Robotics and Autonomous Systems 75 (2016) 409–421

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Joint probabilistic approach for real-time face recognition with transfer learning

Noel Nuo Wi Tay *, János Botzheim, Naoyuki Kubota

Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, 191-0065, Japan

HIGHLIGHTS

- Real-time open universe face identification system is developed for robot partner.
- Efficient reformulated probabilistic method and transfer learning are used.
- Real-time domain mergence is developed and discussed.
- Able to outperform established methods under low training data.
- Lab test shows its adaptive learning and feasibility on low powered processors.

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form
31 July 2015
Accepted 1 September 2015
Available online 8 September 2015

Keywords:
Real-time
Robot partner
Face recognition
Probabilistic
Transfer learning

ABSTRACT

Face recognition is an integral part in robot partner's interaction with its human subject. With the task expected, open universe scenario face recognition needs to be performed under uncontrolled and uncooperative environment in real-time without imposing strain on its low powered processors. With the concept of informationally structured space where additional samples can contribute to constructing the prior for recognition, joint probabilistic face method is used for its accuracy, simplicity and transparency in data handling. A system is built upon the method that enables transfer learning from across different domains and real-time adaptive learning for continual information collection. Test result shows that the approach can perform well compared to other established methods under low number of prototype number. Test result on adaptive learning sheds light to the direction of future design.

1. Introduction

The presence of domestic robots is quite common nowadays, which ranges from household robots to edutainment robots. However, one type of robot that is also rising to prominence is the robot partner that provides security, assistance and companionship to human occupants. As human life-style is becoming increasingly hectic, complemented by the demographic shifts towards elderly people for developed and some developing countries, a non-human assistant becomes necessary. It is estimated that the population of those above 60 years old in the world will double to 2 billion in the year 2050 [1]. The problem is particularly serious in Japan as the elderly population is expected to consume 25.2% of the total population pie in Japan, many of which will lose the ability to live independently [2,3]. Therefore, robot partners become a necessity in order to maintain the welfare and comfort of their human occupants. They provide temporary companionship and also a focus point for other services like tele-presence, news announcement and entertainment.

To ensure favorable interaction with robot partner's human occupants, face recognition is one of the fundamental requirements as interaction mode and style differ between individuals. Apart from just identification or verification, it also needs to discern whether a face is considered familiar or not, which is termed as open universe scenario recognition. This paper will concentrate on open universe real-time face recognition. Given unconstrained and uncooperative condition during recognition, it also needs to be able to run on low powered processors, which is the case for most robots' onboard processors. Designing a good robot partner means seriously taking the tradeoff between accuracy and computational load into account.

One alternative over the tradeoff is to exploit surrounding information via informationally structured space that provides an environment for information gathering, storage, processing and control [4]. Surrounding information can then be supplied to the

* Corresponding author.
E-mail addresses: tay-noelnuowi@ed.tmu.ac.jp (N.N.W. Tay), botzheim@tmu.ac.jp (J. Botzheim), kubota@tmu.ac.jp (N. Kubota).

http://dx.doi.org/10.1016/j.robot.2015.09.002
0921-8890/

robot implicitly [5], so that it can perform the appropriate task, instead of solely relying on its onboard sensors. Readers can refer to the following works [6–8] for more information related to informationally structured space. It should be emphasized that this work does not deal explicitly with informationally structured space, but dwell on the implication it brings to designing a robust face recognition system.

Three issues we are addressing are different modality, recognition under uncontrolled condition and low computational load.

One can think of different modality as capturing images under different hardware like Kinect, webcam and security cameras, which imposes different scale and resolution. Recognition under different modality is limited to comparing normal photo images, and does not involve heterogeneous recognition [9]. Robot partners need to recognize faces under unconstrained and uncooperative situation [10]. It includes recognizing faces under different poses, expression and illumination. An efficient and fast method needs to be devised for the task without imposing constraints to its human occupants as this will hamper robot–human interaction.

As low powered processors normally run robot partners, face recognition process cannot be too heavy. For this work, we will use the iPhone 5s as a case study to function as a robot, which is from our previous work that is called the iPhonoid [11]. The motivation behind the iPhonoid is that robot partners should be able to accompany their owners, helping and collecting information during the course of everyday life, while being able to dispense services depending on the devices that are attached to it. Fig. 1 shows an iPhonoid that is attached to a body device with servo-powered limbs. Smartphone is a perfect choice as implementation medium for such robots due to their wide adoption by general public and interfacing capabilities with various devices from standardization. Face recognition system devised needs to be able to fit into the robot, among other modules such as motion control, speech recognition and generation, interface and gesture recognition (the other modules will not be covered in this paper). Although we use iPhone 5s for robot execution, it is by no means the best and we believe this work (face recognition with transfer learning) can traverse to other low-powered processors as well, which is the main motivation of this paper.

