

Correspondencia de características utilizando esquemas clásicos en el espectro visible

Pablo Ricaurte¹, Carmen Chilán¹, Cristhián A. Aguilera-Carrasco²
Boris X. Vintimilla¹ and Angel D. Sappa^{1,2}

¹*CIDIS-FIEC, Escuela Superior Politécnica del Litoral (ESPOL),
Campus Gustavo Galindo, Km 30.5 vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*

²*Computer Vision Center, Universitat Autònoma de Barcelona,
08193 Bellaterra, Barcelona, España*

{paricaur, cchilan, boris.vintimilla}@espol.edu.ec, {caguilera, asappa}@cvc.uab.es

Keywords: Imágenes infrarrojas, Descriptores de puntos característicos.

Abstract: This paper presents a comparative evaluation of classical feature point descriptors when they are used in the long-wave infrared spectral band. Robustness to changes in rotation, scaling, blur, and additive noise are evaluated using a state of the art framework. Statistical results using an outdoor image data set are presented together with a discussion about the differences with respect to the results obtained when images from the visible spectrum are considered.

1. INTRODUCTION

En general, las aplicaciones de visión por computadora se basan en el uso de cámaras que trabajan en el campo visible del espectro. Recientemente avances en la captura de imágenes infrarrojas, así como también la reducción en los costos de esas cámaras, ha abierto nuevas oportunidades para desarrollar soluciones excelentes que trabajan en la banda infrarroja del espectro o entre varios dominios del espectro entre el campo infrarrojo y visible (ejm., (Barrera et al., 2013), (Aguilera et al., 2012), (Barrera et al., 2012) and (Felicísimo and Cuartero, 2006)).

Las bandas espectrales de las imágenes infrarrojas van desde $0.75 \mu m$ a $15 \mu m$, del cual se divide en las siguientes categorías: Infrarrojo cercano (NIR: $0.751.4 \mu m$), Infrarrojo de onda corta (SWIR: $1.43 \mu m$), Infrarrojo de onda media (MWIR: $38 \mu m$) o Infrarrojo de onda larga (LWIR: $815 \mu m$). Imágenes de cada una de estas categorías tienen ventajas particulares para ciertas aplicaciones dadas; por ejemplo las imágenes NIR son generalmente usadas para aplicaciones que usan detección de la mirada y seguimiento de los ojos (Coyle et al., 2004); la banda espectral SWIR ha mostrado su uso en ambientes con mucha neblina (Hansen and Malchow, 2008); MWIR es usa-

do generalmente para detectar temperaturas de alguna forma superiores a la del cuerpo en aplicaciones militares; finalmente, las imágenes LWIR han sido usadas en video vigilancia y asistencia en el manejo (Krotosky and Trivedi, 2007). El presente trabajo se concentra en el campo LWIR, el cual corresponde a la banda espectral más lejana con respecto al espectro visible.

Siguiendo la evolución del espectro visible basado en vision por computadora, en las imágenes del dominio Infrarrojo han surgido temas tales como el registrado de imágenes, reconocimiento de patrones o vision estereó los cuales han sido enteramente abarcados. Como primera instancia, herramientas clásicas del espectro visible han sido solo usadas o poco adaptadas en el nuevo dominio. Una de estas herramientas es la descripción de puntos característicos, los cuales han sido un tema de bastante investigación en la pasada década en la comunidad de vision por computadora. Dado la gran cantidad de contribuidores en este tópico han habido muchos trabajos en la literatura sobre evaluar y comparar su rendimiento en el caso del espectro visible (ejm., (Miksik and Mikolajczyk, 2012), (Mikolajczyk and Schmid, 2005), (Bauer et al., 2007) y (Schmid et al., 2000)).

