más clara, se decide añadir a los experimentos de la sección 4.6.2 (entre pares) una nueva prueba incluyendo ahora la derivada. A falta de espacio se ha decidido mostrar únicamente las tasas medias de las 45 combinaciones de dos materiales en la Tabla 16.

A la vista de los resultados, sacar conclusiones no contradictorias resulta difícil ya que la única mejoría considerable en este caso resulta de la aplicación de la matriz de corrección espectral. Es posible que la mejoría sea debida a la eliminación de la correlación entre bandas. Esto se vería reflejado en el cálculo de la derivada, ya que la mayor radiancia en una banda no implicaría mayor radiancia en otra mediante ninguna relación de proporcionalidad. En contraposición, como veíamos anteriormente, en el caso de distinto número de

Tabla 16: Tasas de acierto medias para la clasificación de dos materiales con los distintos métodos considerados en esta sección.

	<i>sin SC</i>	<i>con SC</i>
<b><i>D</i></b>	97.5168	98.4031
<b><i>D+BN</i></b>	97.3908	98.5957
<b><i>D+MN</i></b>	97.4670	98.5835
<b><i>D+DN</i></b>	97.4343	98.5000
<b><i>MN+D</i></b>	97.4213	98.5108
<b><i>DN+D</i></b>	97.4397	98.6015

materiales a clasificar obteníamos peores resultados. Bien es cierto que las tasas de clasificación en este caso son de extrema precisión y quizás la conclusión resultante sea que la clasificación de dos materiales no aporta información relevante. Evidentemente, las tasas de clasificación más interesantes son las del caso más complicado (9 materiales en distintas condiciones de iluminación) que son las que más se asemejan a la realidad. Por otro lado, para esta última clasificación utilizamos imágenes de la *Base de datos-materiales* por lo que eran previsible resultados similares para los distintos métodos de normalización.

#### 4.6.5. Aumento de la complejidad del modelo: cambio de la topología de la red

En vista de que para casos más simples como la clasificación de dos materiales con la derivada la matriz de corrección espectral aporta información relevante, se ha considerado modificar la topología de la red a una como la que se comenta en el *setup* experimental. El motivo principal es que quizás en casos complejos la red esté subentrenada, además de no contar con la topología adecuada. Es decir, tenemos que asegurarnos de que los resultados contradictorios en cuanto a la matriz de corrección espectral no sean por la falta de entrenamiento. Para ello, se ha repetido el experimento de la derivada con y sin corrección espectral para distinto número de materiales a clasificar: 4, 6, 7 y 9. Los resultados obtenidos se muestran en la Tabla 17.

Tabla 17: Tasas de acierto medias para la clasificación de distinto número de materiales haciendo uso de la derivada con y sin corrección espectral.

	<b>9 materiales</b>	<b>7 materiales</b>	<b>6 materiales</b>	<b>4 materiales</b>
<b><i>D</i></b>	78,46	88,65	91,88	94,83
<b><i>SC+D</i></b>	84,99	91,21	93,19	95,58

Efectivamente, el principal problema era la topología de la red que en un caso de semejante complejidad no estaba extrayendo la totalidad de la información. Asimismo, a la vista de estos resultados podemos concluir que la ortogonalización aporta información que mejora la clasificabilidad en caso de utilizar conjuntamente la derivada. La mejoría está latente en todos los casos por lo que no se trata de una mera casualidad. Además es una mejoría notoria sobre todo en el caso de 9 materiales. Concretamente, una mejora en la tasa de acierto media del 6.53%. Como ya habíamos comentado esto podría deberse a la eliminación de la proporcionalidad entre bandas debida a la ortogonalización.

## 5. Conclusiones y futuros trabajos

### 5.1. Conclusiones

En el ámbito automovilístico el *Machine Learning* así como el *Data Mining* son los pilares que fundamentan el paradigma de la conducción autónoma. Idea hoy en día en desarrollo y en la que grandes empresas están involucradas.

El trabajo se ha enfocado en el tratamiento de imágenes multiespectrales tomadas a partir del sensor CMV2K-SM5x5-NIR de tecnología CMOS. A lo largo de éste se detallan las características del sensor, los fundamentos de las redes neuronales, conceptos algebraicos y los índices probabilísticos utilizados en este tipo de aplicaciones.

El principal objetivo de este trabajo era determinar cual de las técnicas de preprocesamiento de imágenes resultaba más efectiva para la clasificación de materiales con imágenes multiespectrales a través de redes neuronales. A lo largo del trabajo se ha empleado una red neuronal *feedforward* simple como método de clasificación para evaluar las técnicas utilizadas. Los parámetros se precisan en el apartado de *setup* experimental.

