Simposio CEA de Robótica, Bioingeniería, Visión Artificial y Automática Marina 2026

RBVM
Simposio CEA de Robótica, Bioingeniería, Visión por Computador y Automática Marina

Fusión temprana de imágenes multiespectrales mediante técnicas de reproyección para tareas de detección mediante algoritmos profundos: Un estudio comparativo

Heredia-Aguado, Enrique[a,*], Flores, María[a], Cabrera, Juan Jose[a], Ballesta, Mónica[a], Valiente, David[a], Paya, Luis[a], Gil, Arturo[a]

[a] *Instituto de Investigación en Ingeniería de Elche (I3E). Universidad Miguel Hernández de Elche. Avda. de la Universidad s/n, 03202 Elche (Alicante), España.*

Resumen

Frente a las limitaciones que suponen las condiciones ambientales y de iluminación para el procesamiento de imágenes del espectro visible, la fusión con imágenes de otros espectros constituye una estrategia prometedora para la mejora de la percepción. Este artículo presenta un estudio comparativo de métodos de fusión por reproyección adaptativa que emplean transformaciones en el dominio de la frecuencia o de la varianza para combinar datos de imágenes RGB y térmicas. Se evalúan métodos basados en Análisis de Componentes Principales (PCA), Análisis Factorial (FA), y fusión basada en Wavelets y Curvelets, todos ellos integrados en un flujo de detección con YOLOv8. Los experimentos se realizan sobre el *dataset* KAIST, con especial atención al rigor metodológico y la reproducibilidad. Los resultados se comparan con métodos previos, mostrando sus fortalezas pero también sus limitaciones y desventajas frente a métodos más clásicos. Finalmente, se discuten las implicaciones para futuras investigaciones y el valor de un diseño experimental robusto como punto indispensable en el avance del estado del arte de la fusión de imágenes multiespectrales.

Palabras clave: fusión de imágenes, imagen térmica, imagen multiespectral, detección en imagen, aprendizaje profundo

Multiespectral image early fusion through reprojection techniques for deep learning based detection tasks: A comparative study

Abstract

Given the limitations that environmental and lighting conditions impose on visible-spectrum image processing, fusion with images from other spectral bands constitutes a promising strategy for improving perception. This paper presents a comparative study of adaptive reprojection fusion methods that employ frequency-domain or variance-based transformations to combine RGB and thermal image data. Methods based on Principal Component Analysis (PCA), Factor Analysis (FA), and Wavelet- and Curvelet-based fusion are evaluated, all integrated into a YOLOv8 detection pipeline. Experiments are conducted on the KAIST dataset with particular emphasis on methodological rigor and reproducibility. Results are compared with previous methods, highlighting their strengths but also their limitations and drawbacks relative to simpler approaches. Finally, implications for future research and the value of robust experimental design as an essential foundation for advancing the state of the art in multispectral image fusion are discussed.

Keywords: image fusion, thermal images, multiespectral image, image detection, deep learning

1. Introducción

La fusión de datos multimodales ha demostrado consistentemente su valor en diversas áreas del conocimiento. Combinando diferentes fuentes de información es posible compensar las limitaciones de cada modalidad individual y aprovechar sus fortalezas complementarias.

El escenario de aplicación principal de esta investigación son las tareas de detección autónoma en escenarios como operaciones de búsqueda y rescate (SAR, *Search And Rescue*), vigilancia y seguridad, contextos donde una percepción robusta bajo condiciones no controlables resulta crítica. La solución propuesta está diseñada para su despliegue a bordo de plataformas robóticas autónomas, requiriendo capacidades de procesamiento en tiempo real bajo restricciones impuestas por el hardware empleado. Aunque no es obligatoria una inferencia de alta frecuencia, alcanzar un rendimiento fiable en torno a 1 Hz resulta suficiente para soportar operaciones efectivas.

Aunque las imágenes del espectro visible (RGB) propor-

*Autor para correspondencia: e.heredia@umh.es

cionan información rica de textura y color, su rendimiento se degrada significativamente bajo condiciones de poca iluminación o en escenarios desestructurados. Las imágenes térmicas infrarrojas (espectro infrarrojo lejano) son menos sensibles a estos cambios y proporcionan información complementaria sobre temperatura y emisividad de los cuerpos de la escena. No obstante, son sensibles a cambios ambientales como la temperatura ambiente y las variaciones estacionales. Tal como se describe en Heredia-Aguado et al. (2025), existen casos límite —como oclusiones presentes solo en una modalidad— que justifican la integración de ambas fuentes de datos. La fusión multiespectral tiene además aplicaciones potenciales en seguridad vial, conducción autónoma o agricultura.