Similar al caso de espectro visible, el trabajo pre-

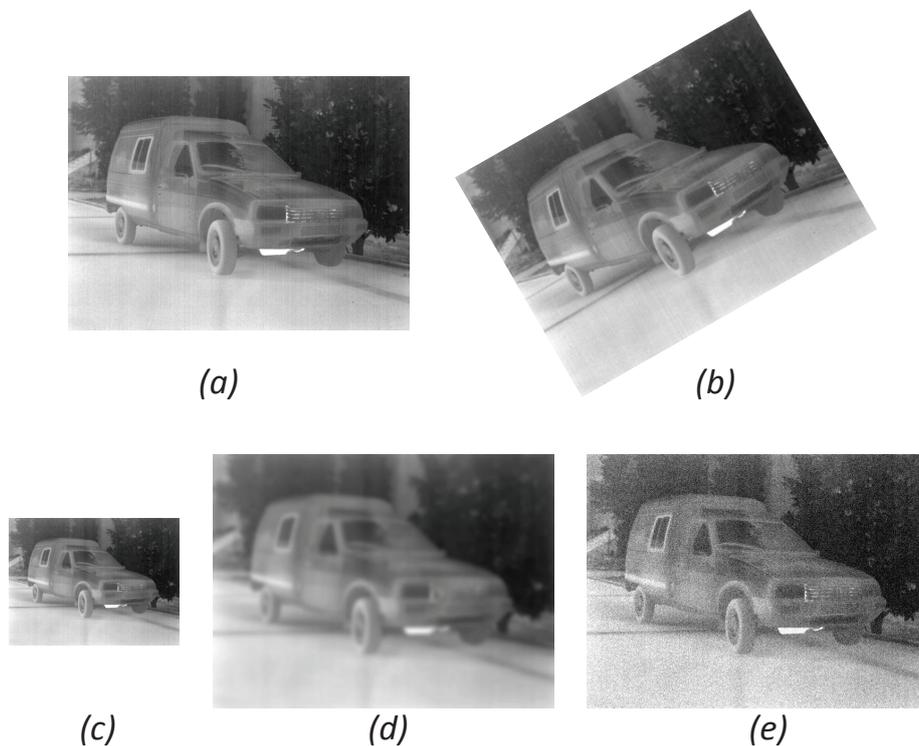


Figura 1: Una imagen LWIR del dataset a evaluar junto con algunas ilustraciones de las transformaciones aplicadas: (a) imagen original; (b) rotación; (c) escala (d) blur; (e) ruido.

sente propone estudiar el rendimiento de los descriptores de puntos característicos cuando son considerados en el dominio Infrarrojo. Como existe una larga cantidad de algoritmos en la literatura, hemos decidido seleccionar los más representativos y recientes. Nuestro estudio incluye: SIFT (Lowe, 1999), SURF (Bay et al., 2006), ORB (Rublee et al., 2011), BRISK (Leutenegger et al., 2011), BRIEF (Calonder et al., 2012) y FREAK (Alahi et al., 2012). Este estudio es motivado por el hecho de que a pesar que las imágenes del campo LWIR su apariencia es parecida a la del espectro visible su naturaleza es diferente, es por ello que consideramos que las conclusiones del espectro visible no pueden ser extendidas directamente al dominio Infrarrojo.

El manuscrito se organiza como sigue. La metodología de evaluación usada para la comparación es presentada en la Sección 2. Luego, el conjunto de datos y los resultados experimentales son detallados en la Sección 3. Finalmente, conclusiones y discusiones son dadas en la Sección 4.

2. ESQUEMA DE EVALUACIÓN

Esta sección resume el esquema usado para evaluar el rendimiento de los diferentes enfoques. Apunta a encontrar el mejor descriptor para correspondencia de puntos característicos cuando transformaciones de imagen comunes son considerados: rotación de la imagen en el plano, cambios del tamaño de la imagen, blur y presencia de ruido en las imágenes. Para poder tomar en consideración todos estos cambios posibles, las imágenes dadas han sido modificadas, luego descriptores diferentes son aplicados. El esquema usado ha sido propuesto por Khvedchenia¹ para evaluar el rendimiento de los descriptores de características en el caso del espectro visible. Los algoritmos han sido evaluados considerando como referencia aquellos puntos en la imagen dada. Un método de fuerza bruta es usado para encontrar las correspondencias, junto con la norma L2 o la distancia de Hamming, como ha sido detallado en la tabla 1. El porcentaje de matches correctos entre la imagen de referencia y la modificada es usado como criterio de evaluación. Las transformaciones aplicadas a la imagen dada son detalladas

¹<http://computer-vision-talks.com/2011/08/feature-descriptor-comparison-report/>

debajo. Figura 1 muestra una ilustración de una imagen LWIR junto con algunas imágenes resultantes después de aplicar las diferentes transformaciones.

