



CHALMERS

Chalmers Publication Library

Multi-view Face Pose Classification by Boosting with Weak Hypothesis Fusion Using Visual and Infrared Images

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012. Kyoto, 25 - 30 March 2012 (ISSN: 1520-6149)

Citation for the published paper:

Yun, Y. ; Gu, I. (2012) "Multi-view Face Pose Classification by Boosting with Weak Hypothesis Fusion Using Visual and Infrared Images". 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012. Kyoto, 25 - 30 March 2012 pp. 1949-1952.

<http://dx.doi.org/10.1109/ICASSP.2012.6288287>

Downloaded from: <http://publications.lib.chalmers.se/publication/154946>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

MULTI-VIEW FACE POSE CLASSIFICATION BY BOOSTING WITH WEAK HYPOTHESIS FUSION USING VISUAL AND INFRARED IMAGES

Yixiao Yun, Irene Y.H. Gu

Dept. of Signals and Systems, Chalmers University of Technology, Sweden
{yixiao, irenegu}@chalmers.se

ABSTRACT

This paper proposes a novel method for multi-view face pose classification through sequential learning and sensor fusion. The basic idea is to use face images observed in visual and thermal infrared (IR) bands, with the same sampling weight in a multi-class boosting structure. The main contribution of this paper is a multi-class AdaBoost classification framework where information obtained from visual and infrared bands interactively complement each other. This is achieved by learning weak hypothesis for visual and IR band independently and then fusing the optimized hypothesis sub-ensembles. In addition, an effective feature descriptor is introduced to thermal IR images. Experiments are conducted on a visual and thermal IR image dataset 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.

Index Terms— multi-class AdaBoost, weak hypothesis fusion, sub-ensemble learning, visual and infrared images, sequential learning

1. INTRODUCTION

Multi-view face pose classification has drawn increasing research interest in recent years, largely driven by many applications such as robotic surveillance [1], monitoring of driver attentiveness [2] or automating camera management [3].

Several face pose classification methods have been proposed and developed recently. [4] uses PCA-based face features and soft margin AdaBoost to detect the frontal views. [5] extracts features inspired by [6] and builds five separate AdaBoost classifiers for face images in each class. [7] presents a nested cascade detector for face poses in 5 classes using confidence-rated AdaBoost [8] based on Haar features. [9] introduces a tree-structured classifier for face poses in 7 classes, and each node is a three-class classifier trained by AdaBoost.MH. [10] suggests a subspace learning approach for feature extraction and classifies five different face poses by k-NN technique. Good results have been achieved, however, these methods mainly adopt one-against-all or one-

against-one strategies 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. [11], [12] and [13] present fusion methods at the sensor level. [14] uses decision fusion of neural classifiers for real time face recognition. [15] 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 face 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 within a boosting framework for multi-view face pose classification. Different from one-against-all or one-against-one strategies, our model is similar to SAMME [16] 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 image. 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 2 gives a big picture of the proposed framework; Section 3 makes some review of AdaBoost algorithms; Section 4 describes our fusion strategy; Section 5 describes feature extraction for thermal IR images; Section 6 shows experiment results on a visual and thermal IR image dataset and comparisons with most relevant existing method; finally Section 7 concludes the paper.

2. PROBLEM FORMULATION: THE BIG PICTURE

As shown in Fig.1, the proposed framework consists of three major parts: (a) independent weak 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 sam-

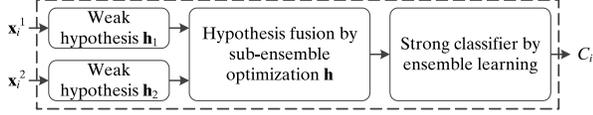


Fig. 1. Block diagram of proposed scheme. The dashed box represents boosting structure. The notations \mathbf{x}_i^1 , \mathbf{x}_i^2 , C_i denote visual features, infrared features, and predicted class labels of i -th object, respectively.

pling 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 weak 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 within multi-class boosting framework, by using visual and infrared information in this interactive manner, which may lead to better classification results.

