A SIMILARITY METRIC FOR QUANTIFYING SYSTEM PERFORMANCE AT PEDESTRIAN DETECTION

Kip Smith and Askar Wubulikasimu
Department of Information and Computer Systems, Linköping University
SE 581 83 Linköping, Sweden
kip51smith@gmail.com

ABSTRACT: We introduce a MaxiMin metric of similarity between sets developed to assess the performance of in-vehicle active safety systems that provide warnings based on analysis of digital images. The metric is an asymmetric variant of the Hausdorff Distance that differentially weights false negatives (misses) and false positives (false alarms). In this paper we discuss an application of the metric to data collected by a night-vision system that automatically detects pedestrians. We discuss two cases with unusually low minimum values of the metric, situations where it is likely that pedestrian detection could be improved. The metric is sufficiently general to be applicable to a broad range of sensor systems and is not limited to night-vision technology.

1 OBJECTIVE

The design of advanced driving assistance systems such as pedestrian detection systems requires the consideration of trade-offs between functionality, cost, durability, and a myriad of other concerns. Typically, trade-off analyses require quantitative metrics of system performance. For example, an analysis of the performance of an infra-red night-vision system requires a metric that is able to quantify the effects of factors that influence pedestrian detection. A short list of such factors includes the distance to the pedestrian and the ambient thermal background. A metric that pinpoints factors that adversely influence functionality, e.g., pedestrian detection, would facilitate comparisons across systems and assessments of the contributions of changes to system design. Performance metric-based assessments can readily be folded into cost-benefit trade-off analyses and would likely be an important driving force for the development of more reliable pedestrian detection systems.

We introduce a MaxiMin metric of similarity between sets developed to assess the performance of in-vehicle active safety systems that provide warnings based on analysis of digital images. The primary application has been the evaluation of ‘night-vision systems’ that use infrared technology to detect pedestrians in the dark and to generate displays that alert drivers to their presence. Our focus here is the metric we use to quantify the quality of pedestrian detection. We do not discuss the algorithm that extracts pedestrians from the scene or the user interface that highlights the location of detected pedestrians on a display screen.

2 METHOD

The raw data are images captured by a far-infrared (FIR) sensor system. FIR systems detect thermal radiation at wavelengths in the interval 8-12 μm. In good conditions, they are capable of detecting and highlighting pedestrians at distances greater than 100 m. Proprietary software generates time-stamped lists of the coordinates of rectangles that define the locations of pedestrians in every frame.

The metric quantifies the similarity between two such lists and is applied to each frame. The rectangles in set S (system) are generated in ‘real-time’ by system software. The rectangles in set G (ground-truth) are drawn frame-by-frame by trained technicians who inspect the FIR images visually. At each frame, the inputs to the metric are the n locations of pedestrians detected by the system and the m locations of pedestrians that constitute the ground truth. We consider only the horizontal component of location because there is little vertical translation in the images and the critical movements, like crossing the street, are horizontal.

The metric is an asymmetric variant of the Hausdorff Distance that differentially weights false negatives (misses) and false positives (false alarms). For the pedestrian detection task, a miss receives much greater weight as this error is by far the more serious. The Hausdorff distance between two sets of points in an arbitrary metric space is the maximum distance of a point in one set to the nearest point in the other set [1]. We calculate the Hausdorff distance from one set to the other using Equation 1:

$$min_{b \in B}\{d(a,b)\}\} \qquad (1)$$

where *a* and *b* are points in the sets *A* and *B*, respectively, and *d(a,b)* is the Euclidian distance between them. When n and m differ*, h(G, S)* is generally not equal to *h(S, G).* The general Hausdorff distance is defined by Equation 2:

$$H(G,S) = max\{h(G,S), h(S,G)\} \qquad (2)$$

The Hausdorff distance is undefined if either *G* or *S* is empty which, in our application, occurs whenever there a false alarm or a missed detection. To account for the possibility of these errors, we augment both sets *G* and S with two points representing the left and right margins of the image. The end points influence the metric only when there is no detection in either set G or *S* and similarity is, accordingly, low. Adding the two end points to both sets has two benefits. First, we can use the Hausdorff distance to quantify the distance between system detections and ground truth when one or the other fails to detect a pedestrian. Second, missed pedestrians have their greatest impact on the metric when in the middle of the frame, directly in front of the vehicle.