- **rotación:** El estudio consiste en evaluar la sensibilidad de las rotaciones de una imagen. Las rotaciones son en el plano de la imagen sobre los 360 grados, una nueva imagen es obtenida cada 10 grados.
- **Escala:** El tamaño de la imagen dada es cambiado y la repetitividad del descriptor es evaluado, la imagen original es escalada de entre 0.2 a 2 veces su tamaño con pasos de 0.1 por prueba. Los pixeles de las imágenes escaladas son obtenidos a través de una interpolación lineal.
- **Blur:** La robustez con respecto a blur es evaluada. Consiste de un filtro Gaussiano aplicado iterativamente sobre una imagen dada. En cada iteración el tamaño del kernel filtro ($K \times K$) usado para realizar el blur en la imagen es actualizado como sigue: $K = 2t + 1$, donde $t = \{1, 2, \dots, 9\}$.
- **Ruido:** Este último estudio consiste en añadir ruido a la imagen original. este proceso es implementado mediante la adición a la imagen original de una imagen personalizada. Los valores de pixel de la imagen personalizada son obtenidos aleatoriamente siguiendo una distribución uniforme con $\mu = 0$ y $\sigma = t$, donde $t = \{0, 10, 20, \dots, 100\}$.

En el esquema original propuesto por Khvedchenia, cambios de iluminación también eran considerados, dado que ese estudio fue orientado para imágenes en el espectro visible. En el presente trabajo, como apunta a estudiar el campo LWIR, cambios en valores de intensidad en la imagen no seguirán el mismo comportamiento a través de toda la imagen (como cambios de iluminación en el espectro visible). Los valores de intensidad en las imágenes LWIR están relacionados con el material de los objetos en la escena. En resumen, un estudio similar a cambios de iluminación no es considerado en el presente trabajo.

3. RESULTADOS EXPERIMENTALES

Un conjunto de 20 imágenes LWIR han sido consideradas con la evaluación presentada arriba. Para cada una de las imágenes LWIR la imagen correspondiente en el campo visible es provisto. Estas imágenes del espectro visible son usadas para comparar los resultados obtenidos en el dominio LWIR. La figura 2 presenta algunas de las imágenes LWIR obtenidas en nuestro conjunto de imágenes de prueba; están

Cuadro 1: Algoritmos evaluados en el estudio.

Alg. descriptor	Norma del matcher
SIFT	norma L2
SURF	norma L2
ORB	Distancia de Hamming
BRISK	Distancia de Hamming
BRIEF (detector SURF)	Distancia de Hamming
FREAK (detector SURF)	Distancia de Hamming

disponibles públicamente para investigación futura a través del sitio web ². Para cada algoritmo y transformación el número de correspondencias correctas por sobre el número total de puntos característicos descritos en la imagen original (el número de correspondencias usadas como referencia) es considerado por la evaluación, similar a (Mikolajczyk and Schmid, 2005):

$$recall = \frac{\#matches\ correctos}{\#correspondencias}. \quad (1)$$

El algoritmo evaluado en el presente trabajo es presentado en la Tabla 1. En el caso de BRIEF y FREAK se usa como detector el algoritmo SURF. En ORB, BRISK, BRIEF y FREAK la distancia de Hamming es usada, en vez de la norma L2, para acelerar el matching. Para cada transformación (Sección 2) un conjunto de imágenes es obtenido; por ejemplo para el caso de rotación 36 imágenes son evaluadas.

La Figura 3 resalta los resultados obtenidos cuando la imagen es rotada 360. Puede ser observado que el algoritmo más robusto es SIFT, tal conclusión se obtiene también en el campo visible. Por otro lado el algoritmo BRISK (usando SURF como detector) es el más sensible a rotaciones; actualmente su rendimiento cae a cero justo después de aplicar ± 25 grados a la imagen dada. Este comportamiento fue también observado en el campo visible. Independiente del algoritmo, un mejor desempeño puede ser apreciado en el caso LWIR. En resumen, el ranking de; mejor al peor es como sigue: SIFT, ORB, FREAK, SURF, BRISK, BRIEF (el mismo ranking puede ser observado en ambos espectros).

En el estudio de escala, a pesar de que resultados similares fueron obtenidos en ambos espectros, el desempeño del algoritmo fue mejor en el caso Infrarrojo, en particular para BRISK, el cual fue el peor algoritmo en ambos casos tuvo un mejor desempeño en el caso LWIR. El ranking de el rendimiento de los algoritmos es como sigue (del mejor al peor): SIFT, FREAK, SURF, ORB, BRIEF y BRISK. La Figura 4 muestra esos resultados.