En la literatura existen diversas estrategias de fusión. Esta investigación se construye sobre técnicas de fusión temprana estática (Heredia-Aguado et al., 2025) para establecer una línea base robusta con la que comparar los beneficios de diversos enfoques. Específicamente, los métodos de fusión explorados en este trabajo reducen una imagen de cuatro canales (RGB+T) a una representación de tres canales, permitiendo el uso de detectores de imagen reconocidos como YOLOv8. Este artículo extiende el estudio de técnicas de fusión temprana estática mediante:

- La propuesta y evaluación de métodos de fusión dinámica basados en proyecciones (PCA, FA) y transformadas en el dominio de la frecuencia (Wavelets, Curvelets).
- Un control experimental robusto: usando el mismo *dataset*, arquitectura de red, inicialización y parámetros de entrenamiento en todos los métodos, mismas condiciones de evaluación que en los casos de fusión estática previos.

El apartado 2 detalla la metodología empleada, incluyendo el algoritmo de detección, el *dataset*, la configuración experimental y las métricas de evaluación que se valoran. El apartado 3 explica cada método de fusión. El apartado 4 presenta y discute los resultados de detección para cada enfoque. Finalmente, el apartado 5 resume los principales hallazgos, limitaciones y posibles direcciones futuras.

2. Metodología

Los algoritmos de fusión que se plantean son parte del enfoque conocido como fusión temprana: la información se fusiona antes de alimentar algoritmos de aprendizaje profundo para la tarea de detección. Tal como se muestra en la Figura 1, este enfoque permite aprovechar arquitecturas ya consolidadas sin modificación. A continuación se detallan los bloques de la Figura 1 junto con la configuración experimental y las métricas de evaluación.

Con el fin de facilitar la reproducibilidad, el código fuente está disponible de forma abierta (`https://github.com/enheragu/yolo_test_utils`, consultado el 20 de febrero de 2026), permitiendo la replicación de los experimentos.

2.1. Algoritmo de Detección

Todos los algoritmos de fusión han sido probados bajo la misma red de detección, YOLOv8 Jocher et al. (2023), que funciona como descriptor común del rendimiento de cada método. Se trata de una arquitectura de detección de una sola etapa que ofrece un buen equilibrio entre velocidad, consumo de memoria y precisión (Jiang et al., 2022; Diwan et al., 2023), resultando adecuada para el despliegue en plataformas robóticas con restricciones de hardware. Se mantiene la versión v8 para asegurar la comparabilidad con los resultados de fusión estática presentados en Heredia-Aguado et al. (2025), donde se detalla la justificación de esta elección frente a alternativas como Faster-RCNN (Ren et al., 2017) o detectores basados en transformers (Carion et al., 2020).

2.2. El Dataset KAIST

El *dataset* KAIST (Hwang et al., 2015) comprende pares de imágenes térmicas (infrarrojo de onda larga) y visibles. Específicamente, el *Multispectral Pedestrian Dataset* contiene aproximadamente 95 000 pares color-térmico (640×480 píxeles, 20 Hz) capturados desde un vehículo en movimiento. Todos los pares de imágenes están anotados con categorías de objeto (persona, grupo de personas, ciclista), proporcionando un total de 103 128 anotaciones densas y 1 182 instancias únicas de peatones adecuadas para tareas de detección. A diferencia de otros *datasets* que solo incluyen imágenes nocturnas, KAIST abarca tanto condiciones diurnas como nocturnas, lo que permite un análisis más completo del rendimiento de los métodos de fusión bajo diferentes escenarios de iluminación. En este trabajo nos centraremos en el caso diurno para poner a prueba las conclusiones extraidas del caso nocturno con LLVIP (Heredia-Aguado et al., 2025) bajo diferentes condiciones de iluminación.

