Image Classification by Multi-Class Boosting of Visual and Infrared Fusion with Applications to Object Pose Recognition

Image Classification by Multi-Class Boosting of Visual and Infrared Fusion with Applications to Object Pose Recognition

Yixiao Yun, Irene Yu-Hua Gu
Dept. of Signals and Systems, Chalmers University of Technology, Sweden
Email: yixiao@chalmers.se

***Abstract*—This paper proposes a novel method for multi-view object pose classification through sequential learning and sensor fusion. The basic idea is to use images observed in visual and infrared (IR) bands, with the same sampling weight under a multi-class boosting framework. The main contribution of this paper is a multi-class AdaBoost classification framework where visual and infrared information interactively complement each other. This is achieved by learning hypothesis for visual and infrared bands independently and then fusing the optimized hypothesis sub-ensembles. Experiments are conducted on several image datasets including a set of visual and thermal IR images containing 4844 face images in 5 different poses. Results have shown significant increase in classification rate as compared with an existing multi-class AdaBoost algorithm SAMME trained on visual or infrared images alone, as well as a simple baseline classification-fusion algorithm.**

I. Introduction

Recognizing poses of objects (e.g., face, vehicle) has been one of the most important research topics in computer vision, largely driven by many applications such as robotic surveillance, visual-based monitoring of driver awareness or automating camera management [1].

Several face pose classification methods have been proposed and developed recently. [2] uses PCA-based face features and soft margin AdaBoost to detect the frontal views. [3] extracts features inspired by [4] and builds five separate AdaBoost classifiers for face images in each class. [5] presents a nested cascade detector for face poses in 5 classes using confidence-rated AdaBoost [6] based on Haar features. [7] introduces a tree-structured classifier for face poses in 7 classes, and each node is a three-class classifier trained by AdaBoost.MH. Good results have been achieved, however, these methods mainly adopt *one-against-all* or *one-against-one* strategies [8] for multi-class problems, so model complexities may be increased.

To improve the classification of objects, approaches are proposed on fusion of visual and infrared information. [9], [10] and [11] present fusion methods at the sensor level. [12] uses decision fusion of neural classifiers for real time face recognition. [13] introduces fusion scheme at different levels for SVM-based obstacle classification. These methods usually combine multiple individual features or decisions in a one-off manner, however, the interactive relations between visual and infrared observations are seldom considered. Despite these efforts, classifying object poses using both visual and infrared observations remains an open issue.

To tackle these problems, we propose a novel method fusing visual and infrared information interactively under a boosting framework for multi-view object pose classification. Different from one-against-all or one-against-one strategies, our model is similar to *SAMME* [14] in true solution to multi-class problems, however, a new part of sensor fusion is introduced. The main contributions of this paper include using sub-ensemble learning for fused hypothesis optimization and suggesting effective feature for thermal IR images. Improved classification results are demonstrated by empirical evaluation compared with SAMME using visual or infrared images alone, as well as a simple baseline classification-fusion algorithm.

The rest of this paper is organized as follows: Section II gives a big picture of the proposed framework; Section III makes some review of AdaBoost algorithms; Section IV describes the proposed classification and fusion scheme; Section V shows experiment results on several datasets including a set of visual and thermal IR images and comparisons with three closely related methods; finally Section VI concludes the paper.

II. Overview of Proposed Scheme

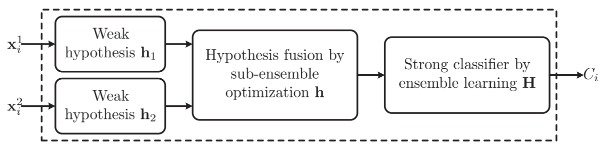

Fig. 1. Block diagram of proposed scheme. The dashed box represents the boosting framework. $\mathbf{x}_i^1$, $\mathbf{x}_i^2$, C_i denote visual features, IR features, and predicted class labels of i-th object, respectively.

As shown in Fig.1, the proposed framework consists of three major parts: (a) independent hypothesis learning using visual and infrared features with the same sampling weight; (b) fusion by optimizing hypothesis sub-ensemble; (c) adding sub-ensemble to a final strong classifier and updating sampling weight distribution with a scale factor. The essence for using the same sampling weight is to force weak classifiers for both visual and infrared bands to focus on the same objects, therefore hypotheses independently learned from visual and infrared features match each other. The basic idea for hypothesis optimization is to add hypotheses for both bands to the sub-ensemble, with sub-ensemble weights according to their accuracies, so that hypothesis sub-ensemble may have enhanced performance based on fusion of visual and infrared information. In this way, the final strong ensemble may have further improved accuracy. The main motivation for using a scale factor to update sampling weights is to make weak classifiers focus on those difficult objects misclassified in both visual and infrared bands. The main novelty lies in two-stage ensemble learning under a multi-class boosting framework, by using visual and infrared information in this interactive manner, which may lead to better classification results.