²<http://www.cvc.uab.es/adass/projects/simeve/>

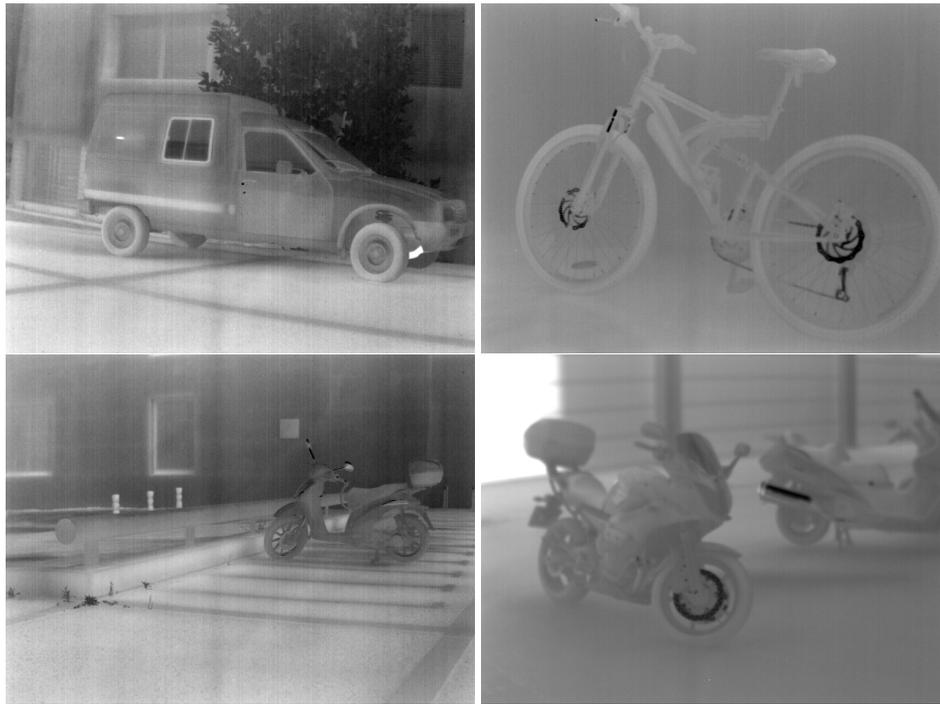


Figura 2: Ilustración de imágenes infrarrojas consideradas en el framework de evaluación.

La Figura 5 presenta el estudio de robustez de los diferentes algoritmos cuando las imágenes dadas son degradadas usando un filtro Gaussiano de tamaño incremental. Similar a el caso previo todos los algoritmos tuvieron un mejor rendimiento en el LWIR que en el espectro visible. En este caso el algoritmo BRIEF es el más robusto, el resto se ubica como sigue: FREAK, SURF, ORB, SIFT y BRISK; siendo BRISK el algoritmo menos robusto al ruido.

Finalmente, la Figura 6 muestra las curvas obtenidas cuando ruido aditivo es considerado. En este caso, diferentemente que en estudios previos, el algoritmo tiene un mejor rendimiento en el espectro visible. Adicionalmente, el rendimiento de SIFT en el espectro LWIR no es tan malo como en el espectro visible. El ranking del rendimiento de los algoritmos es como sigue (del mejor al peor): BRIEF, FREAK, ORB, SIFT, SURF y BRISK.

4. CONCLUSIONES

Este trabajo presenta una evaluación empírica del rendimiento de los descriptores del estado del arte cuando son usados en el dominio LWIR. El principal objetivo fue estudiar qué conclusiones obtenidas en el espectro visible son también válidas para el espectro LWIR. A pesar de que se obtuvieron resultados

similares a los obtenidos en el espectro visible puede ser apreciado que el rendimiento del algoritmo BRIEF (usando SURF como detector) es mejor en el espectro LWIR cuando es comparado con su rendimiento en el espectro visible. El ranking relativo entre algoritmos se mantiene igual para ambas bandas del espectro.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente apoyado por el Gobierno de España bajo el proyecto de investigación TIN2011-25606 y proyecto PROMETEO de la "Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación de la República del Ecuador". Cristhián A. Aguilera-Carrasco apoyo por permiso de "Universitat Autònoma de Barcelona". A los autores les gustaría agradecer a Mr. Khvedchenia Ievgen por proveerles con el esquema de evaluación.