Sin embargo, los pares de imágenes del *dataset* KAIST no están perfectamente alineados píxel a píxel, debido a la desincronización entre las dos cámaras. Aunque un pequeño desfase temporal puede parecer de menor impacto, esta diferencia se vuelve significativa con objetos en movimiento, especialmente considerando que las imágenes fueron capturadas desde un vehículo en marcha. Para los experimentos de este estudio se empleó la corrección de alineamiento propuesta en Heredia-Aguado et al. (2025), basada en flujo óptico entre imágenes visibles consecutivas. Dicha corrección mejoró los resultados y alineó las conclusiones con LLVIP; aunque no es perfecta, resulta suficiente para usar KAIST como fuente complementaria de información bajo condiciones distintas.

El propio *dataset* incluye una propuesta de división en subconjuntos de entrenamiento y test. En este estudio se ha optado por un división 80-20, como se detalla en la Tabla 1.

Tabla 1: Distribución del *dataset* KAIST: subconjuntos originales y division 80-20 empleado.

Subconjunto	**Imágenes**	**Fondos**	**Instancias**
Original KAIST			
test-day-01	29 178	15 191	34 492
train-day-02	16 694	10 803	12 521
Remuestreo 80-20			
Test day	12 515	7 043	11 667
Train day	50 062	28 942	57 052

Algunos ejemplos pueden verse en la Figura 2. Todas las imágenes fueron tomadas desde un vehículo, lo que implica personas a distancias variables con tamaños diversos. La cámara LWIR (FLIR-A35) no fue calibrada para medir temperaturas

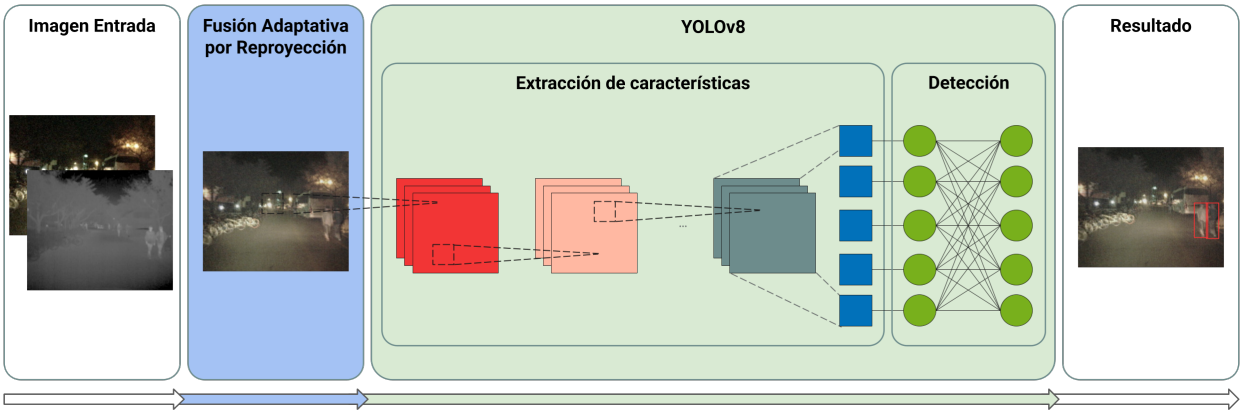

Figura 1: Diseño arquitectónico para la evaluación de algoritmos de fusión.

específicas sino que se empleó un tiempo de exposición constante, manteniendo una correspondencia consistente entre temperatura y nivel de gris a lo largo del *dataset*.

2.3. *Configuración Experimental*

Como ya se ha indicado, la idea de esta investigación es establecer un marco de comparación entre diferentes métodos de fusión. Dado que algunos de ellos (fusión intermedia, fusión tardía) implican cambios en la arquitectura del modelo profundo, no se utilizarán técnicas de *transfer learning*, sino que todos los modelos se entrenan desde cero, haciendo uso de las imágenes fusionadas resultantes, y basándose en los mismos pesos de inicialización e hiperparámetros idénticos, incluyendo estrategias de aumento de datos y esquemas de inicialización. Esto garantiza una comparación robusta y justa entre los diferentes algoritmos de fusión.

