Pose-Specific Pedestrian Classification using Multiple Features in Far-Infrared Images

Dong-Seok Kim
Center for Integrated Smart Sensors
312 IT Convergence Center(N1)
291 Daehak-ro, Yuseong-gu
305-701, Daejeon, Korea
kds1130@kaist.ac.kr

Ki-Yeong Park
Center for Integrated Smart Sensors
312 IT Convergence Center(N1)
291 Daehak-ro, Yuseong-gu
305-701, Daejeon, Korea
cpky0@kaist.ac.kr

ABSTRACT

We present a multiple feature-based, pose-specific pedestrian classification approach to improve classification performance for far-infrared (FIR) images. Using pose-specific classifiers and multiple features has proved to be beneficial in visible-spectrum-based classification systems; therefore, we adapt both to an FIR-based classification system. For pose-specific classifiers, we separate poses into sets of front/back and right/left poses and estimate the pose using template matching. For feature extraction, we use histograms of local intensity differences (HLID) and local binary patterns (LBP). Experiments showed that the proposed approaches improve the classification performance of a baseline HLID/linSVM approach.

Keywords

Pedestrian classification, Multi-features, Pose templates, Far-infrared images

1. INTRODUCTION

Far-infrared (FIR) cameras (or thermal cameras), capture the heat emitted from objects, so pedestrians typically appear brighter than backgrounds in FIR images (see Fig. 1). Thus, FIR technology is advantageous for pedestrian detection at nighttime. Because of this characteristic, a major subject of previous studies has been candidate generation processes for finding regions in images that are highly likely to contain a pedestrian. For classification, most studies have simply used single feature-based classifiers, such as histograms of oriented gradients (HOG) [1, 2]. For pedestrian classification in visible spectrum images, it has been shown that an ensemble of classifiers that has been trained for particular pedestrian poses outperforms a single classifier that has been trained for the entire data set [3]. It has also been shown that methods that use a combination of multiple features are able to improve the classification performance as compared to methods that only use a single feature [3].

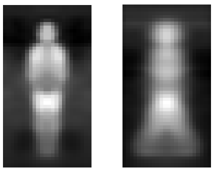
Despite the benefits of using pose-specific classifiers and multiple features that are expected in FIR images, only a few studies have dealt specifically with this subject. In [4], appearance-based classifiers, which consisted of “along-street,” “across-street,” and “bicyclist,” were utilized and decisions were simply based on the logical OR process for all classifiers. However, the combination of these classifiers risks generating numerous false positives. In [5], multiple features were investigated, but decisions were based on the classification result using only one of the features after the feature selection process, instead of fusing the multiple features. Thus, it is difficult to expect better performance from this method as compared to that from single feature-based classification methods.

In this paper, we investigate the benefit of using pose-specific classifiers and multiple features in FIR images. For pose-specific classifiers, pose estimation is performed simply by correlation with pose

(a) (b)

Figure 1. Sample images of pedestrians. (a) Visible spectrum image. (b) Far-infrared image.

(a)



(b)

Figure 2. Pedestrian example images. (a) Samples of pose variations of front, back, left, and right. (b) Templates of front/back and left/right

templates. For feature extraction, we introduce a novel combination of HLID [6] and cell-structured LBP [7] features. We compare the proposed approach to the HLID/linSVM classifier approach because the motivation of this study is to gain performance improvements over our previous work that used an FIR-specified feature [6].

2. PEDESTRIAN CLASSIFICATION

To improve classification performance, we adopt pose-specific classifiers and multiple features to FIR-based pedestrian classification. For the pose model, only two poses of front/back and right/left are considered instead of four poses (front, back, right, and left), as is the case in visible spectrum image-based pedestrian classification because of the similar contours between the combined poses. Furthermore, it makes the classification problem simple and reduces the computational power. Fig. 2 shows the samples as pose variations and the pose templates that are generated by averaging the intensity of manually separated positive samples of poses from training sets. Regarding features, we chose HLID features and cell-structured LBP features. HLID was selected because it has been shown to outperform HOG in FIR-based classifications [6]. Further, LBP was selected because it was expected to compensate the problems (sensitivity to noisy background edges) of HLID using its uniformity constraints [7].

