Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 123 (2018) 104–109

Procedia Computer Science

www.elsevier.com/locate/procedia

8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017

Looking at faces in the wild

Eugene Borovikov[1*], Szilárd Vajda[2†], Michael Bonifant[1‡], and Michael Gill[1]

[1]National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

[2]Department of Computer Science, Central Washington University, WA, USA

{Eugene.Borovikov, Michael.Bonifant, MGill}@NIH.gov, Szilard.Vajda@CWU.edu

Abstract

Recent advances in the face detection (FD) and recognition (FR) technology may give an impression that the problem of face matching is essentially solved, e.g. via deep learning models using thousands of samples per face for training and validation on the available benchmark data-sets. Human vision system seems to handle face localization and matching problem differently from the modern FR systems, since humans detect faces instantly even in most cluttered environments, and often require a single view of a face to reliably distinguish it from all others. This prompted us to take a biologically inspired look at building a cognitive architecture that uses artificial neural nets at the face detection stage and adapts a *single image per person* (SIPP) approach for face image matching.

Peer-review under responsibility of the scientific committee of the 8th Annual International Conference on Biologically Inspired Cognitive Architectures

Keywords: face detection, face matching, artificial neural network, single image per person

1 Introduction

Faces play a critical role in social interactions presenting a very convenient and non-intrusive way for visual identification and non-verbal communication. Although recent research on macaques indicates that facial identity may be encoded via a simple neural code that relies on the ability of neurons to distinguish facial features along specific axes in face space [1], we still do not understand how humans detect and read faces with little visual sampling per individual, generalizing their recognition ability to a vast variety of lighting, poses and expressions.

Modern face recognition (FR) systems have become quite advanced in recent years, showing near-human abilities to recognize faces [2]–[4] on very challenging face datasets [5]–[7]. Nearly all of them rely on deep neural nets (DNN), whose we recently observed due to the availability of affordable graphics processing units (GPU) allowing to train DNNs in hours rather than days.

DNNs originally have been inspired by biological perceptual systems [8] and have been shown to solve complex pattern recognition problems [9], but they appear to learn statistical patterns very

* Mastermind behind the approach and system integration

† Crucial methodology research and development

‡ Essential implementation components and evaluation

Peer-review under responsibility of the scientific committee of the 8th Annual International Conference on Biologically Inspired Cognitive Architectures

10.1016/j.procs.2018.01.017

differently from humans, as primates typically require just a few visual samples of an object, to start recognizing it from various view-points, while their artificial counterparts require thousands of samples per object to start approaching human-level recognition accuracy.

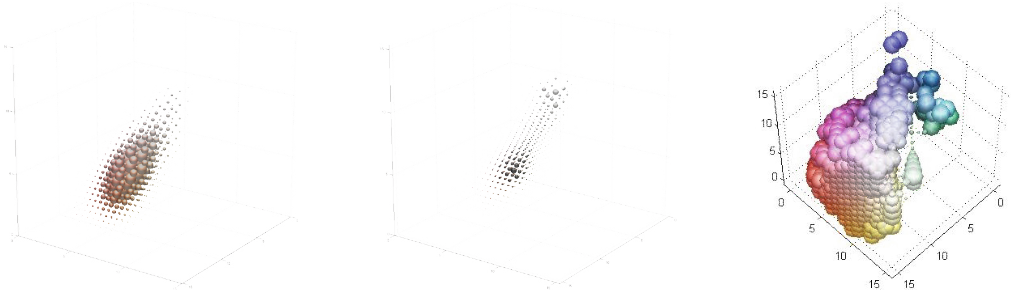
That prompted us to research and develop (R&D) a light weight (yet accurate) face detector and a *single image per person* (SIPP) face matcher, which is less complex than modern DNN systems, yet it is able to (a) use a single visual sample per subject, (b) be comparably accurate on unconstrained images, (c) adapt to the test visuals, and (d) run in near real-time requiring minimal computing power.

Our method cannot claim near-human level detection or recognition accuracy, but it does use several biologically inspired elements and it is utilized in a real-world face image retrieval system [10]. As biological systems inherit and then build up their perceptual abilities from the sensory experience, we proceed by R&D of a data-driven perceptual modules modeling inheritance (via coded algorithms) and experience (via statistical models).

2 Face localization in unconstrained images

Any real world face recognition system requires a reliable face localization (detection) stage that needs to be accurate and quick at the same time. Finding faces in unconstrained images presents many challenges to FR systems due to large variations of intrinsic (head pose, face expression, makeup, jewelry, etc.) as well as extrinsic (lighting, occlusions, blur, defocus, etc.) face image formation factors. To remedy these variations, we proceed by augmenting a baseline color-blind rotation-sensitive detector [11] by taking into account skin color, facial landmarks and face geometry.

2.1 Skin color mapping



(a) (b) (c)

Figure 1: (a) skin vs. (b) non-skin tone distribution and (c) resulting skin likelihood given color in RGB axes

We approach the problem of *skin mapping* by determining a real-valued skin likelihood map over any given image with a pixel-wise mapping function $s : C \rightarrow [0,1]$, where C is some color space and the skin likelihood values are real numbers in the range [0,1]. Researchers studied various color spaces [12], but for simplicity we start with RGB and use other spaces, as needed.

We compute skin and non-skin color histograms shown in Figure 1 using skin labeled data [10], [13]: (a) skin color forming a near-normal cluster in RGB, (b) non-skin color grouped around the gray-scale diagonal, and (c) conditional probability of skin (given color). The axes correspond to the color components that are quantized into 16 bins each. Each sphere has its bin's color with its size reflecting the bin's likelihood. Note that there is not much overlap between the skin and non-skin clouds, thus one can build a robust skin color classifier.

Bayesian skin mapper is based on conditional probability estimate from the source skin and non-skin pixels: $P(s|c) = P(c|s)p(s) / (P(c|s)P(s)+P(c|n)P(n))$, where $P(s|c)$ is the probability of skin given