III. AdaBoost: Review

This section briefly reviews AdaBoost algorithms, with emphasis on SAMME, which our proposed classification and fusion scheme is built upon.

AdaBoost is an ensemble learning method originally intended only for binary problems. Many extensions of AdaBoost for multi-class problems exist, and most of them have been restricted to using one-against-all or one-against-one strategies [8]. SAMME, one of the true multi-class AdaBoost algorithms, is a true multi-class classifier that solves multi-class problems without reducing them to multiple binary subproblems.

Let $\mathbf{X} = \{\boldsymbol{\chi}_i\}$, $i = 1, 2, ..., N$ be the entire training set containing feature vectors of objects. Let the class label (denoted by C) be represented as a K-dimensional vector $\mathbf{y} = (y_1, y_2, ..., y_K)^T \in \mathcal{Y}$, where $y_k = 1$ if $C = k$, otherwise $y_k = -1/(K-1)$, $k \in \{1, 2, ..., K\}$ and $K \geq 3$ is the number of classes. In such a way, $\mathbf{Y} = \{\mathbf{y}_i\}$ is an equivalent set of class labels corresponding to $\mathbf{X}$. The output of weak classifier for each feature vector, the hypothesis $\mathbf{h} = (h_1, h_2, ..., h_K)^T \in \mathcal{Y}$, is encoded in the same way.

The goal is to minimize the objective function as exponential loss function $L(\mathbf{Y}, \mathbf{H}) = \sum_{i=1}^{N} \exp\left(-\frac{1}{K}\mathbf{y}_i^T \mathbf{H}(\boldsymbol{\chi}_i)\right)$ by learning a strong ensemble

$$\mathbf{H}^{(t)}(\boldsymbol{\chi}_i) = \mathbf{H}^{(t-1)}(\boldsymbol{\chi}_i) + \alpha^{(t)}\mathbf{h}^{(t)}(\boldsymbol{\chi}_i) \tag{1}$$

subject to the constraint $\sum_{k=1}^{K} H_k(\boldsymbol{\chi}_i) = 0$. Several boosting rounds $t = 1, ..., T$ is applied. In each boosting round, the sampling weight $D_i^{(t)}$ for each feature vector of objects, weighted errors $\epsilon^{(t)}$ for the weak classifier and the ensemble weight $\alpha^{(t)}$ for each hypothesis that is added to the ensemble are updated as follows:

$$D_i^{(t)} = \exp\left(-\frac{1}{K}\mathbf{y}_i^T \mathbf{H}^{(t-1)}(\boldsymbol{\chi}_i)\right) \tag{2}$$

$$\epsilon^{(t)} = \sum_{i=1}^{N} D_i^{(t-1)} \mathbb{I}(\mathbf{y}_i^T \mathbf{h}^{(t)}(\boldsymbol{\chi}_i) \leq 0) / \sum_{i=1}^{N} D_i^{(t-1)} \tag{3}$$

$$\alpha^{(t)} = \frac{(K-1)^2}{K}\left(\log\frac{1-\epsilon^{(t)}}{\epsilon^{(t)}} + \log(K-1)\right) \tag{4}$$

where $\mathbb{I}(A)$ is an indicator function which equals 1 if event A is true, and 0 otherwise.

IV. Multi-class boosting with hypothesis fusion

A sub-ensemble learning method fusing hypotheses learned from visual and infrared features under a multi-class AdaBoost framework is introduced in this section. Each object feature vector $\boldsymbol{\chi}_i$ contains two component feature vectors $\{\mathbf{x}_i^1, \mathbf{x}_i^2\}$, corresponding to visual and infrared bands, respectively.

In the proposed method, we enforce a same set of sampling weights to the weak classifiers for both visual and infrared bands on the same objects, therefore hypotheses independently learned from visual and infrared features match each other, yielding $\mathbf{h}_m(\mathbf{x}_i^m)$, $m = 1, 2$. Different from multiple AdaBoost classifiers trained on single-band features with independent sampling weights, the interaction between visual and infrared information in our case is conducted at each boosting round inside the boosting structure.