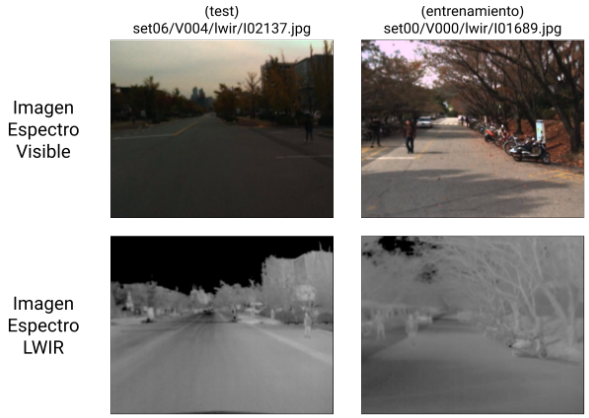

Figura 2: Ejemplos de pares de imágenes visible-LWIR del *dataset* KAIST en condición diurna.

Todas las pruebas de entrenamiento y validación se realizaron utilizando el mismo hardware: una GPU NVIDIA, modelo GeForce RTX 4090 con 24 GB junto con un procesador Intel(R) Core(TM) i7-11700 de 11ª generación (2.50GHz).

2.4. *Métricas de Evaluación y Detalles de Implementación*

Para tener una comprensión más clara del rendimiento y enfocándose en el caso de uso ya presentado, el rendimiento de cada método se comparará basándose en las métricas de *precisión* y *recall*. La *precisión* proporciona una medida de la capacidad del modelo entrenado para evitar falsos positivos, mientras que el *recall* informa sobre la capacidad del modelo para detectar todas las instancias sin dejar ninguna sin detectar.

El análisis también incluye la métrica estándar *mean Average Precision* (mAP) en umbrales de IoU (*Intersection over Union*) estándar. También se reporta el tiempo de fusión para cada método, ya que la eficiencia computacional puede ser crítica para aplicaciones en tiempo real.

3. Fusión de Imagen

Los métodos de fusión que se presentan cubren diferentes alternativas de reproyectar la información de cuatro canales en una salida de tres canales. Los dos primeros métodos se basan en reproyectar los datos de imagen en un marco diferente basado en la variación de los datos. Los siguientes métodos incluyen fusión en el dominio de la frecuencia antes de que los datos se reproyecten de vuelta al formato de imagen de tres canales.

Este apartado cubre PCA y FA como métodos de reducción de dimensionalidad aplicados a la fusión de imágenes; y la transformada Wavelet y Curvelet para fusión de imágenes. Las imágenes resultantes son alimentadas a la red de detección profunda, YOLOv8, para entrenar y validar el modelo resultante.

3.1. *Fusión basada en la varianza de los datos*

3.1.1. *Análisis de Componentes Principales (PCA)*

El Análisis de Componentes Principales involucra una herramienta matemática que transforma un número dado de variables correlacionadas en un número de variables no correlacionadas. Con este enfoque, y partiendo de datos de cuatro canales, a través de PCA se calculan las direcciones de máxima varianza. Tomando los componentes más relevantes (tres componentes para una salida de tres canales en este caso), la imagen se reproyecta al espacio de imagen. Este método ha sido propuesto con diferentes variantes y aplicaciones en el campo del procesamiento de imágenes (Kumar and Muttan, 2006; Elmasry et al., 2020).

Para esta investigación se sigue el enfoque genérico: para cada imagen los datos se reproyectan basándose en los tres componentes más relevantes de esa misma imagen y luego se alimentan al algoritmo de aprendizaje profundo.

3.1.2. *Análisis Factorial (FA)*

Siguiendo un enfoque similar a la herramienta PCA, el Análisis Factorial es otra herramienta para reducción de dimensionalidad basada en la varianza compartida de los datos (Joliffe and Morgan, 1992). Con esta herramienta se calculan un conjunto de factores (con cierta similitud a los componentes en PCA), de manera que las variables de entrada se asumen como combinaciones lineales de estos factores más, para cada variable, un término de error. La ventaja clave del método es que permite la separación de la varianza común en los datos de la varianza atribuible al error. De esta manera, la reproyección se hace solo a través de los factores calculados basándose en la varianza común. Aunque no es un método comúnmente usado en procesamiento de imágenes, creemos que aporta un enfoque interesante al problema ya que el ruido o incluso los valores atípicos no son algo desconocido en el procesamiento de imágenes.

3.2. *Fusión en el Dominio de la Frecuencia*

3.2.1. *Fusión por Transformada Wavelet*

La Transformada Wavelet Discreta (DWT) (Sifuzzaman et al., 2009) es una técnica derivada basada en la Transformada de Fourier. La transformada de Fourier analiza una señal dada basándose en sus componentes de frecuencia, pero al hacerlo pierde información espacial sobre los datos. La DWT bidimensional asegura mantener la información espacial (crítica al analizar una imagen) mientras se centra en el análisis de frecuencia (Zhang, 2019). Con este enfoque una imagen dada puede descomponerse en componentes de frecuencia.