For combining information from multiple poses and multiple features, we employed a mixture-of-experts (MoE) framework introduced in [3]. In the MoE approach, the posterior probability that a given sample (x_i) is a pedestrian class ((ω_0)) is ($P(\omega_0|x_i)$), which is approximated with a sample-dependent weight $w_k(x_i)$ and a pose-specific classifier output $H_k(x_i)$ with pose clusters k as

$$P(\omega_0|x_i) \approx \sum_k w_k(x_i) H_k(x_i) \qquad (1).$$

Given the pose-specific MoE model, the pose-specific expert classifier $H_k(x_i)$ was modeled in terms of our multiple feature set (*f*) as

$$H_k(x_i) = \sum_f v_k^f I_k^f\left(x_i^f\right) \qquad (2).$$

Here, $I_k^f\left(x_i^f\right)$ denotes a local expert classifier for the *k*th pose cluster with features *f* from a feature set, and v_k^f represents a pose and feature dependent weight with $\sum_f v_k^f = 1$. For expert classifiers I_k^f, we used a linear support vector machine (linSVM) to train the classifiers from the corresponding pose and feature only. Given K (2 of front/back and right/left) pose clusters and F (2 of HLID and LBP) features, we trained $K \times F$ classifiers I_k^f on the pose-specific training set. Weights v_k^f were used to model the contribution of the individual classifiers. Hence, we derived the weights by the discriminative power of the individual expert classifiers using a training dataset. The sample-dependent weight $w_k(x_i)$ was decided using similarity between pose templates t_k and the sample x_i as

$$\omega_k(x_i) = \frac{corr(x_i, t_k)}{\sum_k corr(x_i, t_k)} \qquad (3).$$

To measure the similarity, we used simple template matching using Pearson's correlation measures. Both templates and samples were normalized before matching. For the weight function, the weights of sample outputs of pose-specific classifiers were determined proportionally by their similarity to the pose template with $0 \le w_k(x_i) \le 1$ and $\sum_k w_k(x_i) = 1$. Using the weight function, this method can lower the risk of degradations in the classification process that are caused by incorrect pose decisions.

3. EXPERIMENTAL RESULTS

The proposed method was evaluated using 6573 FIR images that were taken from moving vehicles in an urban area at nighttime. We split the set of images into training sets and test sets according to the days of images captured: 4668 images for training and 1905 images for test. The training samples were cropped from the training set and then divided into two different pose sets of front/back and right/left. Test samples were cropped automatically from the test set using the sliding window technique based on the overlap ratio between the current window and the manually labeled pedestrian ground truth (we choose the current window as test sample when the overlap ratio exceeds over 70%). Table 1 gives an overview of the dataset. Samples were resized to 24 × 48 pixels.

	Pedestrians (front/back)	Pedestrians (right/left)	Non-pedestrian
Training set	2209	1879	8555
Test set	13123		10173

Table 1. Datasets for evaluation

We computed $HLID_{8,2}$ features using a cell size of 6 × 6 pixels, block size of 2 × 2 cells, overlap of 0.5 blocks, and *L2-norm* block normalization. To extract LBP features, we computed $LBP_{8,2}$ features using a cell size of 8 × 8 pixels and a maximum transition number of 2. Expert weights v_k^f were estimated by a linSVM on the training set (0.52 for HLID and 0.48 for LBP for both poses). To quantify the performance, we plotted the detection error tradeoff (DET) curves on a log-log scale on both a per-window and per-image evaluation. We followed the evaluation method used in [8].

First, we compared our proposed pose-specific and multiple feature-based classification approach to the single feature-based classifier, pose-specific classifier, and multiple feature-based classifier approaches. We selected HLID as the baseline feature to show the performance improvements over the FIR-specified feature. Multiple feature-based classification was conducted by concatenating two features into a single feature vector. The results are shown in Fig. 3(a). As expected, the proposed approach that combines pose-specific and multiple feature-based classifiers outperformed other approaches. We also found that both the pose-specific classifier approach and the multiple feature-based classifier approach outperformed the single feature-based classifier. The results confirm that the pose-specific classifier approach performs better because the pose variations are relatively smaller than the classifier trained on a whole dataset irrespective of pose. Further, the combination of multiple complementary features boosts the performance.

Next, we compared the MoE framework to simple combination rules to see the performance if the fusion method is varied. For simple combination rules, the concatenated multiple features were classified based only on the selected pose-specific classifier of having maximum pose similarity. Fig. 3(b) shows that the MoE approach outperforms the simple combination approach. The differences were mainly caused by errors in pose estimation and by the use of the same weights for features without consideration of the discriminative power of each feature.