The objective criterion of the proposed scheme is to minimize the exponential loss function

$$L(\mathbf{Y}, \mathbf{h}) = \sum_{i=1}^{N} \exp\left(-\frac{1}{K}\mathbf{y}_i^T \mathbf{h}^{(t)}(\boldsymbol{\chi}_i)\right) \tag{5}$$

through learning a sub-ensemble of hypotheses

$$\mathbf{h}^{(t)}(\boldsymbol{\chi}_i) = \sum_{m=1}^{M} \beta_m^{(t)} \mathbf{h}_m^{(t)}(\mathbf{x}_i^m) \tag{6}$$

subject to the constraints $\sum_{k=1}^{K} h_k(\boldsymbol{\chi}_i) = 0$ and $\sum_{m=1}^{M} \beta_m^{(t)} = 1$, $M = 2$. The solution is shown to be:

$$\beta_m^{(t)} = \frac{\log\left(\frac{1-\epsilon_m^{(t)}}{\epsilon_m^{(t)}}(K-1)\right)}{\log\left((K-1)^M \prod_{m=1}^{M} \frac{1-\epsilon_m^{(t)}}{\epsilon_m^{(t)}}\right)} \tag{7}$$

where

$$\epsilon_m^{(t)} = \sum_{i=1}^{N} \mathbb{I}(\mathbf{y}_i^T \mathbf{h}_m^{(t)}(\mathbf{x}_i^m) \leq 0)/N \tag{8}$$

β_m is the sub-ensemble weight for each single-band hypothesis that is added to the sub-ensemble and ϵ_m is the error rate for each single-band hypothesis.

A scale factor $\gamma_i^{(t)}$ is then introduced for $\boldsymbol{\chi}_i$, which is exponentially proportional to the count of misclassification by the two weak classifiers

$$\gamma_i^{(t)} = \lambda^{\eta_i} \tag{9}$$

where $\lambda > 0$ is a constant, and $\eta_i = \sum_{m=1}^{M} \mathbb{I}(\mathbf{h}_m^{(t)}(\mathbf{x}_i^m) \neq \mathbf{y}_i)$. In such a way, objects correctly classified by weak classifiers in both visual and infrared bands lose more weights, and objects misclassified by both weak classifiers are treated as difficult objects by gaining more weights:

$$D_i^{(t)} = \gamma_i^{(t)} D_i^{(t-1)} \exp\left(-\frac{1}{K}\beta^{(t)}\mathbf{y}_i^T \mathbf{h}^{(t)}(\boldsymbol{\chi}_i)\right) \tag{10}$$

The scheme corresponds to a single-band classifier if only visual or infrared features are used.

V. Experimental Results

Datasets: The proposed classification scheme is tested on three datasets.

- *Dataset-1*: a total of 2554 vehicle images in visual band are used. The images containing various types of vehicles are collected from Internet. Detail about the dataset split to each class is given in Table V. Fig.2 shows some example images.

Class#	Vehicle pose	#Visual images
1	Front	515
2	Left	451
3	Right	595
4	Rear	993

TABLE I

Dataset-1: Visual vehicle image dataset containing four poses.

Fig. 2. *Dataset-1*: Example visual vehicle images with four poses.

- *Dataset-2*: a total of 1176 face images in thermal infrared band are used. The images are collected from OTCBVS dataset [15]. Detail about the dataset split to each class is given in Table V. Fig.3 shows some example images.

Class#	Face pose	#IR images
1	Front	350
2	Left	443
3	Right	383

TABLE II

Dataset-2: Thermal infrared face image dataset containing three poses.

- *Dataset-3*: A total of 2422 visual and 2422 thermal infrared images are used. Detail about the dataset split to each class is given in Table V. Fig.4 shows some example images.

Class#	Face pose	#Visual images	#IR images
1	Front	506	506
2	Left	500	500
3	Right	500	500
4	Up	456	456
5	Down	460	460

TABLE III

Dataset-3: Visual and IR face image dataset containing five poses.

Setup: All vehicle and face images are manually cropped and normalized to 48×32 and 32×32 pixels in gray-scale images, respectively. The constant in (9) is $\lambda = 2$. Gabor wavelets with 3 frequency bands (1.5 octave bandwidth) are used for extracting visual and infrared features. The number of orientations is 8 for each image. The down-sampling rate is 4 in each (horizontal/vertical) direction. PCA is applied to Gabor feature vectors retaining average of 95% energy. Images in the dataset are partitioned into 2 sets, i.e. 60% of images in each class are used for training, the remaining 40% are used for testing.