Para cada canal de la imagen de entrada de cuatro canales, la DWT proporciona un conjunto de componentes de frecuencia. Estos componentes se dividen en dos sub-bandas: coeficientes de aproximación (cA) y sub-bandas de coeficientes de detalle. Estos coeficientes son los que se mezclan entre imágenes: el coeficiente de aproximación RGB se mezcla con el coeficiente de aproximación térmico; lo mismo aplica a los coeficientes de detalle. Una vez fusionados, se aplica una transformada inversa para reconstruir una imagen de tres canales.

Las sub-bandas de detalle, el componente de alta frecuencia de la imagen, capturan principalmente cambios locales, texturas e información de bordes, mientras que las sub-bandas de aproximación, los componentes de baja frecuencia, contienen la mayoría de la estructura general e información espectral de la imagen. Hay diferentes enfoques sobre cómo estos componentes deberían combinarse:

- Valor máximo: Entre dos componentes dados se mantiene el valor máximo descartando el otro. Este enfoque asegura mantener información de textura y bordes, pero puede incluir mayor ruido en la imagen resultante.
- Valor promedio: Aunque promediar ambos componentes puede difuminar cambios locales y bordes, mantiene una imagen más suave y limpia.

Ambas versiones se han implementado en dos fusiones Wavelet: promediada y máximo valor. Para la primera versión los componentes de cada canal RGB se promedian con los de la imagen térmica, tanto para sub-bandas de aproximación como de detalle. El segundo enfoque combina los componentes de detalle de cada canal RGB con los componentes térmicos manteniendo el valor máximo. En este caso el coeficiente de aproximación se combina siguiendo una Fusión por Mezcla α (Ofir, 2023): $C_{Aprox} = \alpha \cdot C_{A-RGB} + (1 - \alpha) \cdot C_{A-TH}$ siendo alpha un coeficiente relativo basado en el valor del píxel térmico. El método de valor máximo tiende a preservar más información que el método de promedio en fusión de imágenes, ya que selecciona el píxel de mayor intensidad de las imágenes de entrada, asegurando que no se pierdan detalles significativos de ninguna fuente, mientras que el promediado puede diluir o difuminar características importantes (Patil et al., 2013; Sahu and Sahu, 2014). Otras técnicas que muestran resultados interesantes se basan en maximizar el contraste (Indira, 2015) pero no se han incluido en este estudio ya que potencialmente aumentarían el consumo de tiempo de fusión.

3.2.2. *Fusión por Transformada Curvelet*

El problema con la transformación Wavelet es que se centra en singularidades puntuales, ignorando algunas de las propiedades geométricas de las estructuras en la imagen. Además, no aprovecha la regularidad de los bordes (Ma and Plonka, 2010). La transformada Curvelet define una transformada diferente de la DWT, ya que realiza un análisis multi-escala y multi-direccional que es particularmente efectivo al representar y comprimir estructuras de bordes y curvas. No es sorprendente que sea ampliamente popular para soluciones de procesamiento de imágenes (Starck et al., 2002).

Como en el enfoque de fusión DWT, se han implementado dos versiones del algoritmo. En ambas, los coeficientes de la primera capa se fusionan promediando tanto los coeficientes RGB como térmicos. En los últimos niveles (hasta cuatro niveles en total) se promedian o se filtran por máximo.

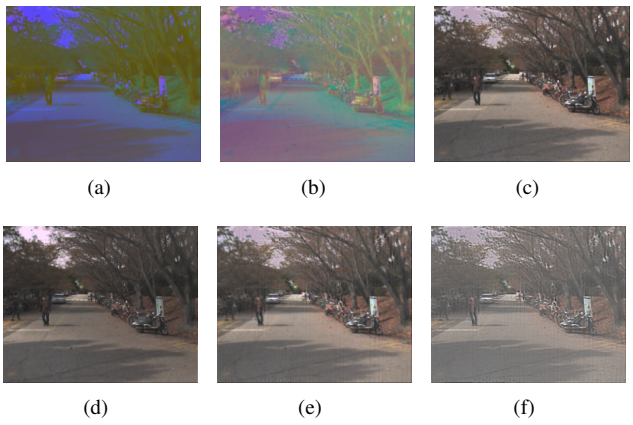

Figura 3: Ejemplos en representación de falso color del resultado de cada método de fusión basado en una imagen del *dataset* KAIST. (**a**) Fusión PCA. (**b**) Fusión FA. (**c**) Fusión Wavelet (promedio). (**d**) Fusión Wavelet (max-value). (**e**) Fusión Curvelet (promedio). (**f**) Fusión Curvelet (max-value).

