CHALMERS

Multi-class ada-boost classification of object poses through visual and infrared image information fusion

Proceedings - International Conference on Pattern Recognition, 21st ICPR 2012, Tsukuba,11-15 November 2012 (ISSN: 1051-4651)

(article starts on next page)

21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Multi-Class Ada-Boost Classification of Object Poses through Visual and Infrared Image Information Fusion

Mohamed H. Changrampadi Yixiao Yun Irene Y.H. Gu
Dept. of Signals and Systems, Chalmers University of Technology, Gothenburg, 41296, Sweden

Abstract

This paper presents a novel method for pose classification using fusion of visual and thermal infrared(IR) images. We propose a novel tree structure multi-class classification scheme with visual and IR sub-classifiers. These sub-classifiers are different from the conventional one-against-all or one-against-one strategies, where we handle the multi-class problem directly. We propose to use an accuracy score for the fusion of visual and IR sub-classifiers. In addition, we propose to use the original Haar features plus an extra one, and a multi-threshold weak learner to obtain weak hypothesis. The experimental results on a visual and IR image dataset containing 3018 face images in three poses show that the proposed classifier achieves high classification rate of 99.50% on the test set. Comparisons are made to a fused one-vs-all method, a classifier with visual band only, and a classifier with IR band only. Results provide further support to the proposed method.

1 Introduction

In recent years, many studies on object pose classification have been reported [1]. [2] introduces a method for classifying face poses in three classes using SVM. [3] extends the face detection framework in [4] to handle profile views and rotated faces using AdaBoost classifier and Haar features. [5] extracts face features inspired by [4] and builds five separate AdaBoost classifiers for each class. Though promising results have been obtained, these methods generally follow one-against-all or one-against-one strategy that is a compromised solution to the multi-class problem.

Other work employs visual and IR information for object classification. [6] proposes an illumination-invariant algorithm by fusing visual and IR images. [7] proposes an IR and visible image fusion at pixel level. [8] applies fusion scheme at different levels for SVM-based classification of obstacles. [9] applies visual-IR fusion for pose classification within a multi-class boosting frame. Despite these promising work, classifying face poses remains a challenging issue.

Motivated by this, we propose a novel approach by fusing visual and IR information for face pose classification. Different from one-against-all or one-against-one strategies, our model uses a tree structure of multi-class sub-classifiers, each of which handles a multi-class problem directly. The main contribution of this paper include (a) using a tree-structure classifier where each end node is a K-class classifier with emphasis on each individual class, trained by AdaBoost.M1 (on either visual or IR images); (b) fusing visual and IR sub-classifiers based on their accuracy scores. Very high classification rate are obtained. Comparisons with single visual / IR classifier and one-against-all classifier with fusion are also conducted.

The rest of this paper is organized as follows: Section 2 gives a big picture of the proposed scheme; Section 3 reviews Haar features (together with a newly proposed feature component) and the basic AdaBoost algorithm; Section 4 describes the proposed method; Section 5 shows experiment results on a visual and IR image dataset and comparisons with two most relevant existing methods. Finally Section 6 concludes the paper.

2 The big picture

As shown in the block diagram of Fig.1, the proposed scheme consists of two level-1 sub-classifiers independently trained by visual and IR images.These two sub-classifiers are then fused based on the proposed accuracy scores. For each visual or IR band, the classifier has K AdaBoost level-2 sub-classifiers, each of which is a multi-class classifier. This method is different from one-against-all or one-against-one strategies, as our method does not decompose the multi-class problem to multiple binary sub-problems. It employs multiple sub-classifiers, each solving the multi-class problem with emphasis on one particular class. To achieve this, a criterion based on the sum of false-positive and false-negative rate for each single class is proposed to find the weak hypothesis in each AdaBoost sub-classifier. Fusion is then performed based on the accuracy scores of visual/IR sub-classifiers.

