Pattern Recognition 83 (2018) 388–400

Contents lists available at ScienceDirect

Pattern Recognition

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

Reimagining the central challenge of face recognition: Turning a problem into an advantage

Ognjen Arandjelović

School of Computer Science, University of St Andrews, St Andrews KY16 9SX, United Kingdom

ARTICLE INFO

Article history:
Received 13 December 2017
Revised 24 April 2018
Accepted 6 June 2018

Keywords:
Meta-algorithm
Paradigm change
Retrieval
Intra-class
Inter-class
Similarity
Dissimilarity

ABSTRACT

High inter-personal similarity has been universally acknowledged as the principal challenge of automatic face recognition since the earliest days of research in this area. The challenge is particularly prominent when images or videos are acquired in largely unconstrained conditions 'in the wild', and intra-personal variability due to illumination, pose, occlusions, and a variety of other confounds is extreme. Counter to the general consensus and intuition, in this paper I demonstrate that in some contexts, high inter-personal similarity can be used to advantage, i.e. it can help *improve* recognition performance. I start by a theoretical introduction of this key conceptual novelty which I term 'quasi-transitive similarity', describe an approach that implements it in practice, and demonstrate its effectiveness empirically. The results on a most challenging real-world data set show impressive performance, and open avenues to future research on different technical approaches which make use of this novel idea.

1. Introduction

Face recognition is often described as one of the most active areas of research in computer vision [1–4]. While I am unaware of attempts to formalize this claim and support it with rigorous empirical evidence, it is beyond doubt that the field has undergone substantial changes over time. By this I am not referring merely to changes in the technical approach which can be naturally expected to take place as advances are made, but rather to the practical paradigms and the context in which face recognition is employed.

Early face recognition work can be described as a proverbial exploratory mission which served to deepen the understanding of the key challenges and features (in an abstract sense) which have the greatest discriminative power [5,6]. Geometric features and the first statistical appearance based methods were described in this period. Thereafter the focus has shifted to the practical challenge of making face recognition useful in real world security oriented applications. It is in this period that the difficulty of the problem has crystallized, with concurrent changes in pose, illumination, resolution, and other extrinsic factors, exposing the limitations of the proposed algorithms [7–10]. Most face recognition work falls under the umbrella of this conceptual period. Despite the immense amount of research effort, both by academia and industry, the highly optimistic predictions expressed in the early years of face recognition research failed to materialize: in unconstrained conditions the performance of face recognition in security applications remains disappointing [11–13]. The key reason lies in the nature of the demands of most security applications on the one hand, and the inherent discriminative weakness of facial biometrics. As regards the former, security applications demand a low false positive rate (allowing an intruder the access to a resource carries a high cost) and often a low false negative rate (denying access to a legitimate user is frustrating, time consuming, and potentially costly). At the same time, on the latter point, there is no compelling evidence that face based biometrics even in principle can be used to attain these demands. Face recognition by humans, often intuitively seen as highly sophisticated, is in fact not very accurate when evaluated in conditions comparable to those in which automatic methods are expected to operate [14,15]. Humans use a variety of constraints, such as knowledge based priors ('whom do I expect to encounter in this place?'), complementary biometrics (height, gait, voice, etc.), and a plethora of others to simplify the task in everyday situations. However, such assumptions are either difficult to incorporate in automatic methods (e.g. due to the semantic gap) or inappropriate in the context of practical applications of interest. While work on the underlying fundamentals continues with unabated effort [16–20], with particularly promising innovations arising from the use of sparse coding [2,19,21], dictionary representations [22,23], and deep learning [24–27], turning point for face recognition research has come in the last decade with the emer-

URL: http://oa7.host.cs.st-andrews.ac.uk/

https://doi.org/10.1016/j.patcog.2018.06.006
0031-3203/

gence of massive amounts of visual data – the focus has shifted to the use of face recognition for the retrieval and organization of photographs and video recordings [4,27,28]. The requirements of these applications contrast the aforementioned requirements of security applications: following the successes of web search engines, by adopting the ranked retrieval presentation of output, both so-called type I and type II errors are much more readily tolerated. The user is often not overly troubled by not every instance of interest being retrieved, or it not being retrieved at rank-1, as long as correct matches are within a reasonable rank (the quantified meaning of 'reasonable' being somewhat dependent on the application).

Thus, to summarise briefly the history of face recognition, the field has largely been characterized by incremental (but important and cumulatively significant) technical advances with major practical leaps which came though by innovative ways of seeing the same problem though a different lens. In the present paper my aim is to achieve the latter. Specifically, I will argue from theory that a characteristic at the heart of all face recognition problems, which is universally considered as the key challenge, can in fact be turned into an advantage in the right context. My case is first put forward on rigorous theoretical grounds, and subsequently demonstrated and discussed using empirical evidence.

