Performance evaluation of FIR sensor systems applied to pedestrian detection

S. Franz, R. Schweiger, O. Loehlein, D. Willersinn, K. Kroschel

Daimler AG, Ulm, Germany; Fraunhofer IOSB, Karlsruhe, Germany

ABSTRACT

Besides resolution, an important performance parameter of a FIR camera is the sensitivity. It depends on the sensitivity of the detector array itself and the characteristics of the optic. The effects of the optic are considerably driven by the f-number, with high values resulting in decreased sensitivity, but providing the possibility for simple lens design and cheaper production costs. In this contribution 4 different sensor setups with different optics are evaluated for their impact on the performance of trained pedestrian classifiers.

To overcome the expensive and time consuming process of ground truth generation for multiple sensors, an approach for reusing available high sensitivity reference data is presented. Classifiers are trained on specially transformed reference data with characteristics of sensors with degraded sensitivity.

For the evaluation of the classifiers, data of real world road scenarios is collected simultaneously with the target sensors mounted in parallel in a test vehicle, following a detailed script for recording a pedestrian scene test catalogue. This allows for a direct analysis and comparison of the different sensors and their impact on the detection performance.

Keywords: performance evaluation, sensor performance, pedestrian detection

1. INTRODUCTION

The recurring question for engineers during the development of a detection system is, to which extent a sensor can fall off in quality without affecting the quality of the system itself. The two main parameters of far infrared (FIR) sensors are the resolution of the imager and its sensitivity. Both have an impact on the detection capabilities of a processing system as they have an effect to texture and shape of objects displayed in images. Reducing costs of single components of a sensor is also combined with a change in the performance parameters of the sensor. The task for the system developer is to determine the maximal reasonable degradation of the sensor, which still yields a performance sufficient for the requirements of the implemented system.

The objective of the EU-project FNIR is to develop an automotive Night Vision Enhancement system using automatic pedestrian recognition, with a system performance beyond of current single sensor systems. The main targets of the project are improved pedestrian detection capability as well as cost reduction. Improved performance can be achieved by fusing a FIR-sensor with a NIR-sensor. Fusing with today’s high resolution and high sensitive FIR-sensors will be too costly to get a high market penetration though. The task of the fusion principle implemented in the FNIR-project is to enable the use of low performance FIR-sensors as well. Although a fusion detection algorithm shows significant better results than in the single sensor case, no direct conclusion can be made with respect to the expected performance of a classifier based on a low performance FIR-sensor. The reason for that is that the characteristic parameters of the sensor design are no input to the classifier.

The classifier is a result of statistical analysis of reference data and the direct influence of a sensor parameter to the classifier result is unknown. So far, the direct impact of the concrete sensor specification to the classifier result is unexplored and is only investigated by readapting classifiers using newly recorded data of different sensors.
The supervised training method used for training of the classifiers requires large sets of reference data. The reference database consists of a representative set of labeled images and is the basic instrument for development and evaluation of a detection algorithm. It must represent scenes with all typical variations of environment conditions regarding weather, temperature and illuminations, and all typical variations of appearances and poses of the class of objects it should detect, which is pedestrians in the task at hand.

Generating a representative reference database is very time consuming and expensive since every pedestrian has to be labeled manually. It is neither feasible nor affordable to generate ground-truth data for every possible sensor setup. The approach to overcome this problem is to use an already available reference database of high sensitive FIR-data for the currently used sensor setup and develop a transformation, that processes the data in such a way, that it reflects the particular design changes of a FIR-sensor with less performance. The expectation of the simulated data is to enable the possibility to train new classifier, in order to investigate to which extend the requirements of the FIR-sensor can be reduced without loosing the discriminative power necessary for a reliable pedestrian detection system. Simulating reduced resolution is considered to be straight forward, as the data only needs to be resampled, but well established algorithms for simulating reduced sensitivity are not available. The evaluation of the effects that degenerated sensitivity of FIR-sensors will have on the detection performance of a classifier is the main focus of this contribution.

Degenerated sensitivity of an FIR-sensor can be simulated with optics that have different f-numbers (f/no). The greater the f/no the less temperature radiation per unit reaches the sensor die. By evaluating the output of these degraded FIR-sensors the knowledge on what impact these effects have on a system has to be obtained. For this the data of the different sensors is analyzed and compared at the image level. Four different FIR-sensor setups are investigated and their performance differences are evaluated on real world scenarios. For the evaluation of the detection performance, a script and a scene catalog for recording comparable test databases for each sensor has been developed.