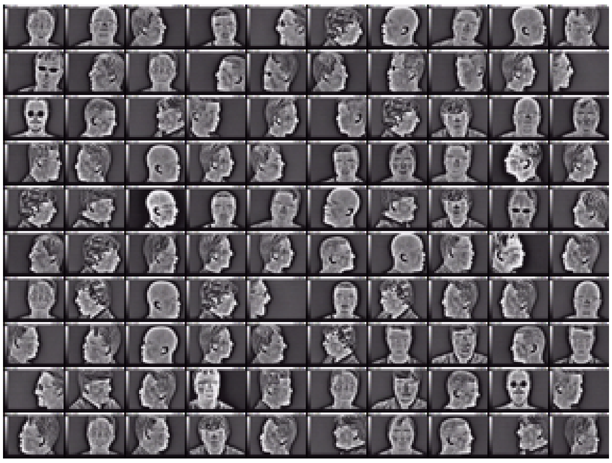

Fig. 3. *Dataset-2*: Example thermal IR face images with three poses.

Fig. 4. *Dataset-3*: Example face images of visual and thermal IR bands with five poses.

Results and comparisons: The performance of the proposed classification and fusion scheme is evaluated according to classification rate, false positive and false negative rate. Observing the results in Table IV and V, one can see that the proposed classifier provides good classification rate while maintaining small false positive and false negative rate on on the testing set of *Dataset-1* and *Dataset-2*.

Method	Dataset	Classification rate (%)
Proposed	*Dataset-1*	**94.18**

Class#	False positive rate (%)	False negative rate (%)
1	2.94	1.43
2	0.55	1.10
3	1.14	0.55
4	1.20	2.75

TABLE IV

Performance: proposed scheme in terms of classification rate, false positive and false negative rate on the testing set of *Dataset-1*.

Method	Dataset	Classification rate (%)
Proposed	*Dataset-2*	**99.86**

Class#	False positive rate (%)	False negative rate (%)
1	0.04	0.09
2	0.03	0.02
3	0.06	0.03

TABLE V

Performance: proposed scheme in terms of classification rate, false positive and false negative rate on the testing set of *Dataset-2*.

Further, comparisons are made with three closely related classification methods on the testing set of *Dataset-3*.

- *Method-1 (M1)*: SAMME using visual images only;
- *Method-2 (M2)*: SAMME using infrared images only;
- *Method-3 (M3)*: fusion of *M1* and *M2* based on confidence.

Method	Dataset	Classification rate (%)
M1	*Dataset-3*(Visual)	87.31
M2	*Dataset-3*(IR)	92.44
M3	*Dataset-3*(Visual+IR)	93.90
Proposed	*Dataset-3*(Visual+IR)	**96.20**

TABLE VI

Performance: proposed scheme and 3 other methods in terms of classification rate on the testing set of *Dataset-3*.

False positive rate (%)					
Method	Front	Left	Right	Up	Down
M1	14.01	10.70	9.55	13.35	16.14
M2	12.18	5.30	3.70	8.30	8.42
M3	12.38	3.00	5.50	4.95	4.35
Proposed	**6.09**	**2.20**	**1.90**	**4.62**	**4.29**
False negative rate (%)					
Method	Front	Left	Right	Up	Down
M1	15.10	7.03	7.28	11.35	22.19
M2	13.15	3.02	1.78	9.64	9.80
M3	9.23	1.52	3.08	8.95	7.85
Proposed	**7.33**	**1.66**	**0.71**	**2.85**	**6.38**

TABLE VII

Performance: proposed scheme in terms of false positive and false negative rate on the testing set of *Dataset-3*.

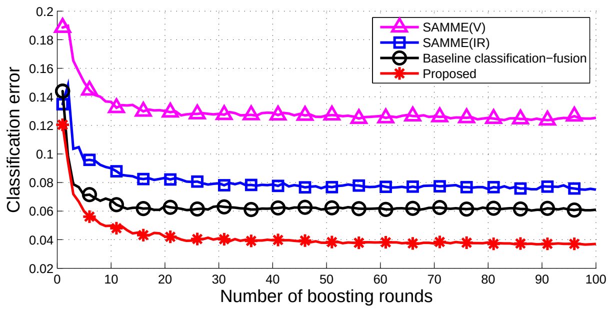

Fig. 5. Classification errors vs. boosting round for the proposed classifier and 3 other classifiers on the testing set.