3. ADABOOST: REVIEW

This section briefly reviews AdaBoost algorithms, with emphasis on SAMME, which our proposed classification method 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 [17]. 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 $\mathbf{X} = \{\mathbf{x}_i\}$, $i = 1, 2, \dots, N$ be the entire training set containing feature vectors of objects. Let the class label (denoted by c) be represented as a K -dimensional vector $\mathbf{y} = (y_1, y_2, \dots, y_K)^T$, where $y_k = 1$ if $c = k$, otherwise $y_k = -1/(K-1)$, $k \in \{1, 2, \dots, K\}$ 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 is encoded in the same way as the weak hypothesis $\mathbf{h} = (h_1, h_2, \dots, h_K)^T$.

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

$$\mathbf{H}^{(t)}(\mathbf{x}_i) = \mathbf{H}^{(t-1)}(\mathbf{x}_i) + \alpha^{(t)} \mathbf{h}^{(t)}(\mathbf{x}_i) \quad (1)$$

subject to the constraint $\sum_{k=1}^K H_k(\mathbf{x}_i) = 0$. Several boosting rounds $t = 1, \dots, T$ 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:

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

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

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

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

4. MULTI-CLASS BOOSTING WITH WEAK HYPOTHESIS FUSION

A sub-ensemble learning method fusing weak hypotheses learned from visual and infrared features under multi-class AdaBoost framework is introduced in this section. Each object feature vector \mathbf{x}_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 weak 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

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

through learning a sub-ensemble of weak hypotheses

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

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

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

where

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

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

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

$$\gamma_i^{(t)} = 2^{\eta_i} \quad (9)$$

where $\eta_i = \sum_{m=1}^M \mathbb{I}(\mathbf{h}_m^{(t)}(\mathbf{x}_i^m) \neq \mathbf{y}_i)$. 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:

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

Table 1 summarizes the pseudo code of the proposed scheme.

Training process:

Input: training set \mathbf{X} , label set \mathbf{Y} and #iteration T .

1. **Initialization:** sampling weights $D_i^{(0)} = 1/N, i = 1, 2, \dots, N$ and ensemble $\mathbf{H}^{(0)}(\mathbf{x}_i) = \mathbf{0} \in \mathbb{R}^K$.
2. **For** $t = 1$ to T (boosting round):
 - (a) Learn single-band weak hypothesis $\mathbf{h}_m^{(t)}(\mathbf{x}_i^m)$ with $D_i^{(t-1)}$;
 - (b) Compute error rate $\epsilon_m^{(t)}$ for each single-band weak hypothesis by (8);
 - (c) Set sub-ensemble weight $\beta_m^{(t)}$ by (7);
 - (d) Fuse optimized hypothesis sub-ensemble $\mathbf{h}^{(t)}(\mathbf{x}_i)$ by (6);
 - (e) Compute weighted error for fused hypothesis $\epsilon^{(t)}$ by (3);
 - (f) Set ensemble weight for fused hypothesis $\alpha^{(t)}$ by (4);
 - (g) Add to ensemble $\mathbf{H}^{(t)}$ according to (1);
 - (h) Update sampling weights $D_i^{(t)}$ by (10) and re-normalize, where the scale factor $\gamma_i^{(t)}$ is obtained from (9);

End

Output: parameters of trained classifier $\mathbf{h}_m^{(t)}, \beta_m^{(t)}, \alpha^{(t)}$.

Test Process:

Input: a new pair of test images \mathbf{x}_j in the test set

Repeat: the boosting round in Step 2 above, using fixed classifier's parameters obtained from the training process;

Output: class label $k^* = \arg \max_k H_k(\mathbf{x}_j)$ for \mathbf{x}_j .

Table 1. Pseudo code of multi-class boosting with weak hypothesis fusion.

5. FEATURE DESCRIPTOR FOR IR IMAGE

Thermal infrared images present different characteristics from those in visual band images, e.g. blurred edges and lack of texture information (as shown in Fig.3). Viewing the

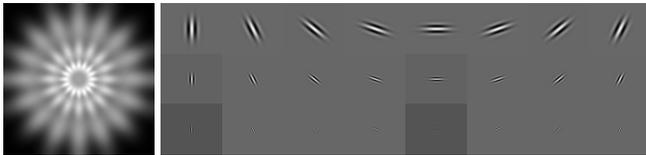


Fig. 2. A bank of Gabor wavelets in (a) frequency domain and (b) real parts in spatial domain.

special nature of thermal IR images, special feature descriptors that are effective should be explored. We propose in this paper to use Gabor wavelet features. The idea here is that a bank of Gabor wavelets with appropriately specified frequency bands and orientations is used to characterize an IR image, which may extract salient features in thermal IR images due to the spatial locality, frequency selectivity and orientation selectivity [18]. DC component is added as a feature component covering the lower frequency band. Fig.2 shows the Gabor wavelets in frequency domain and real part

in spatial domain. To further reducing the feature dimension, we then apply PCA (principal component analysis) to Gabor features from each IR image.