2. PEDESTRIAN DETECTION WITH CASCADED CLASSIFIERS

The main cost-drivers of the FIR-sensor which have an impact on the pedestrian detection capabilities are its resolution and its overall sensitivity. Both will have an effect to the features used for the classification of objects in a detection system. Low resolution has the effect that only large filters can be used for classification. A decrease in sensitivity results in less discriminative features.

2.1 Feature selection

The performance of the used sensor is an important factor for the selection of features for a successful separation of different classes. The implemented detection system for FNIR uses combinations of standard Haar-wavelet features, with extensions for statistical and histogram features, at varying position and scales (see feature in Fig. 2) to build a high dimensional feature space. Together they are able to describe texture and shape of objects in images and also achieve the task of an effective computation in real-time on current existing automotive hardware solutions.

2.2 Cascaded Classifiers

In order to realize a realtime detection system the implemented system incorporates cascaded classifiers. These classifiers are capable of separating regions in the input space into two classes for objects and background using a chain of classifier stages. With an increasing number of stages, the complexity of the classifiers is raised with the number of used features. Depending on the discriminative power of the features the resulting cascade can perform very efficiently. Each classifier of the cascade is scalar linear combination of weak classification functions. These weak learners simply compare one feature value against a trained threshold. The final strong classifier, or strong learner, forms a weighted combination of weak learners followed by a threshold (Fig. 2). The selection and parameterization is done using the supervised training algorithm AdaBoost, which incrementally selects those features with the lowest weighted error on the training set, until a predefined correct and false detection rate is achieved on a validation set.

The question is, how much performance the FIR sensor has to provide without loosing significant detection performance. Therefore classifiers based on data of degenerated FIR sensors would have to be evaluated. The
acquisition of the data needed for training and evaluation however is not realizable, which is why the approach of reusing available reference data is investigated in this paper.

3. FIR SENSOR PERFORMANCE EVALUATION

One performance parameter and cost-driver of a FIR-sensor is its sensitivity. It depends on the sensitivity of the classifier array itself and the characteristics of the optics. A typical measure of sensitivity is the noise equivalent delta temperature (NEdT), which describes the smallest temperature difference a sensor can measure. The classifier array of the sensor at hand has a resolution of $325 \times 240$ with $25\mu m$ pixel pitch and $14\text{bit/pixel}$ bit depth at a refresh rate of $33\text{ms}$. The spectral range of the sensor is at $7\mu m - 14\mu m$ with an absolute sensitivity of $0.1^\circ\text{C}$. The sensor output is not calibrated to a known truth but with focus on equal responsitivity of all
pixels for a good image uniformity with low spatial noise and to ensure temperature stability. During operation the sensor frequently resets its working point to satisfy these requirements.

The sensitivity of the sensor is affected by the $f$/no of the optical configuration. For example a sensor which achieves a NEdT of $40mK$ on chip has an overall sensitivity of $80mK$, when using an optic with $f$/no $1.4$ ($80mK = 40mK \cdot 1.4^2$). In the FNIR project four FIR-sensors with different optics have been provided in order to evaluate the influence of their effects. Based on a FIR-sensor with a NEdT of about $30mK$ on chip, the different $f$/no of the optics decrease the sensitivity accordingly (Table 1). The FIR sensor, which was used to record the existing reference database has a higher sensitivity with a NEdT of $42mK$ at $f$/no $1.4$.

$$\begin{array}{|c|c|}
\hline
f$/no$ & NEdT \\
\hline
1.4 & 52mK \\
1.55 & 71mK \\
1.8 & 94mK \\
2.2 & 130mK \\
\hline
\end{array}$$

Table 1. Impact of the optical configuration on a sensor with an on chip sensitivity of $\approx 30mK$.

### 3.1 Sensor analysis

A first impression of the differences of the sensors due to the reduced sensitivity can be investigated by capturing a homogeneous scene with a heat target and evaluating a histogram of it for each sensor. For each $f$/no increment there is a significant change in width and height of the different lobes (Fig. 3). This relationship describes the change in the sensors performance to resolve and distinguish the temperature radiation of the scene. However this kind of scene information and the histogram does not provide sufficient data to gain enough knowledge on how to describe the effects of the different $f$/no from one sensor to another.

