Pattern Recognition 52 (2016) 1–16

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A design framework for hierarchical ensemble of multiple feature extractors and multiple classifiers

Kyounghoon Kim [a,b], Helin Lin [a,c], Jin Young Choi [a], Kiyoung Choi [a,*]

[a] Department of Electrical Engineering and Computer Science, Seoul National University, 151-744 Seoul, Republic of Korea
[b] DMC R&D Center, Samsung Electronics, 443-742 Suwon, Republic of Korea
[c] System LSI, Samsung Electronics, 446-711 Giheung, Republic of Korea

ARTICLE INFO

Article history:
Received 2 May 2014
Received in revised form
17 August 2015
Accepted 10 November 2015
Available online 30 November 2015

Keywords:
Ensemble of detection systems
Multiple feature extractors
Multiple classifiers
Pedestrian detection
Reinforcement learning
Bayesian network

ABSTRACT

It is well-known that ensemble of classifiers can achieve higher accuracy compared to a single classifier system. This paper pays attention to ensemble systems consisting of multiple feature extractors and multiple classifiers (MFMC). However, MFMC increases the system complexity dramatically, leading to a highly complex combinatorial optimization problem. In order to overcome the complexity while exploiting the diversity of MFMC, we suggest in this paper a hierarchical ensemble of MFMC and its optimizing framework. By constructing local groups of feature extractors and classifiers and then combining them as a global group, the approach achieves a better scalability. Both reinforcement machine learning and Bayesian networks are adopted to enhance the accuracy. We apply the proposed method to vision based pedestrian detection and recognition of handwritten numerals. Experimental results show that the proposed framework outperforms the previous ensemble methods in terms of accuracy.

1. Introduction

Ensemble of multiple classifiers is one of promising approaches for constructing a more accurate classifier. It can handle difficult problems where a single classifier easily makes a wrong decision due to lack of training or parameter optimization. Combining the decisions of participating classifiers statistically reduces the risk of wrong decision. In addition, such an ensemble system can generate a sensible solution in a special environment, where several classifiers should be trained with different training datasets due to temporal or spatial constraints. It can also solve instability problems that frequently occur in a single classifier like neural networks with different initial conditions. Another benefit comes from a fact that no single classifier solution can tackle all problems according to the no free lunch theorem (NFL) [1–3]. Due to these advantages, classifier ensemble has been an active research area in the literature of machine learning and pattern recognition [4–6]. According to these researches, an ensemble system generates more stable and accurate results compared to conventional single classifier systems.

Considering multiple feature extractors (FEs) and classifiers used for constructing an ensemble system, classification systems can be classified into three categories according to the number of FEs and classifiers as shown in Fig. 1. Conventional classification systems use a single FE and a single classifier as shown in Fig. 1(a). There are systems that use multiple classifiers sharing the same feature vector generated by an FE as shown in Fig. 1(b). AdaBoost is a well-known machine learning algorithm that supports this model [7]. There have been various researches to extend the concept of AdaBoost for better performance [8,9]. The third category shown in Fig. 1(c) has been introduced [10–12]; it uses multiple FEs as well as multiple classifiers and thus we call it MFMC.

One of most representative applications of classifier ensemble is pedestrian detection, which is a key problem in transportation, surveillance, robotics, entertainment systems, and other systems that need to recognize and interact with human [13–16]. In pedestrian detection, vision based approach is the most effective and popular way. However, it is still quite challenging due to large variations in many aspects such as human clothing, pose, size, background, weather, and illumination. In order to overcome the difficulty, many studies have been conducted in many different ways [17]. However, the achieved accuracy is still insufficient to be used for real applications including advanced driver assistance system (ADAS), thus leaving room for improvement as mentioned by Dollar et al. [18]. Especially, in case that a pedestrian is far from the camera or under partial occlusion, the accuracy degrades dramatically. In order to improve the accuracy or detection rate, many studies have tried to find more effective extractors and classifiers such as those in [19–23] and [24]. The researches have focused on finding good features as well as good classifiers.

* Corresponding author.

http://dx.doi.org/10.1016/j.patcog.2015.11.006
0031-3203/

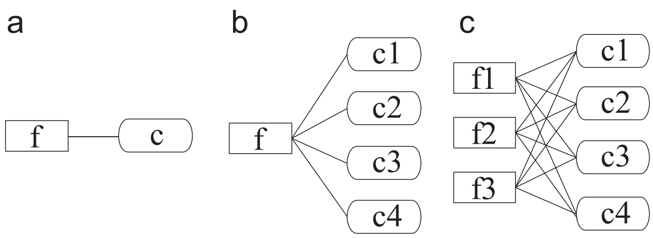

Fig. 1. Example of detection system regarding the number of components. ‘f’ stands for a feature extractor; ‘c’ stands for a classifier. (a) Traditional single feature extractor and single classifier. (b) Single feature extractor and multiple classifiers represented by AdaBoost. (c) Multiple feature extractors and multiple classifiers (MFMC).

Meanwhile, utilizing combination(s) of multiple FEs and classifiers has also been studied; it has a strong advantage compared to single FE, single classifier counterparts.