6. EXPERIMENTAL RESULTS

Dataset: A total of 2422 visual and 2422 thermal infrared images are used. Detail about the dataset split to each class is given in Table 2. Fig.3 shows some example images.

Face pose	#Visual images	#IR images
Front	506	506
Left	500	500
Right	500	500
Up	456	456
Down	460	460

Table 2. Visual and thermal IR face image dataset containing five poses.



Fig. 3. Example face images of visual and thermal IR bands with five poses.

Setup: All face images are manually cropped and normalized to 32×32 pixels in gray-scale images. 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.

Results and comparisons: Table 3 and 4 show the classification results from the proposed scheme on the testing set by using visual and thermal IR images as compared with (a) SAMME using visual images only; (b) SAMME using infrared images only; (c) a simple baseline classification-fusion algorithm. The simple baseline classification-fusion algorithm is obtained by training sub-classifiers $\mathbf{H}_m(\mathbf{x}_i^m)$ with independent sampling weights for visual features \mathbf{x}_i^1 and IR features \mathbf{x}_i^2 . The class label is then determined according to $k^* = \arg \max_{m,k} H_{m,k}(\mathbf{x}_i^m)$, where $H_{m,k}(\mathbf{x}_i^m)$ is the k -th element of sub-classifier $\mathbf{H}_m(\mathbf{x}_i^m)$. Fig.3 shows the curves of the corresponding classification error as a function of boosting rounds for all these four cases on the testing set. Results from Table 3 and 4 show that the proposed classifier improves the average classification rate as comparing with

Method	Dataset	Classification rate (%)
SAMME(V)	Visual	87.31
SAMME(IR)	IR	92.44
Baseline classification-fusion	Visual+IR	93.90
Proposed	Visual+IR	96.20

Table 3. Comparison of different methods: average classification rate on the testing set (V: Visual).

Method	False positive rate (%)				
	Front	Left	Right	Up	Down
SAMME(V)	14.01	10.70	9.55	13.35	16.14
SAMME(IR)	12.18	5.30	3.70	8.30	8.42
Baseline classification-fusion	12.38	3.00	5.50	4.95	4.35
Proposed	6.09	2.20	1.90	4.62	4.29

Method	False negative rate (%)				
	Front	Left	Right	Up	Down
SAMME(V)	15.10	7.03	7.28	11.35	22.19
SAMME(IR)	13.15	3.02	1.78	9.64	9.80
Baseline classification-fusion	9.23	1.52	3.08	8.95	7.85
Proposed	7.33	1.66	0.71	2.85	6.38

Table 4. Comparison of different methods: false positive rate and false negative rate for each class on the testing set.

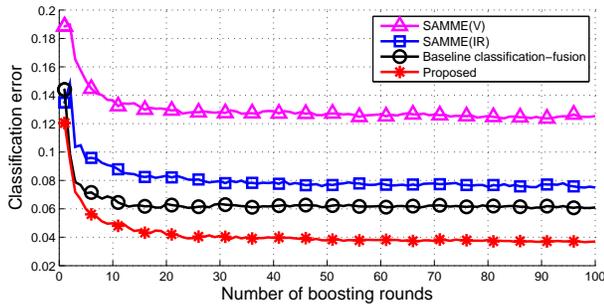


Fig. 4. Classification errors vs. boosting round for the proposed classifier and 3 other classifiers on the testing set.

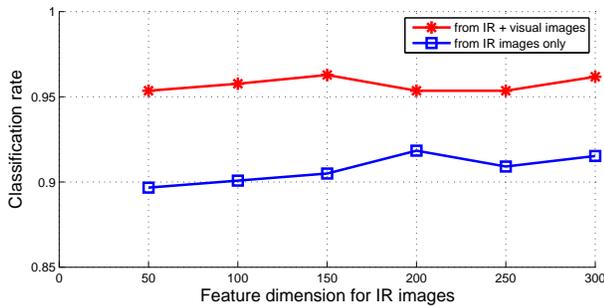


Fig. 5. Dimension of IR image features vs. the average classification rate. Red curve: final classification rate from proposed scheme when the feature dimension of IR images changes meanwhile the feature dimension of visual band (386 in our tests) is fixed; Blue curve: the classification rate when the classifier only uses IR images with specified feature dimension.

