Statistical Learning of Multi-view
Face Detection

Stan Z. Li\(^1\), Long Zhu\(^1\), ZhenQiu Zhang\(^2\)*, Andrew Blake\(^3\)
Hongjiang Zhang\(^1\), Harry Shum\(^1\)

\(^1\) Microsoft Research Asia, Beijing, China
\(^2\) Institute of Automation, Chinese Academy Sinica, Beijing, China
\(^3\) Microsoft Research Cambridge, Cambridge, UK
Contact: szli@microsoft.com, http://research.microsoft.com/~szli

Abstract. A new boosting algorithm, called FloatBoost, is proposed to overcome the monotonicity problem of the sequential AdaBoost learning. AdaBoost \([1, 2]\) is a sequential forward search procedure using the greedy selection strategy. The premise offered by the sequential procedure can be broken-down when the monotonicity assumption, \textit{i.e.} that when adding a new feature to the current set, the value of the performance criterion does not decrease, is violated. FloatBoost incorporates the idea of Floating Search \([3]\) into AdaBoost to solve the non-monotonicity problem encountered in the sequential search of AdaBoost.

We then present a system which learns to detect multi-view faces using FloatBoost. The system uses a coarse-to-fine, simple-to-complex architecture called detector-pyramid. FloatBoost learns the component detectors in the pyramid and yields similar or higher classification accuracy than AdaBoost with a smaller number of weak classifiers. This work leads to the first real-time multi-view face detection system in the world. It runs at 200 ms per image of size 320x240 pixels on a Pentium-III CPU of 700 MHz. A live demo will be shown at the conference.

1 Introduction

Pattern recognition problems has two essential issues: (i) feature selection, and (ii) classifier design based on selected features. Boosting is a method which attempts to boost the accuracy of an ensemble of weak classifiers to a strong one. The AdaBoost algorithm \([1]\) solved many of the practical difficulties of earlier boosting algorithms. Each weak classifier is trained one stage-wise to minimize the empirical error in a given distribution re-weighted according classification errors of the previously trained classifier. It is shown that AdaBoost is a sequential forward search procedure using the greedy selection strategy to minimize a certain margin on the training set \([4]\).

A crucial heuristic assumption made in such a sequential forward search procedure is the monotonicity, \textit{i.e.} that when adding a new weak classifier to the

\* The work presented in this paper was carried out at Microsoft Research Asia.
current set, the value of the performance criterion does not decrease. The premise offered by the sequential procedure can be broken-down when the assumption is violated, \textit{i.e.} when the performance criterion function is non-monotonic. This is the first topic to be dealt with in this paper.

Floating Search [3] is a sequential feature selection procedure with backtracking, aimed to deal with non-monotonic criterion functions for feature selection. A straight sequential selection method like sequential forward search (SFS) or sequential backward search (SBS) adds or deletes one feature at a time. To make this work well, the monotonicity property has to be satisfied by the performance criterion function. Feature selection with a non-monotonic criterion may be dealt with by using a more sophisticated technique, called plus-$\ell$-minus-$r$, which adds or deletes $\ell$ features and then backtracks $r$ steps [5, 6].

The Sequential Floating Search methods [3] allows the number of backtracking steps to be controlled instead of being fixed beforehand. Specifically, it adds or deletes $\ell = 1$ feature and then backtracks $r$ steps where $r$ depends on the current situation. It is such a flexibility that amends limitations due to the non-monotonicity problem. Improvement on the quality of selected features is gained with the cost of increased computation due to the extended search. The SFFS algorithm performs very well in several applications [3, 7]. The idea of Floating Search is further developed in [8] by allowing more flexibility for the determination of $\ell$.

The second topic is an application of booting learning in face detection. Learning based methods have so far been the most effective ones for face detection, \textit{e.g.} [9–12]. There, face detection is treated as an intrinsically two-dimensional (2-D) problem. Taking advantage of the fact that faces are highly correlated, it is assumed that human faces can be described by some low dimensional features which may be derived from a set of prototype face images. Large amount of variation and complexity brought about by changes in facial appearance, lighting and expression makes the face manifold highly complex [13, 14]. Changes in facial view (head pose) further complicate the situation. From pattern recognition viewpoint, two issues are essential in face detection: (i) feature selection, and (ii) classifier design based on selected features.

Applied to face detection [15], AdaBoost is adapted to solving the following three fundamental problems in one boosting procedure: (1) learning incrementally crucial features from a large feature set, (2) constructing weak classifiers each of which is based on one of the selected features, and (3) boosting the weak classifiers into a stronger classifier using a linear combination derived during the learning process. The work of Viola and Jones results in the first real-time frontal face detection system which runs at about 14 frame per second for a 320x240 image [15]. This work, like [9–12], deals with frontal faces only.

However, ability to deal with non-frontal faces is important for many real applications because statistics show that approximately 75\% of the faces in home photos are non-frontal [16]. A reasonable treatment for multi-view is the view-based method [17], in which several face models are built, each describing faces in a certain view. This way, explicit 3D modeling is avoided. Feraud \textit{et al.} [18]