En la Figura 3 se presenta un ejemplo de cada fusión para una imagen dada. Cabe destacar que las opciones PCA y FA suponen una reconstrucción de la imagen con canales que no son coherentes con los canales RGB en los que se representa la imagen. Como puede observarse, la fusión FA proporciona una solución más limpia respecto a la fusión PCA, aislando mejor la instancia y la información relevante de la imagen. En cuanto

a la fusión Curvelet la versión max-value tiende a proporcionar una imagen con mayor contraste y definición, pero a costa de mayor ruido y pérdida de información. El caso de la transformada Wavelet es algo más sutil, aun así puede verse como la versión de valor máximo ofrece mejores contrastes mientras que la promediada da mayor uniformidad.

4. Resultados

Los resultados se centran en la condición diurna del *dataset* KAIST, comparándolos con los obtenidos en condición nocturna en (Heredia-Aguado et al., 2025). La Tabla 2 resume los tiempos de fusión: solo PCA y Wavelet (<1 s) cumplirían el requisito de 1 Hz, mientras que FA y Curvelet resultan significativamente más lentos. En Wavelet y Curvelet ambos enfoques (máximo y promediado) comparten la transformación, computándose conjuntamente.

Tabla 2: Resumen del tiempo medio de computación para realizar la fusión de cada uno de los métodos basados en las imágenes del *dataset* KAIST.

Método de Fusión	**Media (s)**	**Std (s)**
Fusión PCA	0.454	0.12
Fusión FA	18.135	2.737
Fusión Wavelet (ambas)	0.304	0.043
Fusión Curvelet (ambas)	22.59	0.405

4.1. *Condición Diurna*

La Figura 4 y la Tabla 3 resumen las métricas de detección en condición diurna, incluyendo las líneas base de modalidad única y dos fusiones estáticas de referencia.

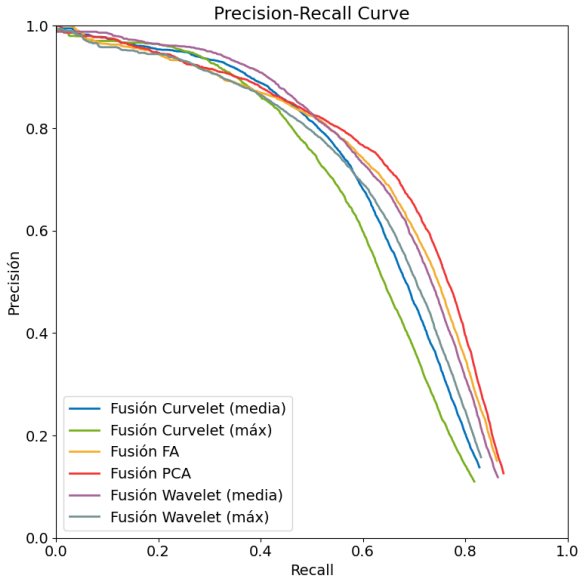

Figura 4: Curva de *precisión-recall* para todos los métodos de fusión en condición diurna.

Como puede observarse, la fusión PCA obtiene los mejores resultados globales en las métricas mAP50 (0.6961) y mAP50-95 (0.2855), superando tanto las líneas base de modalidad única (Visible, LWIR) como las fusiónes estáticas RGBT y VTHS (Heredia-Aguado et al., 2025). Aunque por separado la *precisión* y el *recall* son mejores en técnicas estáticas, podemos ver que la técnica de PCA ofrece mejor equilibrio entre ambas. De las fusiónes por reproyección la fusión Curvelet (promedio) alcanza la mayor *precisión* (0.7359), pero a costa de un *recall* significativamente más bajo, lo que limita su utilidad en operaciones SAR. En general, los métodos basados en proyección (PCA, FA) ofrecen un mejor equilibrio entre *precisión* y *recall* que los métodos en el dominio de la frecuencia en condición diurna. La figura respalda estas conclusiones, mostrando como los métodos PCA y FA pierden cierta *precisión* más rápido conforme aumentamos el recall, pero mantienen un rendimiento más consistente a valores más altos de recall.