Inicialmente, para hacer una estimación de la eficacia de cada técnica y dada la gran cantidad de posibilidades se decidió utilizar una medida de separabilidad estadística para determinar cuáles eran más prometedoras. Entre las cuales se destacaron tres de ellas: *Matrix Normalization*, *Band Normalization* y *Normal Distribution*. Siempre en presencia de la ortogonalización de Gram-Schmidt en la cual se depositaron grandes esperanzas. Por ese motivo se escogieron éstas para el preprocesamiento y posterior clasificación de los materiales a través de la red neuronal.

En la clasificación de materiales por pares con las imágenes de la *Base de datos-materiales* no se detectó ninguna mejora por lo que se decidió realizar un experimento más en profundidad. Tomando imágenes de la *Base de datos-vehículo* se clasificaron distinto número de materiales con las técnicas de preprocesamiento consideradas. A la vista de que los resultados mejoraban de manera poco significativa introdujimos como tratamiento adicional, la derivada o los incrementos banda a banda. Esta si produjo una mejora más notoria. El caso más exitoso fue el que combinaba la *Matrix Normalization* junto con la derivada. Sin embargo, en contra de lo esperado la matriz de ortogonalización empeoraba los resultados. Para mayor contradicción tomamos la derivada y las distintas técnicas de preprocesamiento para la clasificación por pares de materiales (*Base de datos-materiales*) y se observó una mejora latente en el uso de la ortogonalización en todos los casos. Por ello intuimos que para casos más complejos la red estaba subentrenada. Se decidió así, cambiar la topología de la red y el número de iteraciones y errores de validación para extraer el máximo de información y determinar el verdadero valor de la ortogonalización.

Realizamos el experimento de la derivada con y sin la corrección espectral y vimos cómo en la totalidad de los casos la ortogonalización mejoraba la clasificación. Así pues, podemos asegurar con cierto rigor que dos técnicas de preprocesamiento efectivas utilizadas conjuntamente son el uso de la derivada así como la ortogonalización de Gram-Schmidt. Los resultados con la red bien entrenada solo los contrastamos entre el uso o no de la corrección espectral entre los cuales se apreció una diferencia en la tasa de acierto media (en el caso más complejo) del 6.53%. La derivada consideramos que aporta información

válida debido a que en el caso de la red subentrenada mejoraba la tasa en todos los experimentos. Sin embargo, esto es necesario comprobarlo realizando todos los experimentos con una red del tamaño adecuado. Por otro lado, aunque la mejoría fuese escasa y los resultados no del todo significativos debido al poco entrenamiento de la red, los tres preprocesamientos aportaban mejoras que es necesario contrastar con nuevos experimentos. Dichos experimentos se podrían derivar a futuras actividades de investigación. Entre los preprocesamientos destacaba la *Matrix Normalization* quizás debido a la similitud con la corrección de reflectancia (siendo además esta última no del todo rigurosa).

Finalmente, concluir diciendo que a lo largo de este trabajo se han abordado materias hoy en día en auge y englobadas dentro de distintos campos o disciplinas. Entre las cuales se encuentra la inteligencia artificial, el *Data Mining*, los sensores multiespectrales, la estadística, el álgebra y los métodos de interpolación lineal. Es decir, es un trabajo muy polifacético en el que se dan a conocer muchos conceptos interesantes, muchos de los cuales constituyen hoy en día ámbitos de investigación de la comunidad científica.

## 5.2. Futuras actividades de investigación derivadas del TFG

Partiendo de los resultados obtenidos en el trabajo se deduce que hay varias ramas de investigación que aún precisan un desarrollo futuro. En general, la problemática principal de la que se derivarían la mayoría de actividades futuras sería la siguiente: encontrar el menor conjunto de variables de entrada así como la mejor topología que proporcione el máximo rendimiento a ser posible a una red neuronal combinada con el mejor preprocesamiento.

Para lograr minimizar el número de variables (por ejemplo tratando de eliminar redundancias) una técnica interesante son los algoritmos evolutivos, en nuestro caso aplicados a redes neuronales. Este tipo de algoritmos ya se han considerado en estas aplicaciones por otros investigadores como por ejemplo en [24], donde consideran aplicar un algoritmo genético para la selección de bandas con el objetivo de maximizar la eficiencia de un sistema SVM. No obstante, en este artículo parten de imágenes tomadas por una cámara hiperespectral con un número mucho mayor de bandas, lo que les obliga a hacer una partición en subgrupos. En cuanto a la topología de la red este tipo de algoritmos también se pueden aplicar para encontrar la que mejor se adapte al problema concreto. El grupo GDED desarrolló en su momento un algoritmo genético para este tipo de aplicaciones [25]. Este algoritmo es aplicable a cualquier problema que involucre reconocimiento de materiales a través de imágenes multiespectrales o hiperespectrales. Por lo que sería posible su utilización en nuestro problema. El algoritmo se aplica a una ELM (Extreme Learning Machine) redes *feedforward* con un método de aprendizaje basado en la matriz inversa de Moore-Penrose que son muy eficientes en cuanto a tiempo de procesamiento. En este artículo también se comenta alguna desventaja referente a la utilización de técnicas PCA (*Principal Component Analysis*) en cuanto a encarar el costo del preprocesamiento.