Results from Table VI and VII show that the proposed classifier improves the average classification rate as comparing with *Method-1*, *Method-2* and *Method-3*. It is observed in Fig.5 that the proposed classifier has a fast convergence speed with the lowest classification errors. Further, Fig.6 shows that using the Gabor feature descriptor for infrared images is very efficient in the proposed classifier. It allows very low dimensional features for infrared images without significantly reducing the final classification rate.

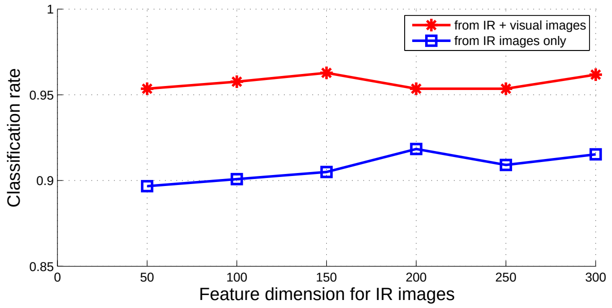

Fig. 6. Dimension of IR image features vs. the average classification rate. Red curve: final classification rate from proposed scheme when the feature dimension of IR images changes meanwhile the feature dimension of visual band (386 in our tests) is fixed; Blue curve: the classification rate when the classifier only uses IR images with specified feature dimension.

VI. Conclusion

The proposed multi-class classification method, using fused hypotheses from visual and infrared information under a unified multi-class AdaBoost framework, is shown to be effective in obtaining high classification rate with low false alarm in our experiments. Our results have also shown that the proposed feature descriptor for infrared images is very effective. Comparison with an existing and closely related AdaBoost algorithm SAMME on visual or infrared face image dataset alone as well as a baseline classification-fusion algorithm has provided further evidence on the effectiveness of the proposed method. Future work will be conducted on testing on more datasets.

References

[1] E. Murphy-Chutorian and M.M. Trivedi, "Head pose estimation in computer vision: a survey," *IEEE Trans. PAMI*, 31(4):607–626, 2009.

[2] Y. Guo, G. Poulton, J. Li, M. Hedley, and R. Qiao, "Soft margin adaboost for face pose classification," *IEEE Proc. ICASSP*, 2003.

[3] S. Baluja, M. Sahami, and H.A. Rowley, "Efficient face orientation discrimination," *IEEE Proc. ICIP*, 2004.

[4] P. Viola and M. Jones, "Robust real-time object detection," *Int'l J. Computer Vision*, 2001.

[5] C. Huang, H. Ai, B. Hu, and S. Lao, "Boosting nested cascade detector for multi-view face detection," *IAPR Proc. ICPR*, 2004.

[6] R.E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, 37(3):297–336, 1999.

[7] Z. Yang, H. Ai, T. Okamoto, and S. Lao, "Multi-view face pose classification by tree-structured classifier," *IEEE Proc. ICIP*, 2005.

[8] E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: a unifying approach for margin classifiers," *Machine Learning Research*, 1:113–141, 2000.

[9] M. Hanif and U. Ali, "Optimized visual and thermal image fusion for efficient face recognition," *IEEE Proc. Fusion*, 2006.

[10] I. Ulusoy and H. Yuruk, "New method for the fusion of complementary information from infrared and visual images for object detection," *IET Proc. Image Processing*, vol. 5, pp. 36 - 48, 2011.

[11] X. Wang and G. Li, "Fusion algorithm for infrared-visual image sequences," *IEEE Proc. ICIG*, 2011.

[12] V.E. Neagoe, A.D. Ropot, and A.C. Mugioiu, "Real time face recognition using decision fusion of neural classifiers in the visible and thermal infrared spectrum," *IEEE Proc. AVSS*, 2007.

[13] A. Apatean, C. Rusu, A. Rogozan, and A. Bensrhair, "Visible-infrared fusion in the frame of an obstacle recognition system," *IEEE Proc. AQTR*, 2010.

[14] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class adaboost," *Statistics and Its Interface*, 2:349–360, 2009.

[15] R. Miezianko, "Terravic Research Infrared Database," *IEEE OTCBVS WS Series Bench*, http://www.cse.ohio-state.edu/otcbvs-bench/.