The broad topic of the present paper is that of face set retrieval and the central contribution relates both to the previous work on set based recognition and the work concerned with recognition in the context of large data collections [4,27,28]. In contrast to most work in the literature herein my principal interest is neither in the representation of face sets nor in the associated similarity measures *per se*. Rather, given a baseline algorithm for measuring the similarity of two face sets, I seek to leverage the structure of the data at a large scale, that of the entire database, to make the best use of the available baseline. In the sense that the proposed method has as its input both data (face image sets) and an algorithm (the 'baseline'), it can be accurately thought of as a *meta-algorithm*.

1.1. Problem statement

Given a query face set the aim is to retrieve image sets of the same person from a large database (the 'gallery'). More specifically, the desire is to order the gallery sets in decreasing order of confidence that they match the query by identity. Thus the ideal retrieval has all sets of the query person first ('matches') followed by all others ('non-matches'). I assume that the gallery is entirely unlabelled and may contain multiple sets of the same person.

2. Learnt transitive similarity

In this section I introduce the main contribution of the paper. In particular, I describe a general framework for face retrieval especially well suited for large collections of face images acquired "in the wild" i.e. in largely unconstrained imaging conditions, and characterized by highly unbalanced amounts of training data per class (person). I start by motivating the intuition behind the proposed method in the section which follows, and subsequently explain how this intuition can be formalized into a general retrieval framework.

2.1. Motivation and the key idea

It is insightful to begin by considering the motivation behind the key idea in the context of related previous work and in particular the Matched Background Similarity (MBS) method of Wolf et al. [29]. In brief, Wolf et al. argue that in building a classifier which discriminates the appearance of a specific person from that of all other people, the focus should be on discriminating between this person and those individuals most similar to them; improvements in discrimination against very dissimilar people matter less as these individuals are unlikely to be conflated with the person of interest anyway. The idea I introduce here can be seen as complementary and builds upon a similarly simple basic principle. Specifically, I make use of the observation that if person A is alike in appearance to person B, and similarly person B to person C, *on average* persons A and C are more likely to look alike than two randomly chosen individuals. I term this Quasi-Transitive Similarity, the prefix "quasi-" capturing the notion that the stated regularity is a statistical rather than a universal one, as I shall explain shortly.

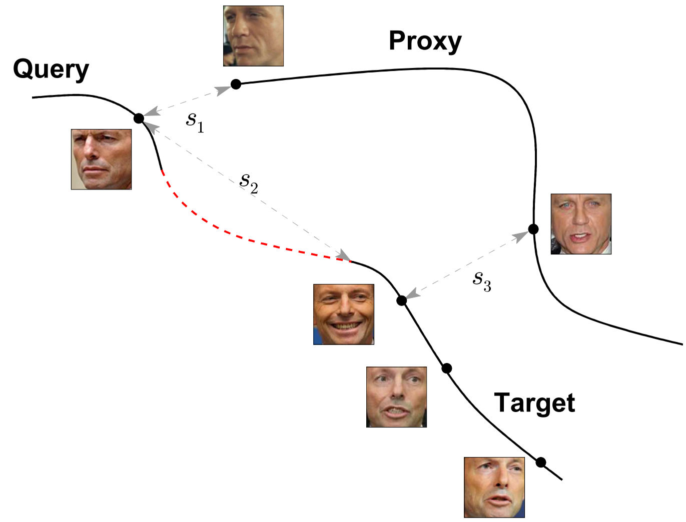

Fig. 1. The similarity between a query and the correct target set (initially poorly matched) may be better estimated indirectly via proxy data. 1D manifolds shown in black represent the appearance variability within sets. The dotted manifold shown in red represents the range of appearance of T Abbott present neither in the gallery nor in the query set. The query is poorly matched to the correct set because the person's pose in the query is vastly different than any of the poses in the target set. However, the query matches well the proxy set which contains more extensive pose variability of a person similar in appearance to the target person, the said similarity being directly inferable from data from the similarity of the matched images in the two sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This is illustrated conceptually in Fig. 1 using images of the former prime minister of Australia, Tony Abbot, and the actor Daniel Craig. For clarity, the variability of a person's appearance is shown as a 1D manifold. Specifically, the manifolds shown in black represent the appearance variability within the corresponding sets. The dotted manifold shown in red represents the range of appearance of Tony Abbott which is present neither in the gallery nor in the query set (in this conceptual example these are left semi-profile to left profile images).

As stated in the introduction above, the transitivity of similarity in appearance does not hold universally. It is possible that persons A and B are similar by virtue of one set of physical features, and B and C by virtue of another. A useful mental picture can be formed by drawing an analogy from statistics (or geometry): random variables (or vectors) A and B, and B and C may be positively correlated (have a positive dot product), yet A and C may be negatively correlated (have a negative dot product) with one another. This is illustrated in Fig. 2.

Lastly, it is worth contrasting the present approach with that of Yin et al. [30]. Unlike the method herein, their method necessitates the localization of face parts, which is problematic and highly likely to fail in severe illuminations, extreme poses, or in poor quality images. Their method also needs to extract estimates