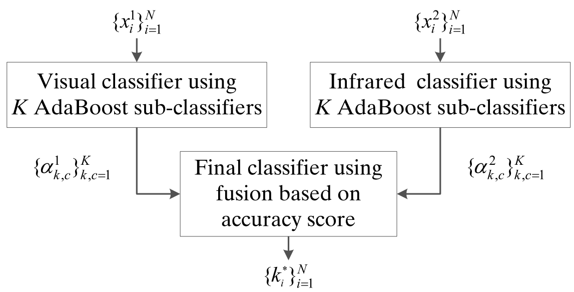

Figure 1. Block diagram of proposed scheme. The notations x_i^1, x_i^2, $\alpha_{k,c}^1$, $\alpha_{k,c}^2$, k_i^* denote visual images, infrared images, the accuracy scores of each sub-classifiers emphasizing k-th class in visual / IR band, and predicted class labels of i-th object, respectively.

The accuracy score is associated with a confidence or reliability measure, so the visual and IR information may compensate each other, thus resulting in improved classification performance.

3 Haar Feature and Adaboost

3.1 Haar features

Haar features, or Haar-like features, are simple image features based on intensity differences between rectangle-based regions (as shown in Fig.2a) that share similar shapes to the Haar wavelets. Haar features computed using integral images are computationally inexpensive as compared to statistics based features. Haar features are initially proposed by [4], where a window of the target size slides over an input image, and the Haar feature is calculated for each subsection of image. This difference is then compared to a learned threshold that separates non-objects from objects. In our study, an additional Haar feature component (Fig.2b) is added to describe the object (face).

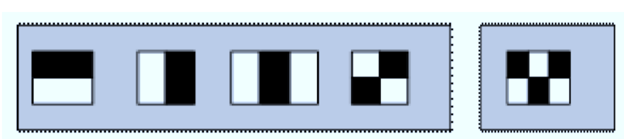

Figure 2. Left: (a) Haar features; Right: (b) a newly added one in our classifier.

3.2 AdaBoost: Review

This part briefly reviews the basic AdaBoost algorithm, with focus on AdaBoost.M1 that our sub-classifiers are built upon. AdaBoost is an ensemble learning method originally designed for binary classes. Many alternatives of AdaBoost for multi-class problems exist, and most of them are restricted to using one-against-all or one-against-one strategies [10]. AdaBoost.M1 [11], one of the multi-class AdaBoost algorithms, is an trivial extension of AdaBoost that solves multi-class problems without reducing them to multiple binary subproblems. It requires the weak classifiers to be multi-class, each having an accuracy greater than 50%.

Let $\boldsymbol{S} = \{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_N, y_N)\}$ be the training set where $\boldsymbol{x}_i$ is the image feature vector and $y_i \in \boldsymbol{Y} = \{1, ..., K\}$ (K: the number of classes) is the class label associated with $\boldsymbol{x}_i$. The goal is to minimize the loss function $L(y, H) = \sum_{i=1}^{N} \mathbb{I}[y_i \neq H(\boldsymbol{x}_i)]$ by learning the classifier:

$$H(\boldsymbol{x}_i) = \arg\max_{y \in \boldsymbol{Y}} \sum_{t: h^{(t)}(\boldsymbol{x}_i) = y} \alpha^{(t)} \tag{1}$$

where $t = 1, ..., T$ is the boosting iteration and $h^{(t)}(\boldsymbol{x}_i) \in \boldsymbol{Y}$ is each weak hypothesis.
In each boosting round, sampling weight $D_i^{(t)}$ for feature vector of object, weighted classification error $\epsilon^{(t)}$ for weak classifier, and weight of weak classifier $\alpha^{(t)}$ are updated as follows:

$$\begin{aligned} D_i^{(t)} &= D_i^{(t-1)} \exp(\alpha^{(t)} \mathbb{I}[y_i \neq h^{(t)}(\boldsymbol{x}_i)]) \\ \epsilon^{(t)} &= \textstyle\sum_{i=1}^{N} D_i^{(t-1)} \mathbb{I}[y_i \neq h^{(t)}(\boldsymbol{x}_i)] / \sum_{i=1}^{N} D_i^{(t-1)} \\ \alpha^{(t)} &= \log \frac{1-\epsilon^{(t)}}{\epsilon^{(t)}} \end{aligned} \tag{2}$$

where $\mathbb{I}[A]$ is an indicator function that equals 1 if event A is true, and 0 otherwise.