Based on the Hausdorff distance, we introduce the following expression to calculate the similarity between the ground truth and the corresponding system output:

$$Similarity = 1 - \frac{[\propto h(G,S) + (1-\propto)\, h(S,G)]}{\frac{D}{2}} \qquad (3)$$

where, *h(G, S)* denotes the distance from the augmented set *G* representing the ground truth to the augmented set *S* of system detections. The free parameter α Є[0, 1] is the weight assigned to the severity of missed detections and *D* is the width of the image in pixels. Because the maximum distance the ground truth and the system output cannot exceed *D/*2, we use that factor to normalize the similarity to a value between 0 and 1. Accordingly, the metric of similarity will be 0 in the worst case - when the pedestrian stands directly right in front of the vehicle but is not detected by the system. The metric will be 1 when the system detects every pedestrian at the same position as the ground truth.

3 ILLUSTRATIVE RESULTS

3.1 Using the metric

During the course of the project we have calculated the metric for hours of data. We use the metric to detect outliers - cases with unusually low minimum values of the metric. We inspect those cases to learn (a) situations where there is likely room to improve system detection, and (b) to identify sequences that are worth further study using post-hoc reviews [2]. Here we discuss two illustrative examples.

3.2 Occlusion

Figure 1 shows a typical situation involving occlusion. As shown in the snapshot on the left of Figure 1, there are four pedestrians in this sequence. Two are walking on the left side of the street and one is walking on the right. The fourth is crossing the street from right to left and become partially occluded by the pedestrian in the right foreground. As the crosser passes behind the pedestrian in the foreground, a discriminating observer can find his legs and torso ‘peeking’ behind the other pedestrian’s legs and arms. However, the sensor system is currently not able to extract such subtle cues. As a result, the crosser is temporarily occluded from detection by the system. In the snapshot shown in Figure 1, the crosser has just emerged from occlusion and the system has yet to (re)detect him. This situation is technically a ‘miss’.

The graph shows the time series of the metric for the entire sequence. The horizontal axis is the value of metric. Note that the minimum value in this instance is greater than 0.5. The vertical axis is frame number, a proxy for time. Frame rate is approximately 30 Hz. Time increases down.

The graph shows approximately 8 s of data. The metric is generally greater than 0.9, indicating a high level of similarity between the system detection and the ground truth. The saw-tooth pattern near frame 250 reflects a sequence of detections: a pedestrian is initially detected in the ground truth but not by the system. The metric drifts down until the system detects the pedestrian. Upon detection, the metric jumps to a higher value. The size of the saw-tooth reflects the lag between the time when a pedestrian first appears in the ground truth and when he or she is detected by the system.