SAMME(V), SAMME(IR) and the baseline fusion-classifier. Observing Fig.4 shows that the proposed classifier has a fast convergence speed with the lowest classification errors. Further, Fig.5 shows that using the Gabor feature descriptor for IR images is very efficient in the proposed classifier. It allows very low dimensional features for IR images without significantly reducing the final classification rate.

7. CONCLUSION

The proposed multi-class classification method, using fused hypotheses from visual and IR information in a unified multi-class AdaBoost, 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 IR images is very effective. Comparison with an existing and most relevant AdaBoost algorithm SAMME on visual or IR 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.

8. REFERENCES

- [1] K.B.J. Axnick and R. Jarvis, "Face and pose recognition for robotic surveillance", *Proc. Australian Conf Robotics and Automation*, 2005.
- [2] X. Liu, Y. Zhu, and K. Fujimura, "Real time pose classification for driver monitoring", *Proc. IEEE Int'l Conf Intelligent Transportation Systems*, pp. 174 - 178, 2002.
- [3] Q. Liu, Y. Rui, A. Gupta, and J.J. Cadiz, "Automating camera management for lecture room environment", *Proceedings of International Conference on Computer Human Interaction, ACM*, 2000.
- [4] Y. Guo, G. Poulton, J. Li, M. Hedley, and R. Qiao, "Soft margin adaboost for face pose classification", *Proc. ICASSP*, vol. 3, pp. 221 - 224, 2003.
- [5] S. Baluja, M. Sahami, and H.A. Rowley, "Efficient face orientation discrimination", *IEEE Int'l Conf ICIP*, vol. 1, pp. 589 - 592, 2004.
- [6] P. Viola and M. Jones, "Robust real-time object detection", *Int'l Journal of Computer Vision*, 2001.
- [7] C. Huang, H. Ai, B. Hu, and S. Lao, "Boosting nested cascade detector for multi-view face detection", *Proc IEEE Int'l Conf on Pattern Recognition*, vol. 2, pp. 415 - 418, 2004.
- [8] R.E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions", *Machine Learning*, vol. 37(3), pp. 297 - 336, 1999.
- [9] Z. Yang, H. Ai, T. Okamoto, and S. Lao, "Multi-view face pose classification by tree-structured classifier", *IEEE Int'l Conf on Image Processing*, vol. 2, pp. 358 - 361, 2005.
- [10] E. Islam, A. Khan, and I. Kim, "Effective face pose classification method", *IEEE Int'l Conf Computer, Control and Communication*, pp. 1 - 6, 2009.
- [11] M. Hanif and U. Ali, "Optimized visual and thermal image fusion for efficient face recognition", *IEEE Int'l Conf on Information Fusion*, pp. 1 - 6, 2006.
- [12] I. Ulusoy and H. Yuruk, "New method for the fusion of complementary information from infrared and visual images for object detection", *IET Image Processing*, vol. 5, pp. 36 - 48, 2011.
- [13] X. Wang and G. Li, "Fusion algorithm for infrared-visual image sequences", *IEEE Int'l Conf on Image and Graphics*, pp. 244 - 248, 2011.
- [14] 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 Int'l Conf on Advanced Video and Signal Based Surveillance*, pp. 301 - 306, 2007.
- [15] A. Apatean, C. Rusu, A. Rogozan, and A. Benserhair, "Visible-infrared fusion in the frame of an obstacle recognition system", *IEEE International Conference on Automation Quality and Testing Robotics*, vol. 1, pp. 1 - 6, 2010.
- [16] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class adaboost", *Statistics and Its Interface*, vol. 2, pp. 349 - 360, 2009.
- [17] E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: a unifying approach for margin classifiers", *Machine Learning Research*, vol. 1, pp. 113 - 141, 2000.
- [18] T.S. Lee, "Image representation using 2d gabor wavelets", *IEEE Trans on PAMI*, vol. 18, pp. 959 - 971, 1996.