4 Proposed method

A tree structure classifier is described in this section. As shown in Fig.3 the final classifier is formed by fusing visual and IR sub-classifiers in the level-1. Each of these sub-classifiers consists of K multi-class AdaBoost sub-classifiers at level-2. In our case, each object feature vector $\boldsymbol{x}_i$ contains two component feature vectors $\{\boldsymbol{x}_i^1, \boldsymbol{x}_i^2\}$, corresponding to visual and IR bands, respectively. The level 2 sub-classifier is trained to emphasize on k^{th} class.

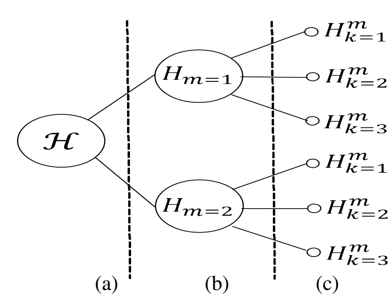

Figure 3. Tree structure of the proposed classification scheme. (a) final classifier with fusion; (b) For visual, IR level-1 sub-classifiers; (c) level-2 sub-classifiers.

4.1 Visual / IR sub-classifier using multiple multi-class AdaBoost

For each visual or IR image set, we train a classifier based on one type (visual/IR images). Our goal for each classifier ($m = 1, 2$) is to minimize the loss function $L(y, H_m) = \sum_{i=1}^{N} \mathbb{I}[y_i \neq H_m(\boldsymbol{x}_i^m)]$ by learning an ensemble:

$$H_m(\boldsymbol{x}_i) = \arg\max_{y \in \boldsymbol{Y}} \sum_{c:H_{k,c}^m(\boldsymbol{x}_i^m)=y} \sum_{k=1}^{K} \alpha_{k,c}^m \quad (3)$$

where $H_{k,c}^m(\boldsymbol{x}_i^m) \in \boldsymbol{Y}$ is the c-th component of the hypothesis made by the AdaBoost sub-classifier emphasizing on the k-th class, and $\alpha_{k,c}^m$ is its corresponding accuracy score as $\sum_{t=1}^{T} \beta_{k,c}^{m,(t)}$ where $\beta_{k,c}^{m,(t)}$ is obtained by (5), T is the number of boosting rounds, and $c = 1, ..., K$. The minimization problem is solved by a stagewise approach. In each step $k = 1, ..., K$, the k-th AdaBoost sub-classifier is trained by enforcing its weak classifiers to minimize the weighted sum of related false-positive and false-negative rate in k-th class:

$$h_k^{m,(t)} = \arg\min_{h^{(t)}} e_k^{m,(t)} \quad (4)$$

where $e_k^{m,(t)} = \sum_{i=1}^{N} D_i^{m,(t-1)} \eta_{k,i}^{m,(t)}$ and $\eta_{k,i}^{m,(t)} = f_{p_{(k,i)}} + f_{n_{(k,i)}}$, and

$$\begin{cases} f_{n_{(k,i)}} = \mathbb{I}[y_i = k \mid h^{m,(t)}(\boldsymbol{x}_i^m) \neq k] \\ f_{p_{(k,i)}} = \mathbb{I}[y_i \neq k \mid h^{m,(t)}(\boldsymbol{x}_i^m) = k] \end{cases}$$

Then the weight of weak classifier with respect to class c ($c = 1, ..., K$) is obtained by

$$\beta_{k,c}^{m,(t)} = \log \frac{1 - \epsilon_{k,c}^{m,(t)}}{\epsilon_{k,c}^{m,(t)}} \quad (5)$$

where $\epsilon_{k,c}^{m,(t)}$ is the weighted miss-classification error for c^{th} class:

$$\epsilon_{k,c}^{m,(t)} = \sum_{i:y_i=c} D_i^{m,(t-1)} \mathbb{I}[h_k^{m,(t)}(\boldsymbol{x}_i^m) \neq y_i] \quad (6)$$

Before the next boosting round, the sampling weight $D_i^{m,(t)}$ is updated as follows:

$$D_i^{m,(t)} = D_i^{m,(t-1)} (\lambda_k^{m,(t)})^{(1-\eta_{k,i}^{m,(t)})} \quad (7)$$

where $\lambda_k^{m,(t)} = \log((1 - e_k^{m,(t)})/e_k^{m,(t)})$. In this way, multiple sub-classifiers are trained for multi-class independently, with a confidence related score θ_c^m described in Section 4.2.