Por otro lado, otra futura actividad imprescindible asociada al trabajo realizado es acelerar la parte del preprocesamiento de imágenes. La solución más evidente a esto es la implementación del tratamiento previo en *Hardware*. Esta necesidad surge del mayor tiempo de cómputo consumido en el tratamiento de las imágenes en vez de en la clasificación, ya que éste todavía se encuentra en *Software* y limita la frecuencia de operación de la cámara.

## Referencias

- [1] K. Basterretxea, V. Martínez, J. Echanobe, J. Gutierrez-Zaballa, and I. del Campo, “Hsi-drive: A dataset for the research of hyperspectral image processing applied to autonomous driving systems,” July 11-17, 2021.
- [2] “10 differences between multispectral and hyperspectral,” -10 2021. Última consulta: 18-04-2023.
- [3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, “Hyperspectral remote sensing data analysis and future challenges,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, pp. 6–36, -06 2013.
- [4] C. C. Aggarwal, *Neural networks and deep learning*. Cham, Switzerland: Springer, 2018.
- [5] C. M. Bishop, *Neural Networks for Pattern Recognition*. United States, New York: Oxford University Press, 13 ed., 2005.
- [6] Iberdrola, “Descubre cómo el ‘data mining’ predecirá nuestro comportamiento.” Última consulta: 20-03-2023.
- [7] E. Bello, “¿qué es el minado de datos o data mininig? técnicas y pasos a seguir,” *Thinking for Innovation*, -12 2021. Última consulta: 21-03-2023.
- [8] M. North, *Data Mining for the Masses*. 2012.
- [9] U. de Alcalá, “Diferencia entre data mining y text mining,” -07 2018.
- [10] Tesla, “AI & robotics.” Última consulta: 15-3-2023.
- [11] H. Martensson, O. Keelan, and C. Ahlstrom, “Driver sleepiness classification based on physiological data and driving performance form real road driving,” *IEEE*, vol. 20, pp. 421–430, 02/ 2019.
- [12] W. Zhang, H. Liu, X. Wu, L. Xiao, Y. Quian, and Z. Fang, “Lane marking detection and classification with combined deep neural network for driver assistance,” *Journal of Automobile Engineering*, pp. 1259–1268, 2019.
- [13] P. Ping, w. Qin, Y. Xu, C. Miyajima, and T. Kazuya, “Spectral clustering based approach for evaluating the effect of driving behaviour on fuel economy,” in *IEEE International Instrumentation and Measurement Technology Conference*, (Houston, TX, USA), pp. 1–6, 2018 2018.
- [14] K.J.Lee, K.Bong, C.Kim, J.Jang, H.Kim, J.Lee, K.-R.Lee, G.Kim, and H.-J. Yoo, “14.2 a 502gops and 0.984 mw dual-mode adas soc with rnn-fis engine for intention prediction in automotive black-box system,” in *IEEE Int. Solid-State Circuits Conference*, pp. 256–257.
- [15] K. Okamoto, K. Berntorp, and S. D. Cairano, “Driver intention-based vehicle threat assessment using random forests and particle filtering,” *IFAC-PapersOnLine*, vol. 50, 2017.

- [16] Imec, “Hyperspectral sensors technology overview,” -2 2018.
- [17] D. Bolles, “Introduction to the electromagnetic spectrum — science mission directorate.” Última consulta: 11-4-2023.
- [18] I. E. Shaaban, A. S. Samra, S. Muhammad, and S. Wageh, “Design of distributed bragg reflectors for green light-emitting devices based on quantum dots as emission layer,” pp. 3–4, 8 February 2022.
- [19] S. J. Leon, Åke Björck, and W. Gander, “Gram-schmidt orthogonalization: 100 years and more,” *Numerical linear algebra with applications*, vol. 20, pp. 492–532, May 2013.
- [20] M. España, “Integración numérica trapezoidal - matlab trapz.” Última consulta: 14-3-2023.
- [21] C. S. Shibi and R. Gayathri, “Hyperspectral target detection using gram-schmidt orthogonal projections,”
- [22] C. Majarrez-Domínguez, D. C. P. Álvarez, D. C. E. V.-S. Verín, and D. A. P. Álvarez, “Separabilidad espectral de cultivos agrícolas con imágenes landsat tm,” *TECNO-CIENCIA Chihuahua*, vol. 1, pp. 48–56, Nov 13, 2018.
- [23] J. Echanobe, I. del Campo, V. Martínez, and K. Basterretxea, “Hsi-drive,” 12/ 2022.
- [24] S. Li, H. Wu, D. Wan, and J. Zhu, “An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine — elsevier enhanced reader,” *ELSEVIER*, 16 July 2010.
- [25] J. Echanobe, I. del Campo, V. Martínez, and K. Basterretxea, “Genetic algorithm-based optimization of elm for on-line hyperspectral image classification,” pp. 4202–4207, 2017-05.