Local Inter-Session Variability Modelling for Object Classification

Kaneswaran Anantharajah
QUT, SAIVT and MILAB

ZongYuan Ge
QUT, CyPhy Lab.

Chris McCool
NICTA, Brisbane

Simon Denman
QUT, SAIVT

Clinton Fookes
QUT, SAIVT

Peter Corke
QUT, CyPhy Lab.

Dian Tjondronegoro
QUT, MILAB

Sridha Sridharan
QUT, SAIVT

Abstract

Object classification is plagued by the issue of session variation. Session variation describes any variation that makes one instance of an object look different to another, for instance due to pose or illumination variation. Recent work in the challenging task of face verification has shown that session variability modelling provides a mechanism to overcome some of these limitations. However, for computer vision purposes, it has only been applied in the limited setting of face verification.

In this paper we propose a local region based inter-session variability (ISV) modelling approach, and apply it to challenging real-world data. We propose a region based session variability modelling approach so that local session variations can be modelled, termed Local ISV. We then demonstrate the efficacy of this technique on a challenging real-world fish image database which includes images taken underwater, providing significant real-world session variations. This Local ISV approach provides a relative performance improvement of, on average, 23% *on the challenging MOBIO, Multi-PIE and SCface face databases. It also provides a relative performance improvement of* 35% *on our challenging fish image dataset.*

1. Introduction

Object classification is a challenging problem due to variations in the appearance of the objects and the environment in which they appear. One of the best known and most well investigated object classification problems is that of face recognition, where variations in subject pose and lighting present significant challenges [6]. A recent state-of-the-art face recognition approach uses session variability modelling [12] to provide a general model that describes the differences that occur between instances of the same class, whether that be from pose, illumination or expression variation. This session variability modelling approach is applied in the context of a free-parts model [16], which discards potentially useful spatial relationships.

The free-parts approach described in [16] divides the face into blocks and each block is considered to be a independent observation of the same object (the face). The distribution of these blocks is described by a Gaussian mixture model (GMM) and has been investigated by several researchers [16, 9, 10, 19]. Lucey and Chen [9] showed that a relevance adaptation approach, similar to the one used for speaker authentication [14], could be used to quickly obtain client (class) specific GMMs by using a universal background model (UBM). Furthermore, Lucey and Chen showed that adding spatial constraints to this free-parts approach could yield state-of-the-art face recognition performance on the BANCA dataset [13]. Sanderson et al. [15] proposed a multi-region probabilistic histogram (MRH) approach which used the free-parts approach as its basis but incorporates spatial constraints and also makes several simplifications for efficiency purposes. This efficient method provided state-of-the-art performance on the labeled faces in the wild (LFW) dataset [1].

Recently in [18, 12] the GMM free-parts (GMM-FP) model was extended to include an inter-session variability (ISV) modelling component. ISV learns a sub-space which models the differences in instances of the same object (the face). Such an approach was initially proposed to cope with similar problems in speaker authentication [17]. This model of session variability is used to estimate session variations in order to suppress, or account, for them. Using this model yielded state-of-the-art performance on several well known face datasets such as MOBIO [11] and Multi-PIE [6]. Despite this state-of-the-art performance, this approach has an obvious limitation as it does not enforce any spatial relationships between the blocks (observations), which discards spatial information which would help to disambiguate between the classes. Furthermore, its general applicability to vision problems has not been shown as it has only ever been applied to face recognition.

Contributions: In this paper we propose a local inter-

[1] http://itee.uq.edu.au/ conrad/lfwcrop/