4.2 Fusion of visual and IR sub-classifiers based on accuracy scores

The accuracy proposed score for each sub-classifier m, and each class c is defined as

$$\theta_c^m = \sum_{c:H_{k,c}^m(\boldsymbol{x}_i^m)=y} \sum_{k=1}^{K} \alpha_{k,c}^m \qquad m = 1, 2 \quad (8)$$

where $H_{k,c}^m(\boldsymbol{x}_i^m), y \in \boldsymbol{Y}$ and $c = 1, ..., K$. This score is proportional to the sum of confidence levels of associated weak learners.
The final classifier is obtained by fusing the information from two classifiers using weighted hypotheses:

$$\mathcal{H}(\boldsymbol{x}_i) = \arg\max_c \sum_{c=1}^{K} \sum_{m=1}^{M} \theta_c^m \mathbb{I}[H_m(\boldsymbol{x}_i^m) = c] \quad (9)$$

In our scheme, the fusion process is performed in the decision level. Table 1 summarizes the algorithm of our proposed scheme.

TRAINING PROCESS:
Input: training set $\boldsymbol{S}$, label set $\boldsymbol{Y}$ and $\#$iteration T.
Initialization: sampling weights $D_i^{(0)} = 1/N, i = 1, 2, ..., N$.
For $m = 1$ to M (visual / IR band):
  For $k = 1$ to K (each sub-classifier):
    For $t = 1$ to T (boosting round):

1. Learn weak hypothesis $h_k^{m,(t)}(\boldsymbol{x}_i^m)$ with $D_i^{(t-1)}$ by (4);
2. Compute weighted misclassification errors $\epsilon_{k,c}^{m,(t)}$ for each class by (6);
3. Set weak classifier weights $\beta_{k,c}^{m,(t)}$ by (5);
4. Update sampling weights $D_i^{(t)}$ by (7) and re-normalize;

    End{t}
  End{k}
End{m}
Output: parameters of trained classifier $h_k^{m,(t)}$, $\beta_{k,c}^{m,(t)}$, $\alpha_{k,c}^m$, and θ_c^m by (8).

TEST PROCESS:
Input: a new pair of test images $\boldsymbol{x}_j$ in the test set.
Repeat: the boosting round in Step 2 above, using fixed classifier's parameters obtained from the training process.
Output: final hypothesis of class label $\mathcal{H}(\boldsymbol{x}_j)$ by fusion in (9).

Table 1. Pseudo codes of the proposed scheme.

5 Experiments and Results

Dataset and experiment setup: The database comprises visual and IR images of human face poses in three classes (front, left and right). The images were captured in uncontrolled environments with large variations, e.g., lighting variations, different background. All images were cropped manually to remove the background and irrelevant information. The cropped images are then normalized to a fixed size before processing. Example images of our dataset are shown in Fig.4.

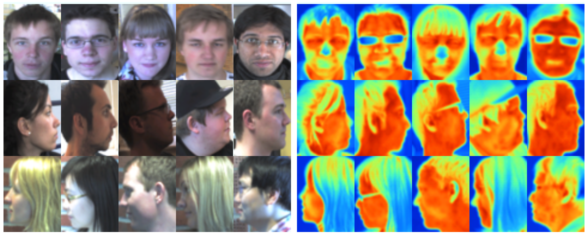

Figure 4. Example images of three poses. Left: visual images; Right: IR images.

Three classes ($K = 3$) in front, left and right, each with 500 images are used in our experiment. 60% images in each class are used for training, the remaining 40% are used for testing. Five Haar feature types shown in Fig.2 are used. For a 19×19 detector window, 35686 features are generated. Each AdaBoost sub-classifier is trained for $T = 50$ boosting rounds.