4.2. *Discusión*

A pesar de la sofisticación de los métodos de fusión dinámica, las mejoras respecto a métodos de fusión previos y más simples son limitadas en KAIST, aunque la corrección de alineamiento aplicada ayuda a reducir parte de las discrepancias y hace más comparable este *dataset* con LLVIP. Comparando los resultados diurnos (Tabla 3) con los nocturnos reportados en (Heredia-Aguado et al., 2025), se observa que todos los métodos rinden mejor de noche, donde la información térmica es dominante. Aun así, esa mejoría no siempre compensa la ofrecida por métodos más simples.

Puede observarse como la fusión basada en promedios supera la fusión basada en máximos, tanto para el caso de Wavelets como de Curvelets, resultado coherente con lo observado en condición nocturna en (Heredia-Aguado et al., 2025). Esta situación sugiere que el promedio puede preservar mejor algunas características sutiles que son relevantes en este caso de uso. Contrariamente a parte de la bibliografía ya presentada, la fusión basada en promedio demostró ser más equilibrada para tareas de detección basadas en fusión de imagen RGB-Térmica.

Aunque tanto FA como Curvelet ofrecen enfoques teóricamente atractivos, su aplicación práctica reveló limitaciones significativas en velocidad computacional (Tabla 2), siendo considerablemente más lentos que PCA o Wavelet. Esto los hace menos adecuados para despliegue en tiempo real a bordo de sistemas robóticos, un compromiso que debe considerarse al seleccionar algoritmos de fusión para aplicaciones sensibles al tiempo.

Un aspecto relevante de este trabajo, junto con (Heredia-Aguado et al., 2025), es la progresiva construcción de un marco comparativo robusto. Al disponer de evaluaciones de fusiones estáticas y por reproyección sobre el mismo *dataset*, con YOLOv8 como herramienta de medida común, es posible extraer relaciones de rendimiento relativo entre métodos dentro de este marco experimental. Estas tendencias deben interpretarse como una guía y validarse para cada detector. En la práctica, replicar un único método —por ejemplo PCA— puede servir para estimar el comportamiento del resto sin repetir toda la batería de experimentos, como se observó de forma razonable entre LLVIP y KAIST.

5. Conclusiones

Presentamos en este trabajo un estudio comparativo robusto de métodos de fusión dinámica para imágenes RGB y térmicas, evaluados con YOLOv8 en el *dataset* KAIST. Los métodos dinámicos no muestran mejoras consistentes frente a fusiones

Tabla 3: Rendimiento de detección en condición diurna. Se incluyen líneas base de modalidad única y también fusión estática (Heredia-Aguado et al., 2025) como referencia. Los mejores resultados por columna están resaltados en negrita.

Método	P	R	mAP50	mAP50-95	Mejor Ép.
Visible	0.699	0.636	0.673	0.257	8
LWIR	0.641	0.495	0.518	0.198	3
Fusión RGBT (no ec.)	0.716	**0.652**	0.616	0.265	4
Fusión VTHS (no ec.)	**0.753**	0.627	0.689	0.283	12
Fusión PCA	0.7239	0.6482	**0.6961**	**0.2855**	14
Fusión FA	0.7122	0.6306	0.6818	0.2846	25
Fusión Wavelet (media)	0.7110	0.6180	0.6842	0.2679	8
Fusión Wavelet (máx)	0.6991	0.5938	0.6513	0.2563	13
Fusión Curvelet (media)	0.7359	0.5689	0.6496	0.2596	23
Fusión Curvelet (máx)	0.6973	0.5432	0.6145	0.2418	14

más simples, especialmente en precisión y recall. Este resultado refuerza la importancia del rigor metodológico y del reporte transparente de resultados negativos para guiar futuras investigaciones en fusión multiespectral. Además, junto con los resultados de (Heredia-Aguado et al., 2025), este trabajo contribuye a un marco comparativo creciente para orientar la selección de métodos en nuevos escenarios.