![Figure 3. Histograms of the FIR sensors imaging the same homogeneous scene with a heat target. With increasing $f$/no the width and height of the histograms changes, implying a decreasing dynamic range.](image)

Reliable comparison of sensor data can only be achieved by comparing data that has been acquired under the exact same conditions while monitoring the same test scenario. For experiments in closed environments (laboratory tests). Thus all system and environment parameters can be easily controlled and fixed. This way
repeatedly execution of the tests produces the exact same results for devices under test. Sensors can be switched and tested at different times, to compare their performance, while all external parameters remain the same.

In an open environment, like open road scenarios, it is not possible to control or fix any external parameters. Even small changes in the environment conditions can have severe impact on the sensors performance and unpredictable influence on the data that has to be analyzed. The same is true for moving objects in the monitored scenario. Switching the different devices under test and repeating the experiments at different times is not an option, when it comes to reliable evaluation of sensor performance. The data that is to be analyzed has to be captured at the exact same time to solely represent differences in sensor performance instead of changes in the scenario or environment.

A basic approach for the analysis is to measure the intensity responses of the sensors pixels. By directly comparing two pixel responses of the same source signal, the resulting deviation represents the effects of differences in the components of the devices. In the case at hand the modified f/no of the optics, which affects the amount of temperature radiation passing through, will lead to lower responses in comparison to the reference sensor. Measuring this deviation is the first step for the evaluation of the different sensor characteristics. To be able to do that, images have to be captured that provide exactly the same scene content for each sensor at the same time. The time requirement can be satisfied by capturing the sensors in parallel.

To assure that each pair of compared pixel responses originates from the same source signal a simple alignment of the sensors is not sufficient. Even small changes of the installation angles of the sensor setup lead to high aberrations for the pixel position of the same source signal in the individual images. To compare the correct pixels the images have to have specifiable features, that allow a definite allocation. Fortunately, images that are normally needed for the calibration of the sensors when placed in the vehicle provide such features in terms of a heatable checker board (Fig.4). For each sensor images with these features were recorded in parallel with the reference sensor. The calibration algorithm itself is able to provide a precise localization of the individual checker board squares, which is eventually the needed information for the image registration process.

![Figure 4. Example of one calibration image with detected square corners used for image registration and analysis.](image)

Each individual square of the analyzed sensors is extracted and fitted with normalized 2-D cross-correlation onto the corresponding square of the reference sensor. Thereby, individual pixels can be directly compared to the reference sensor, which produces a large set of comparative values from the reference sensors response to the measured sensor response of the degenerated optics. A representation of these mappings from the codomain of the reference sensor to the codomain of the degenerated sensors produces a scatter-plot, where each entry represents a mapping between two intensity occurrences of the analyzed images. Each sensor has a distinctive scatter-plot with noticeable variation for the orientation of the lobes (Fig.5). With increasing f/no the slope of the plots is decreasing, which depicts how the codomain of the reference sensor is mapped on the smaller domain of the devices under test. This reflects the presumption from the investigated histograms in Fig. 3, that the dynamic range of the sensors suffers from the reduced sensitivity.
3.2 Development of a transfer-function

The scatter-plots imply, that a linear relationship between the sensors exists. The coefficients for a function that satisfies the data mapping can be estimated with a robust multilinear regression using iteratively reweighted least squares with a bisquare weighting function. This produces linear functions that fit into the different scatter-plots (Fig. 5).