Finally, we evaluated our proposed method on a per-image basis to compare with the single feature-based classifier described in [6], and checked for

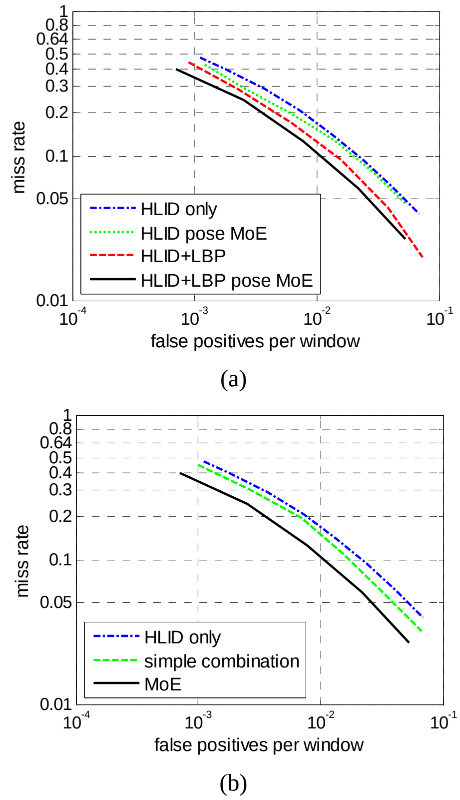


Figure 3. Performance comparison results with per-window evaluations. (a) Performance based on information of classification variations. (b) Performance based on fusion method variations.

improvements for pedestrian detection in FIR image sequences. Except for the classification method, all of the other procedures and evaluation methods are the same. Fig. 4 shows that our proposed method improves the pedestrian detection performance by reducing the miss rate by approximately 4% at 10^{-1} false positive per image (FPPI). These results

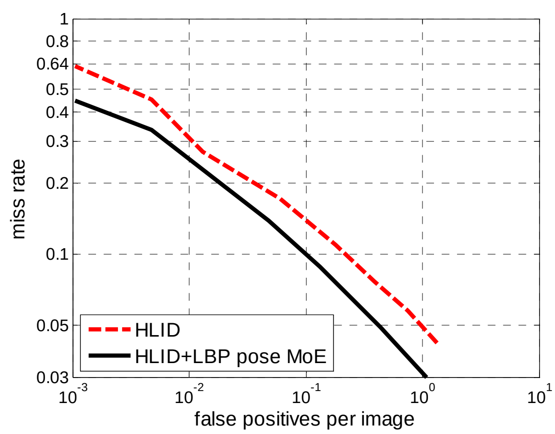


Figure 4. Pedestrian detection performance comparison between the proposed classifier and the baseline HLID/linSVM classifier.

Figure 5. Detection examples of representative scenarios of single or multiple pedestrians with pose variations (the red boxes indicate the detection results).

demonstrate the benefit of our proposed method. In order to gain more performance improvements, it will be necessary to upgrade pose estimation accuracy. This will be investigated in a future work. Fig. 5 shows some detection examples in FIR images.

4. CONCLUSION

We proposed a pose-specific pedestrian classification using multiple features in FIR images. Experiments showed the proposed approaches outperform single feature-based classifier. Reducing pose variation is helpful for FIR-based pedestrian classification. Further, the newly introduced combination of HLID and LBP features proved to be beneficial. We hope that our results will help promote further research on classifiers in FIR-based pedestrian detection systems.

5. ACKNOWLEDGMENTS

This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Education, Science and Technology as Global Frontier Project (CISS-2011-0031863).

6. REFERENCES

[1] O'Malley, R. Jones, E. Glavin, M.: 'Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation', I P & T, Vol.53, pp.439-449, Nov 2010.

[2] Dalal, N. Triggs, B.: 'Histograms of oriented gradients for human detection'. IEEE Conf., CVPR. 2005, Vol.2, pp.886–893, June 2005.

[3] Enzweiler, M. Gavrila, D, M.: 'A Multilevel Mixture-of-Experts Framework for Pedestrian Classification,' Image Processing, IEEE Trans., Vol.20, pp.2967-2979, Oct 2011.

[4] Fengliang, X. Xia, L. Fujimura, K.: 'Pedestrian detection and tracking with night vision,' ITS, IEEE Trans., Vol.6, pp.63-71, 2005.

[5] Li, Z. Bo, W. Nevatia, R.: 'Pedestrian Detection in Infrared Images based on Local Shape Features,' CVPR., IEEE Conf., pp.17-22, 2007.

[6] Kim, D, S. Kim, M. Kim, B, S. Lee, K, H.: 'Histograms of local intensity differences for pedestrian classification in far-infrared images,' Electronics Letters, Vol. 49, issue. 4, pp. 258-260, 2013.

[7] Wang, X. Han, T. X. Yan, S.: 'An HOG-LBP human detector with partial occlusion handling', CV, IEEE Conf., pp.32-39, 2009.

[8] Dollar, P. Wojek, C. Schiele, B. and Perona, P.: 'Pedestrian Detection: An Evaluation of the State of the Art,' Pattern Analysis and Machine Intelligence, IEEE Trans., Vol. 34, No. 4, pp. 743-761, Apr. 2012.