Lessons from Collecting a Million Biometric Samples

P. Jonathon Phillips, Patrick J. Flynn, Kevin W. Bowyer

PII: S0262-8856(16)30128-7
DOI: doi: 10.1016/j.imavis.2016.08.004
Reference: IMAVIS 3539

To appear in: *Image and Vision Computing*

Received date: 18 October 2015
Revised date: 11 August 2016
Accepted date: 16 August 2016

Please cite this article as: P. Jonathon Phillips, Patrick J. Flynn, Kevin W. Bowyer, Lessons from Collecting a Million Biometric Samples, *Image and Vision Computing* (2016), doi: 10.1016/j.imavis.2016.08.004

Lessons from Collecting a Million Biometric Samples

P. Jonathon Phillips[a], Patrick J. Flynn[b], Kevin W. Bowyer[b]

[a] *National Institute of Standards and Technology*
100 Bureau Drive MS 8490
Gaithersburg, MD 20899, USA
jonathon@nist.gov
[b] *Computer Science and Engineering*
University of Notre Dame
Notre Dame, IN 46556, USA
flynn@nd.edu, kwd@cse.nd.edu

Abstract

Over the past decade, independent evaluations have become commonplace in many areas of experimental computer science, including face and gesture recognition. A key attribute of many successful independent evaluations is a curated data set. Desired aspects associated with these data sets include appropriateness to the experimental design, a corpus size large enough to allow statistically rigorous characterization of results, and the availability of comprehensive metadata that allow inferences to be made on various data set attributes. In this paper, we review a ten-year biometric sampling effort that enabled the creation of several key biometrics challenge problems. We summarize the design and execution of data collections, identify key challenges, and convey some lessons learned.

Keywords: face recognition, algorithm performance, human performance, challenge problem

1. Introduction

The creation of designed and curated data sets for grand challenges and independent evaluations has been an important driving force behind progress in biometrics over the last two decades [1],[2],[3],[4],[5],[6],[7],[8],[9]. Data sets distributed to the research community foster the development of new algorithms and technologies, and they allow independent evaluations of the state-of-the-art. Data sets also contribute to the identification of future research