From this follows that for each sensor a linear function $f_{S_{\text{REF}} \rightarrow S_{N}}$ can be defined that characterizes the mapping rule for intensity values from the reference sensor $S_{\text{Ref}}$ to one of the target sensors $S_{N}$ with f/no $N = [1.4, 1.55, 1.8, 2.2]$. The transformation of the reference image $I_{S_{\text{REF}}}$ into the target image $I_{S_{N}}$ of the sensor with f/no $N$, is performed by applying the mapping rule

$$I_{S_{N}}(i, j) = a_{0_{S_{N}}} + a_{1_{S_{N}}} \cdot I_{S_{\text{REF}}}(i, j)$$

for all pixels. The scaling and offset is done by the coefficients gained from the regression analysis (Table 2). Since the offset simply adds a fixed value it does not provide any information and could be omitted.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$a_0$</th>
<th>$a_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>833</td>
<td>0.89</td>
</tr>
<tr>
<td>1.55</td>
<td>2840</td>
<td>0.58</td>
</tr>
<tr>
<td>1.8</td>
<td>3758</td>
<td>0.40</td>
</tr>
<tr>
<td>2.2</td>
<td>4575</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table 2. Coefficients of the scatter-plot regression analysis for each f/no.

Figure 5. Scatter-plots of the intensity mappings, overlaid with the regression analysis results.

The performance of the transformation can be analyzed with the calibration database as well. Fig. 6 visualizes the signal response of one of the heated checker board squares for the recorded image of sensor $S_{1.55}$ compared to the transformed signal of the same square of the reference sensor. The deviation between the transformed and the real signal is less than 2% compared to the domain of the measured signal. These results qualify for a
classifier training that could provide the targeted information about the performance differences resulting from degenerated FIR sensors.

Figure 6. Comparison of one checker board square signal after applying the transformation with target sensor \( S_{1,55} \) (left) and the real data recorded with sensor \( S_{1,55} \) (right). The absolute value of the mean difference between the two signals is less than 2%.

4. DETECTION PERFORMANCE EVALUATION

The developed transformation process finally enables the possibility to generate significant databases for degenerated sensor data. For each target sensor a database was created, by transforming an existing database (more than 160,000 images and 160,000 labels of pedestrians) recorded with the reference sensor. Each database was used for a training of one single stream classifier.

For the evaluation of these new classifiers it is not possible to simply use an existing test database and transform it as well. This would only provide the basic information that the used database is able to produce a working classifier. In the case of these classifiers not only the detection performance itself is a question to be clarified, but also that the concept of the developed transformation process produces valid data and each classifier has the characteristics as if it was trained with real data. The logical consequence is, that the evaluation has to be performed with real data of the individual target sensors. This workflow is presented in Fig. 7. With test data of the reference and a target sensor recorded synchronously the detection performance can be directly compared.

4.1 Benchmark data generation

To provide a significant and fair database for the evaluation of the classifiers, the data recorded with one sensor has to have a comparable base of pedestrian scenes and complexity. For this purpose a recording session has to follow a script which exactly defines the extent and complexity of pedestrian scenarios that are to be recorded. The script covers significant appearances of pedestrians in road scenes in front of the car, divided in static scenes, focusing on distance performance and dynamic scenes, for availability performance. The session is once performed in an off road scenario with homogeneous background and sparse environment. The experiments are repeated in an urban area with a more complex scene to introduce realistic noise in the background.
Figure 7. The workflow for the training of a target classifier based on transformed data and the evaluation on real data of the target sensor.

Another focus of the experiments is a high pose variability to compete with the manifold of possible poses pedestrians can strike when standing, moving or performing different tasks like changing a tire, where a person as to hunker down on the side of the road. Most of the characteristic occurrences can be covered with a subset of scenes that are part of the catalogue. One specific test focuses on the impact of motion blur on the sensor performance, which occurs when the car is driving through road curves (Fig. 8).

Figure 8. Examples of scenes for the recording sessions of a test database for detection performance analysis.

To achieve comparable data a capturing system has been developed, that is capable to acquire 3 sensors in parallel, limited by space and hardware capabilities. Since 4 different sensors have to be analyzed several recording sessions have to be conducted. The reference sensor is always recorded to provide the benchmark data. The other sensors are paired into three different sessions. Table 3 shows the captured sensors for the different sessions. This allows the comparison of two sensors against the reference sensor for each campaign.
### 4.2 Classifier Evaluation

The recorded test data of all sensors has to be manually labeled to provide the ground truth for the classifier evaluation. In each campaign 75 sequences were recorded with more than 23,000 images for each recorded sensor in one campaign. This produced more than 760,000 labeled boxes of pedestrians overall. Each classifier is presented with all images from the corresponding sensors recording session and the detection results are validated against the ground truth. This produces three different ROC-curves for each recording session, where two of the degenerated classifiers can be directly compared against the performance of the reference classifier (see Fig. 9).