REFERENCES

- Aguilera, C., Barrera, F., Lumbreras, F., Sappa, A., and Toledo, R. (2012). Multispectral image feature points. *Sensors*, 12(9):12661–12672.
- Alahi, A., Ortiz, R., and Vandergheynst, P. (2012). FREAK: Fast retina keypoint. In *IEEE Conference on Com-*

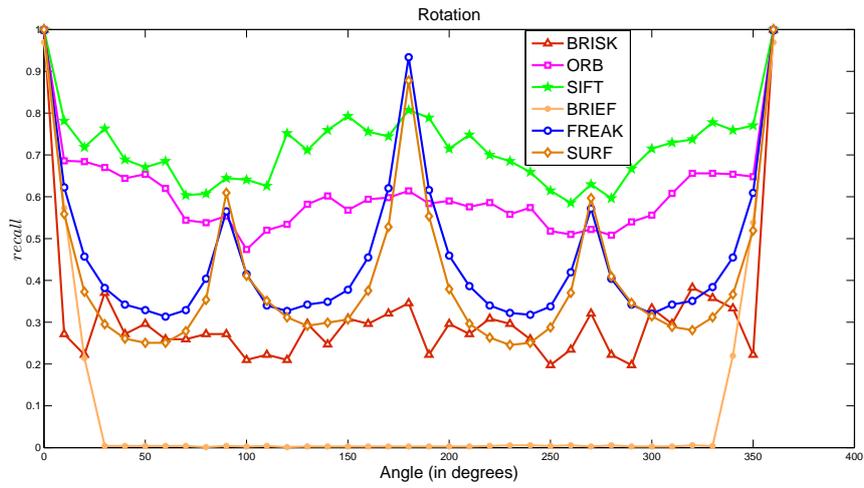


Figura 3: Caso de estudio rotación: resultados promedio obtenidos del conjunto de datos de prueba.

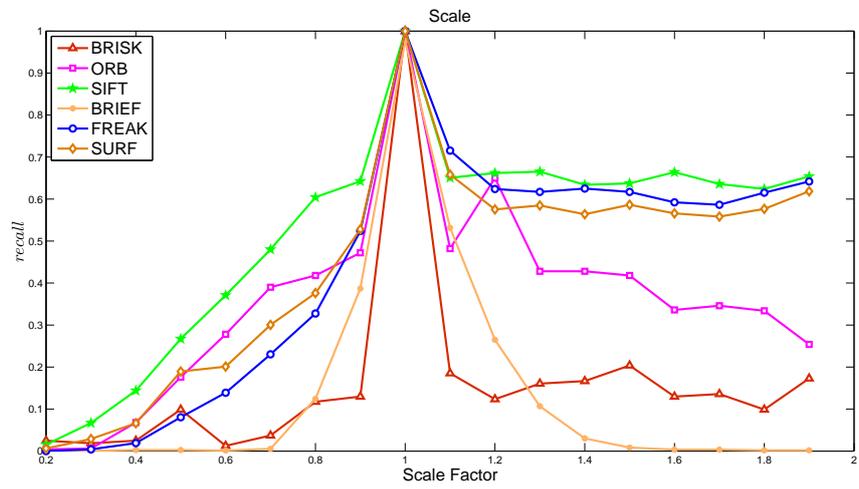


Figura 4: Caso de estudio Escala: resultados promedio obtenidos del conjunto de datos de prueba.

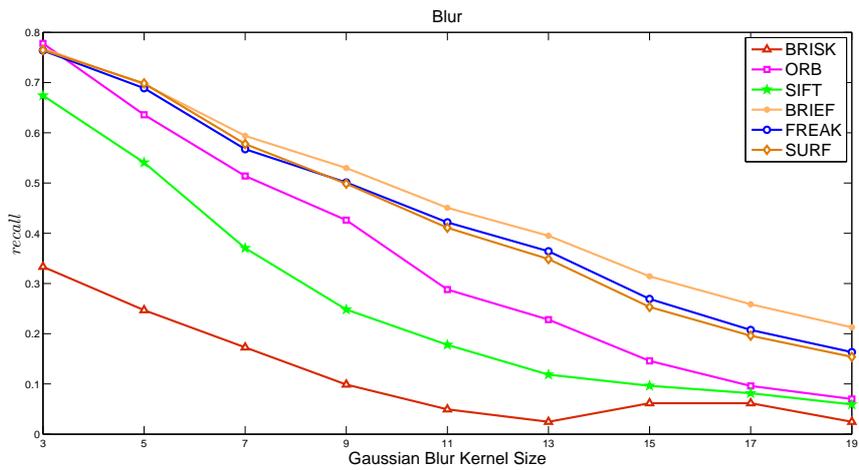


Figura 5: Caso de estudio Blur: resultados promedio obtenidos de evaluar los datos de prueba.