Given the three issues, a system built based on the reformulated joint probabilistic face matching method to support transfer learning is devised [12,13]. Besides, online learning based on spatio-temporal association is used to support initial low number of samples.

The outline of the paper is as follows: Section 2 will discuss the recent approaches in face recognition relevant to our issues and justifications. In Section 3, the proposed face recognition system is explained, where sub-sections will provide more detailed information on the different compartments of the whole system. Experimental results are provided in Section 4, as well as discussions. Finally, conclusion is made in Section 5.

2. Face recognition

As robot partners are expected to function in uncontrolled situation, their recognition system must be invariant to changes due to pose, illumination, and expression difference. Despite the requirement for such significant level of invariance, it needs to be discriminative enough to distinguish between individuals. Moreover, as robot partners are expected to come across people whom they have never met before, they need to be able to tell whether a person is familiar or not, and not just the nearest expected identity. This is termed as an open universe identification problem [14–16]. Recent progress in recognition for open universe scenario has been rapid both in verification [17,12,18,19] and identification [16]. Given the scope of robot partners, where population size is not too massive,

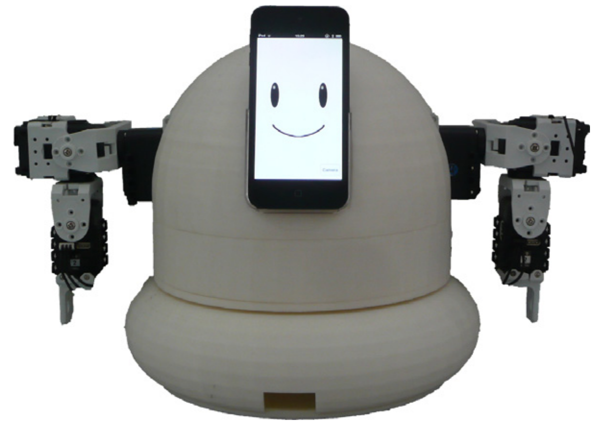

Fig. 1. iPhonoid attached to a peripheral device (body with servo-powered limbs).

identification can be performed through the scores of independent verification [20,21]. This paper will concentrate on face recognition with 2D image as input. For 3D face recognition, interested readers can refer to works by [22–25].

Current face recognition algorithms can reach near human level accuracy, or even seemed to surpass it [26–28]. This includes 3D morphable faces [29], sparse representation method [30,16] and deep networks [26].

Multi-layered networks and deep convolution networks have received significant attention recently in the field of machine learning and pattern recognition. They consist of multiple layers of networks that are greedily learned through each layer. Data abstraction increases with level of layers. In terms of face recognition, the higher level layers retain the identities of a person, which is invariant to distortions from illumination and pose. Deep architecture is currently applied in the state-of-the-art system in pair-wise comparison test for the challenging (Labeled Faces in the Wild) LFW database [17,31], which achieves near human level face recognition performance. Deep learning method achieves success through obtaining the appropriate representation that is abstract enough to be invariant to intra-class variations, but retains its discriminative power between inter classes. They can also be easily incorporated into other algorithms like the joint probabilistic verification [12] by performing the role of representation extraction [32]. But a drawback is that they are extremely computationally intensive and require huge amount of training data to construct.

On the other side of the continuum for face recognition approach are the model-based methods. A notable example is the morphable face method, which can estimate illumination, pose, and texture parameters of the face through the use of 3D face model [29]. Given the 3D face model of a person (estimated from a 2D image given a trained 3D face prior), various variations such as lightning, pose, expression and even weight difference can be simulated to compare with input face image for recognition purposes. It also achieves state of the art performance for Labeled Faces in the Wild test [33]. However, the computational load is very high, and is reported to take several minutes per face. The use of 2D models, complemented by prior information to compensate for the missing dimension [34,35], though faster, faces issues like self-occlusion. But recently, a fast variant of the 3D morphable face model has been developed [36]. It relies on transforming the probing filter instead of building a model of the face.

Although recognition system trained using large amount of data can produce good results [37], this is not practical for robot partners. Individual templates need to be learnt incrementally overtime, yet at the same time, they do not face over-fitting issue from the lack of samples. Besides, continual learning and recognition are to be performed in real-time. Therefore, computational