In this paper, we focus on MFMC since it provides results superior to both single classifier and multiple classifiers with single FE as shown in the previous researches [10–12,25]. Contrary to the previous studies that try to find a manually optimized fixed combination of existing FEs and classifiers, we try to optimize automatically the FEs and classifiers as well as their combinations. In particular, we suggest a novel ensemble framework that can manage the complexity generated from MFMC by using a hierarchical method that integrates reinforcement machine learning and Bayesian network modeling. This paper is organized as follows. Section 2 gives a brief overview of the related work. Section 3 presents the proposed hierarchical ensemble framework for MFMC. Section 4 shows experimental results and Section 5 concludes the paper.

2. Related work

Many ensemble methods have been proposed for a past few decades in the literature of pattern recognition and machine learning. The methods for combining multiple classifiers include weighted majority vote (WMV), naïve Bayes combination (NB) [5], behavior knowledge space (BKS) [26], Wernecke [27], and SVD combination [28]. Moreno-Seco et al. [29] suggested extensions to weighted majority vote such as rescaled weighted vote (RSWV), best-worst weighted vote (BWWV), and quadratic best-worst weighted vote (QBWWV). One of other distinguishing approaches is based on genetic algorithm (GA) [30,31] which use GA to find a better combination without exhaustive search. One study [32] presented an ensemble method with a trained fuser using weights of classifiers, where the optimization problem and solver are proposed. Fuzzy combiner is also used in order to aggregate multiple instances of a classifier for multi-class classification [33]. There are also probabilistic models with posterior estimators [34,35].

However, most of existing ensemble schemes to combine decisions do not consider multiple FEs but consider multiple classifiers with a single FE. For MFMC, only a few ensemble methods were presented. Chen et al. [36] proposed three general methods – linear combination, winner-take-all, and evidential reasoning – to combine multiple classifiers with different features. They applied them to text-independent speaker identification, where the linear combination with different features (LCDF) outperformed the other two methods. LCDF performs linear combination of the decisions from multiple classifiers, where the learning algorithm uses maximum likelihood estimation with the expectation maximization (EM) algorithm. Zenobi et al. [37] presented ensemble of classifiers with different feature subsets considering their diversity. To find the best members of the ensemble, they suggested a hill-climbing algorithm based on the relationship between ensemble accuracy and ambiguity.

There have been a few studies on MFMC in the context of pedestrian detection. Ludwig et al. [10] employed both histogram of oriented gradients (HOG) and covariance matrices (COV) consisting of pixel coordination, derivatives, magnitude, and gradient as FEs. They adopted neural networks (NN) and support vector machines (SVM) as classifiers. They also suggested using an ensemble method called ‘Training of Fusion Algorithm (TFA)’. Oliveira et al. [11] used HOG and local receptive fields (LRF) provided by convolutional neural networks (CNN) as FEs, in which NN and SVM were also employed as classifiers. The work in [12] used HOG and HOF (histogram of optical flow) as FEs and SVM and MPLBoost extended from AdaBoost as classifiers. Those approaches rely on manual optimization of the combinations of MFMC and thus make it hard to have many FEs and classifiers (only two or three of them are allowed). However, our observation is that the FEs and classifiers are complementary with one another, which makes it necessary to combine many of them in order to realize a high accuracy system. Thus the scalability of an ensemble system is important. Moreover, whenever a new FE or a classifier is added, the entire process of finding the best combination will be repeated and thus automating the optimization process is crucial. The aim of this paper is to construct a framework to overcome the limitations of the previous approaches to MFMC. We investigate this challenging problem and suggest efficient methodologies including experiments for the application of pedestrian detection.

In case of ensemble schemes based on weak-learners (or simple prediction rules) such as bagging [38], boosting [39], and error correcting output codes (ECOC) [40–42], a large number of the weak-learners commonly participate in the ensemble (i.e., the number of weak-learners can be several thousands), where even a single feature vector can have its own classifier. They focus on selecting better ones among many weak-learners or reformulating the dimensionality such as principal components analysis (PCA) and linear discriminant analysis (LDA). Whereas, in our approach, we focus on finding the best combination among multiple FEs and classifiers. We use a relatively small number of relatively strong learners; simultaneously using a small number of different FEs (e.g., HOG [24], CENTRIST [43], HAAR [44]) and classifiers (e.g., SVM, KNN, and decision tree [45]) can effectively increase the accuracy in practical applications such as pedestrian detection. In the experimental section, we compare the performance of our scheme to that of AdaBoost, a representative ensemble scheme based on weak-learners.

The objective of our work is not to find just the best performed combination of a few FEs and classifiers for pedestrian detection as the work done in [10–12], but to suggest a general *framework* for designing an ensemble system like the ones in [5,36,29]. Note, however, that our approach allows arbitrary numbers of FEs as well as classifiers, whereas the approaches in [7] are limited to systems with a single FE (although they have multiple classifiers), which are much easier to optimize.

3. Proposed hierarchical ensemble system

As illustrated in Fig. 2, the proposed ensemble system consists of three steps: constructing all possible FE-classifier pairs, building a set of local combinations from the set of pairs using reinforcement machine learning, and making a final decision by constructing a global combination based on Bayesian network. In the first step, each FE generates a feature set in a vector format from an input image. The feature vectors from an FE are used by each classifier pairing with the FE for training and testing, which is identical to conventional approach for creating individual recognizers. In the