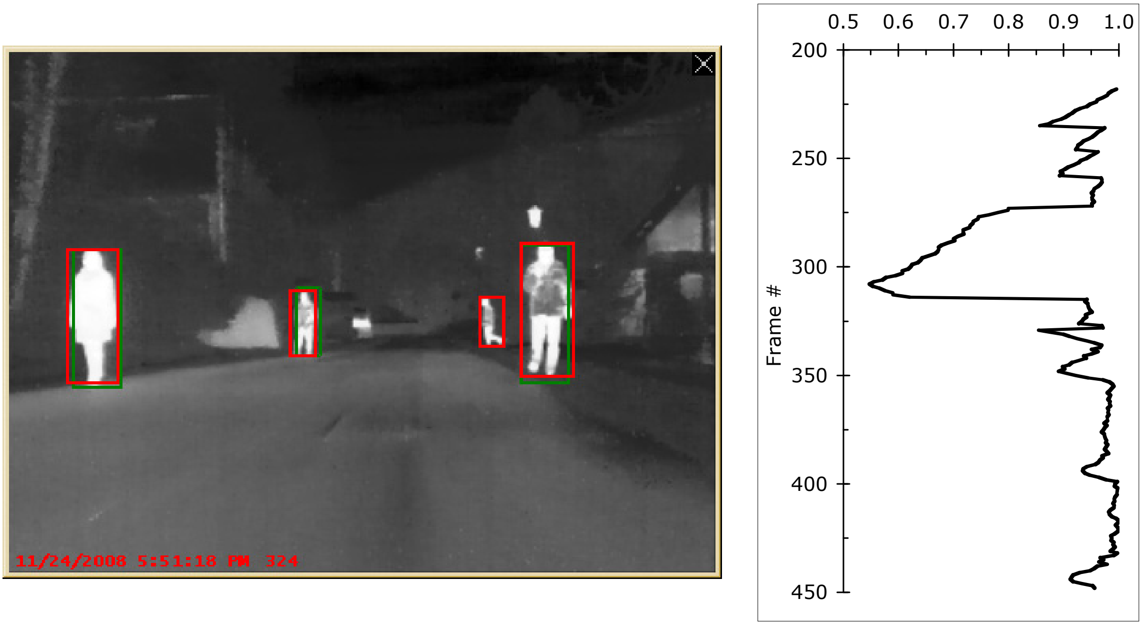

Fig. 1. An infrared snapshot of a pedestrian emerging from occlusion (crossing from right to left) and the time series the metric in response to that occlusion

The prominent excursion at frame 300 occurs when the crosser becomes hidden behind the pedestrian in the foreground. The metric immediately drops because (a) the system can no longer 'find' the crosser but (b) the trained technician who provided the ground truth was able to identify the crosser's legs and torso 'peeking' behind the other pedestrian's legs and arms. The excursion indicates the relative seriousness of the missed detection for the 3 s when the ground truth found four pedestrians and the system only three.

The magnitude of the excursion is determined by the location of the crosser in the image. When detection is lost, his location is near the right side of the image. The initial drop in the metric to approximately 0.8 reflects the loss of detection. As time passes, the crosser moves to the left into the image and metric steadily declines. If he had reached the center of the image (street) without being detected, the metric would have declined to zero (0.0). Instead, the system detected him shortly after the snapshot in Figure 1, well before he was one-quarter of the way into the image. Upon detection, the metric immediately jumps back above 0.9.

3.3 Background thermal radiation

Figure 2 shows an example where background thermal radiation impaired the system's ability to extract pedestrians from the scene. The snapshot on the left shows three pedestrians in an urban environment. One is on the left sidewalk, a second is crossing the street, and the third is on the right sidewalk. The storefronts and sidewalks are emitting a lot of heat. As a result, the pedestrians on the sidewalks are occasionally dropped from detection by the system. In the snapshot shown in Figure 2, the system has dropped its detection of the pedestrians on both sidewalks. This situation is another example of a ‘miss’.