<table>
<thead>
<tr>
<th>Session no.</th>
<th>$S_{\text{REF}}$</th>
<th>$S_{1.4}$</th>
<th>$S_{1.55}$</th>
<th>$S_{1.8}$</th>
<th>$S_{2.2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$x$</td>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Acquired sensors for each recording session of evaluation data.

The recorded test data of all sensors has to be manually labeled to provide the ground truth for the classifier evaluation. In each campaign 75 sequences were recorded with more than 23,000 images for each recorded sensor in one campaign. This produced more than 760,000 labeled boxes of pedestrians overall. Each classifier is presented with all images from the corresponding sensors recording session and the detection results are validated against the ground truth. This produces three different ROC-curves for each recording session, where two of the degenerated classifiers can be directly compared against the performance of the reference classifier (see Fig. 9).

The results of the recording sessions (Fig. 9) show that the classifiers itself are capable of performing their task. The performance difference to the reference classifier is well established. To get an impression on how the classifier relatively perform against the reference this is satisfying. However, the differences of the results between the sessions show how much impact environment parameters like temperature or small scene changes have. The first and the third session vary only in $2^\circ C$ ($20^\circ C$ to $22^\circ C$). For the second session the classifier suffer from the circumstance, that the recordings for the test catalogue in the constant scenery introduced a static false alarm (Fig. 10), which is picked up by the classifier for $S_{1.55}$ and thus affects the ROC curve dramatically. This shows, that the analysis of the ground truth data is essential for the correct interpretation of ROC curves.

The recorded test scenes still do not provide enough variation and significance for all external impacts on the detection performance to achieve a satisfying comparison of the different classifiers. To improve the comparison of the classifiers the evaluation data is therefore expanded with random scenes of pedestrians in urban and suburban environment (about 25 sequences for each sensor) recorded during several sessions to provide a significant mixture of environment and background conditions. For this expanded evaluation set the detection and false alarm rates are more realistic and don’t show any significant external impact like for the homogeneous scenes of the first recording sessions alone (Fig. 11).

### 5. CONCLUSION

The presented results (Fig. 11) show that the degraded classifiers are capable of achieving a significant result for real data of the target sensors. The performance difference between the reference classifier an the classifiers

Figure 9. ROC curves of the classifiers for data of the three recording sessions. For each session the detection rates show the expected lower performance compared to the reference classifier. In session 2 classifier 1.55 is affected by a static false alarm due to the composition of the test data.
Figure 10. Correct detections on street level and static false alarm on a street light picked up by the classifier for $S_{1.55}$ (recording session no.2)

Figure 11. ROC curves of the classifiers for an expanded evaluation set of random urban and suburban pedestrian scenes recorded during several sessions.

trained on degraded data are preserved and meet the expected results. This means, that the developed transformation process is able to preserve the needed information for pedestrian features and provide the characteristics of a target sensor for a successful training.

The desired recycling of already recorded and with large investment cost generated ground truth data is possible with the presented approach of developing transfer functions for degenerated sensor data. The presented results of the impact of performance parameters of FIR sensors on the detection performance of pedestrian classifiers could only be achieved with the help of the developed transformation process. To which extend and accuracy this approach can be extended to other performance parameter of sensors of sensor classes is beyond the scope of the FNIR project.

6. FUTURE WORK

The transformation process is not limited to the presented sensitivity simulation alone. With the incorporation of the a performance parameter, like resolution, it is possible to widen the field of possible degenerated sensors and demonstrate performance limits of sensor configurations. Other effects like different noise sources or effects of the optical configuration could yield more accuracy for the transformation process.

From this point forward the gained results will be used for extended analysis of the impact of the FIR sensor performance on the fusion system developed in FNIR, by integrating the transformation process into the training
and evaluation process of fusion classifiers. This will ultimately help with the aspired goal of giving assistance to identify performance limits of sensor components to improve night vision pedestrian detection systems.

ACKNOWLEDGMENTS

Part of this research was supported by EU 7th Framework project 216384 FNIR, Fusing Far-Infrared and Near-Infrared Imaging for Pedestrian Injury Mitigation. All statements and opinions in this paper are the authors and should not be construed as representing official positions of the EU or project members.

REFERENCES