Results and comparisons: Table 2 and 3 show the classification results of the proposed scheme on the test set. The proposed method is compared with the classifier that uses visual images only and classifier that uses IR images only. Further, the method is compared with the one-against-all classifier using both visual and IR images. Fig.5 shows the classification error as a function of boosting rounds for these four classifiers. Observing the figure, the proposed classifier has shown the highest classification rate and the lowest classification error.

Method	Dataset	Classification rate (%)
Visual (P)	Visual	91.67
Infrared (P)	IR	98.67
Fused one-vs-all	Visual+IR	98.67
Proposed	Visual+IR	99.50

Table 2. Comparison of different methods: average classification rate on the test set. (P) refers to the proposed method

False positive rate (%)			
Method	Front	Left	Right
Visual (P)	2.96	5.36	4.66
Infrared (P)	0.51	0.75	0.75
Fused one-vs-all	1.96	0	0
Proposed	0.5	0.25	0
False negative rate (%)			
Method	Front	Left	Right
Visual (P)	17.11	13.94	14.95
Infrared (P)	2.94	2.48	2.48
Fused one-vs-all	0	3.85	3.85
Proposed	0.5	1.0	1.48

Table 3. Comparison of different methods: false positive rate and false negative rate for each class on the test set.

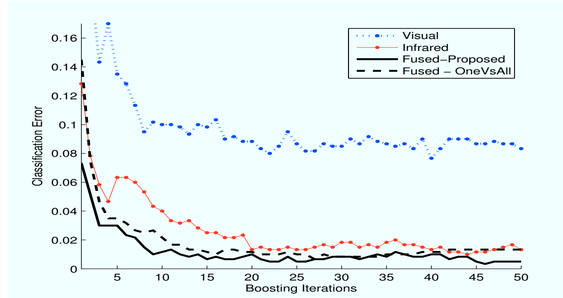

Figure 5. Classification errors vs. boosting round for the proposed classifier, single visual / IR classifier and fused one-vs-all on the test set.

6 Conclusion

The proposed tree structure classifier with multiple multi-class AdaBoost and the accuracy score-based fusion is tested on a visual and IR face images. The experiments from the proposed scheme have yielded high classification rate (99.50%) and low false alarm rate (1.48%). Comparing with the visual-only or IR-only classifier, the proposed method has achieved better classification rate with lower false alarm. The proposed method is also shown to have improved performance than the one-against-all classifier using both visual and IR images. Future work will be extended to test larger dataset and more classes.

References

[1] E. Murphy-Chutorian, M.M. Trivedi, "Head pose estimation in computer vision: A Survey", *IEEE Trans. PAMI*, vol. 31(4), pp.607 - 626, 2009.

[2] J. Huang, X. Shao, H. Wechsler, "Face pose discrimination using Support Vector Machines (SVM)", *Proc. 14th Int'l Conf. Pattern Recognition*, pp.154 - 156, 1998.

[3] M. Jones, P. Viola, "Fast multi-view face detection", *Technical Report 096, Mitsubishi Electric Research Laboratories*, 2003.

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

[5] S. Baluja, et al, "Efficient face orientation discrimination", *IEEE Int'l Conf ICIP*, vol.1, pp.589-592, 2004.

[6] S. G. Kong, J. Heo, et al., "Adaptive fusion of visual and thermal IR images for illumination-invariant face recognition", *Int'l Journal of Comput. Vis.*, 71(2), pp.215-233, 2007.

[7] G. Bebis, et al., "Face recognition by fusing thermal infrared and visible imagery", *Image and Vis. Computing*, 24(7), pp.727-742, 2006.

[8] A. Apatean, C. Rusu, et al, "Visible-infrared fusion in the frame of an obstacle recognition system", *IEEE Int'l Conf. Automation Quality and Testing Robotics*, 2010.

[9] Y. Yun, I.Y.H. Gu, "Multi-view face classification by boosting with weak hypothesis fusion using visual and infrared images", *IEEE Int'l Conf. ICASSP*, 2012.

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

[11] Y. Freund, R.E. Schapire, "Experiments with a new boosting algorithm", *AT&T Research*, 1996.