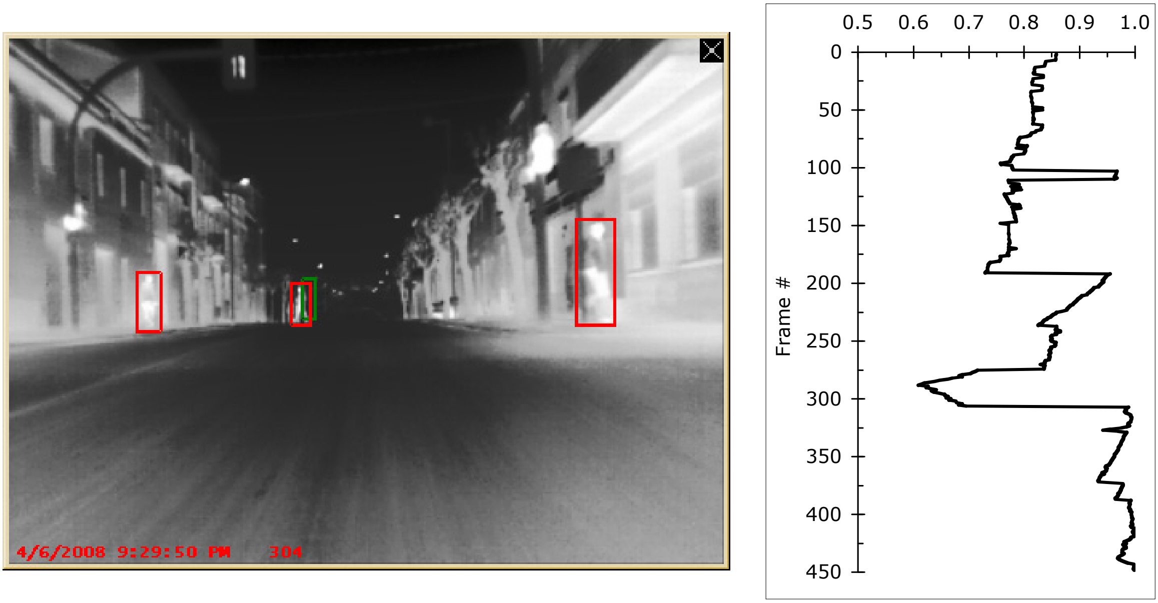

Fig. 2. An infrared snapshot of pedestrians obscured by background thermal radiation and the corresponding time series the metric

The graph shows the time series of the metric for approximately 15 s. The metric hovers near 0.8, indicating that the metric has not detected pedestrians who are relatively close to the edge of the image. The spikes to the right near frames 100 and 200 reveal transient system detections of one of the pedestrians. The excursion to the left near frame 300 occurs when the system drops detection of both pedestrians on the sidewalk, as shown in the snapshot.

This example identifies an opportunity for system refinement. The metric takes a large set of images and pinpoints the exact times and locations where improvements are required, enabling focused analysis of the reasons for the differences in performance.

4 CONCLUSIONS

The metric is sensitive to the location and distribution of pedestrians and to occlusion. Background thermal radiation remains a challenge. These findings suggest that the metric and its formulation are fundamentally sound.

We are in the process of introducing a weighting factor for the relative level of risk posed by the vehicle to the pedestrian. Our proxy for risk is proximity as revealed by the height of a pedestrian in the image. Closer (taller, more pixels in the vertical dimension) pedestrians are assumed to be at greater risk. Introducing a weighting factor presents issues of scaling that require making trade-offs between the differential influences of proximity and location. As there is no objectively correct balance between proximity and location, we are collecting subjective input from drivers to inform our selection of weights [3].

The metric is sufficiently general to be applicable to a broad range of sensor systems and is not limited to night-vision technology. A potential application of the metric is to compare the outputs of successive generations of sensor systems. If two or more systems are mounted on the same vehicle and metrics calculated for all, the ratio of the metrics provides an index of the improvement provided by the next-generation systems. The metric's characterization of the differences in performance informs design decisions that are translated into real world functionality.

5 ACKNOWLEDGEMENTS

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. The authors thank Anders Grimvall for his inspiration.

Kip Smith is now a Senior Lecturer in the Operations Research Department at the U.S. Naval Postgraduate School, Monterey, CA. Askar Wubulikasimu is now in the PhD program in Statistics at the Free University of Amsterdam.

6 REFERENCES

[1] Munkres, J.: 'Topology' (Prentice Hall, 1999)

[2] Smith, K., and Källhammer, J.-E.: 'Driver acceptance of false alarms to simulated encroachment', Human Factors. (In review).

[3] Källhammer, J.-E., Smith, K., Karlsson, J., and Hollnagel, E.: 'Shouldn't the car react as the driver expects?' Proc. 4th Int. Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Stevenson, WA. June, 2007.