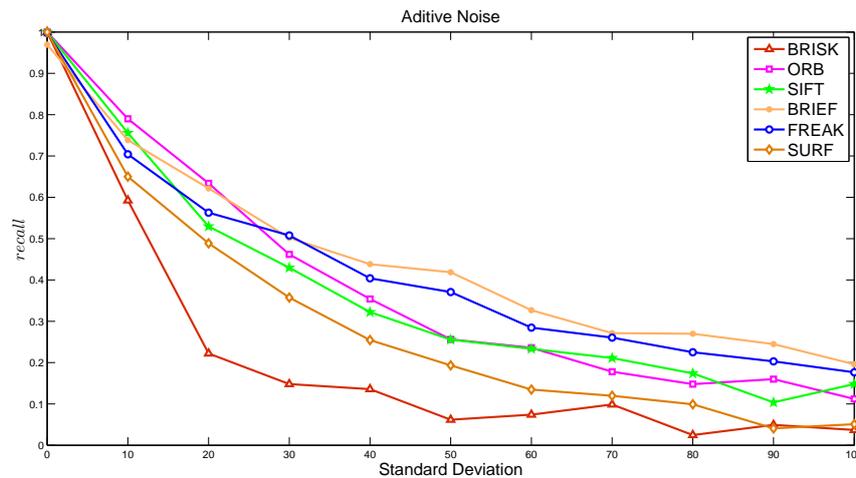


Figura 6: Caso de estudio ruido: resultados promedios obtenidos de evaluar datos de pruebas.

puter Vision and Pattern Recognition, Providence, RI, USA, June 16-21, pages 510–517.

- Barrera, F., Lumbreras, F., and Sappa, A. (2012). Multi-modal stereo vision system: 3d data extraction and algorithm evaluation. *IEEE Journal of Selected Topics in Signal Processing*, 6(5):437–446.
- Barrera, F., Lumbreras, F., and Sappa, A. (2013). Multispectral piecewise planar stereo using manhattan-world assumption. *Pattern Recognition Letters*, 34(1):52–61.
- Bauer, J., Snderhauf, N., and Protzel, P. (2007). Comparing several implementations of two recently published feature detectors. In *Proceedings of the International Conference on Intelligent and Autonomous Systems, Toulouse, France*.
- Bay, H., Tuytelaars, T., and Gool, L. J. V. (2006). SURF: Speeded Up Robust Features. In *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, May 7-13, pages 404–417*.
- Calonder, M., Lepetit, V., Özuysal, M., Trzcinski, T., Strecha, C., and Fua, P. (2012). BRIEF: Computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(7):1281–1298.
- Coyle, S., Ward, T., Markham, C., and McDarby, G. (2004). On the suitability of near-infrared (NIR) systems for next-generation braincomputer interfaces. *Physiological Measurement*, 25(4).
- Felicísimo, A. and Cuartero, A. (2006). Methodological proposal for multispectral stereo matching. *IEEE Trans. on Geoscience and Remote Sensing*, 44(9):2534–2538.
- Hansen, M. P. and Malchow, D. S. (2008). Overview of swir detectors, cameras, and applications. In *Proceedings of the SPIE 6939, Thermosense, Orlando, FL, USA, March 16*.
- Krotosky, S. and Trivedi, M. (2007). On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 8:619–629.
- Leutenegger, S., Chli, M., and Siegwart, R. (2011). BRISK:

Binary Robust Invariant Scalable Keypoints. pages 2548–2555.

- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the IEEE International Conference on Computer Vision, Kerkyra, Greece, September 20-27, pages 1150–1157*.
- Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630.
- Miksik, O. and Mikolajczyk, K. (2012). Evaluation of local detectors and descriptors for fast feature matching. In *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba, Japan, November 11-15, pages 2681–2684*.
- Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. R. (2011). ORB: An efficient alternative to SIFT or SURF. In *IEEE International Conference on Computer Vision, Barcelona, Spain, November 6-13, pages 2564–2571*.
- Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.