El estudio presentado está limitado a imágenes con corrección de alineamiento aplicada. El trabajo futuro debería evaluar estos métodos en *datasets* más desafiantes y con diferentes técnicas de ecualización, que demostraron ser relevantes en fusión estática (Heredia-Aguado et al., 2025). Aunque los métodos propuestos no superaron las líneas base de modalidad única, los resultados sugieren potencial de mejora con la incorporación de estrategias de ecualización y técnicas de fusión más avanzadas.

Agradecimientos

Este trabajo es parte del proyecto CIPROM/2024/8, financiado por la Generalitat Valenciana, Conselleria de Educación, Cultura, Universidades y Empleo (programa PROMETEO). También forma parte del proyecto PID2023-149575OB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER,UE

Referencias

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: Computer Vision – ECCV 2020. Springer International Publishing, pp. 213–229.

Diwan, T., Anirudh, G., Tembhurne, J. V., 2023. Object detection using yolo: challenges, architectural successors, datasets and applications. Multimedia Tools and Applications 82 (6), 9243–9275.
DOI: 10.1007/s11042-022-13644-y

Elmasry, S. A., Awad, W. A., Abd El-hafeez, S. A., 2020. Review of different image fusion techniques: Comparative study. In: Internet of Things—Applications and Future. Springer Singapore, pp. 41–51.

Heredia-Aguado, E., Alfaro-Pérez, M., Flores, M., Paya, L., Valiente, D., Gil, A., 2025. A robust comparative study of adaptative reprojection fusion methods for deep learning based detection tasks with rgb-thermal images. In: Proceedings of the 22nd International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO. INSTICC, SciTePress, pp. 313–320.
DOI: 10.5220/0013761800003982

Heredia-Aguado, E., Cabrera, J. J., Jiménez, L. M., Valiente, D., Gil, A., 2025. Static early fusion techniques for visible and thermal images to enhance convolutional neural network detection: A performance analysis. Remote Sensing 17 (6).
URL: https://www.mdpi.com/2072-4292/17/6/1060
DOI: 10.3390/rs17061060

Hwang, S., Park, J., Kim, N., Choi, Y., Kweon, I. S., 2015. Multispectral pedestrian detection: Benchmark dataset and baseline. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1037–1045.
DOI: 10.1109/CVPR.2015.7298706

Indira, K., 2015. Image fusion for pet ct images using average maximum and average contrast rules. Int J Appl Eng Res 10 (1), 673–80p.

Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022. A review of yolo algorithm developments. Procedia Computer Science 199, 1066–1073.
DOI: https://doi.org/10.1016/j.procs.2022.01.135

Jocher, G., Chaurasia, A., Qiu, J., january 2023. YOLOv8 by Ultralytics.
URL: https://github.com/ultralytics/ultralytics

Joliffe, I., Morgan, B., 1992. Principal component analysis and exploratory factor analysis. Statistical Methods in Medical Research 1 (1), 69–95.
DOI: 10.1177/096228029200100105

Kumar, S. S., Muttan, S., 2006. PCA-based image fusion. In: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII. Vol. 6233. SPIE, p. 62331T.
DOI: 10.1117/12.662373

Ma, J., Plonka, G., 2010. The curvelet transform. IEEE Signal Processing Magazine 27 (2), 118–133.
DOI: 10.1109/MSP.2009.935453

Ofir, N., 2023. Multispectral image fusion based on super pixel segmentation. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5.
DOI: 10.1109/ICASSP49357.2023.10095874

Patil, V., Sale, D., Joshi, M., 2013. Image fusion methods and quality assessment parameters. Asian Journal of Engineering and Applied Technology 2, 40–45.
DOI: 10.51983/ajeat-2013.2.1.643

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6), 1137–1149.
DOI: 10.1109/TPAMI.2016.2577031

Sahu, V., Sahu, D., 2014. Image fusion using wavelet transform: A review. Global Journal of Computer Science and Technology 14 (F5), 21–28.

Sifuzzaman, M., Islam, R., Ali, M., 2009. Application of wavelet transform and its advantages compared to fourier transform. Journal of Physical Science 13, 121–134.

Starck, J.-L., Candes, E., Donoho, D., 2002. The curvelet transform for image denoising. IEEE Transactions on Image Processing 11 (6), 670–684.
DOI: 10.1109/TIP.2002.1014998

Zhang, D., 2019. Wavelet Transform. Springer International Publishing, pp. 35–44.
DOI: 10.1007/978-3-030-17989-2_3