Pattern Recognition 47 (2014) 208–215

Contents lists available at ScienceDirect

Pattern Recognition

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

Salient and non-salient fiducial detection using a probabilistic graphical model

C.F. Benitez-Quiroz*, Samuel Rivera, Paulo F.U. Gotardo, Aleix M. Martinez

Computational Biology and Cognitive Science Laboratory, The Ohio State University, 469 Dresse Laboratories, 2015 Neil Avenue, Columbus, OH 43210, United States

ARTICLE INFO

Article history:
Received 6 June 2012
Received in revised form
17 April 2013
Accepted 10 June 2013
Available online 21 June 2013

Keywords:
Shape modeling
Detailed face shape detection
Face detection
Probabilistic graphical model
Landmark detection

ABSTRACT

Deformable shape detection is an important problem in computer vision and pattern recognition. However, standard detectors are typically limited to locating only a few salient landmarks such as landmarks near edges or areas of high contrast, often conveying insufficient shape information. This paper presents a novel statistical pattern recognition approach to locate a dense set of salient and non-salient landmarks in images of a deformable object. We explore the fact that several object classes exhibit a homogeneous structure such that each landmark position provides some information about the position of the other landmarks. In our model, the relationship between all pairs of landmarks is naturally encoded as a probabilistic graph. Dense landmark detections are then obtained with a new sampling algorithm that, given a set of candidate detections, selects the most likely positions as to maximize the probability of the graph. Our experimental results demonstrate accurate, dense landmark detections within and across different databases.

1. Introduction

Shape detection is an important problem in computer vision and pattern recognition with applications in recognition, tracking, and classification, amongst others. The goal is to accurately detect the 2D position of specific shape landmarks, or *fiducials*, in an image. Some applications such as 3D reconstruction and the recognition of facial expressions require that the deformable shape be described by a *dense* set of *salient and non-salient* landmarks for satisfactory results [1–3]. Unfortunately, current detection algorithms are typically tailored to locating only a few salient landmarks [4–6]. Some exceptions are the 3-dimensional morphable model (3DMM) [7] and the 3D model of [8] which find a dense set of face landmarks. However, these methods require a 3D database to construct the model and, thus, cannot be learned directly from an image collection.

In this paper, we propose a novel statistical pattern recognition algorithm to accurately detect a dense set of salient and non-salient landmarks in an image. Our methodology does not require 3D object databases and can be used to design landmark detectors for different types of objects – *e.g.*, faces, hands, and structures in medical images. Our approach utilizes the fact that many object classes exhibit a homogeneous structure such that any detected landmark provides contextual information that facilitates the detection of the other landmarks. For instance, Fig. 1(a) shows a classical example with human faces as objects of interest. While at first sight one may not perceive the 10 faces in this image, once a few fiducials have been detected (*e.g.*, an eye or the nose), the remaining facial parts become readily apparent. Thus, the location of a fiducial automatically provides information on where to find the others.

In our new method, the relationship between every pair of landmark positions is encoded by the edges of a probabilistic graphical model, where each node represents a landmark position and its local texture. The local texture information of salient landmarks allow them to be detected reliably, whereas non-salient landmark detection is unreliable from just the local texture. Fortunately, the reliable detections constrain the position of non-salient landmarks and vice versa. In addition, the coarsely localized non-salient fiducials aid estimation of other non-salient and misdetected landmarks. As a key result, our detection algorithm can robustly estimate the positions of fiducials in areas such as face cheeks, where a simple local feature detector would generally fail, Fig. 1(b)–(c). Hence, the resulting algorithm can be used to estimate dense landmark maps in 2D images as in 3DMMs, without requiring prior 3D models.

Our proposed methodology is depicted in Fig. 2. Using our graphical model (presented in Section 3), landmark detection amounts to maximizing the joint probability of the graph's nodes

* Corresponding author. Tel.: +1 614 961 7212.
E-mail addresses: benitezc@ece.osu.edu (C.F. Benitez-Quiroz), riveras@ece.osu.edu (S. Rivera), gotardop@ece.osu.edu, paulo.gotardo@disneyresearch.com (P.F.U. Gotardo), aleix@ece.osu.edu (A.M. Martinez).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.06.013

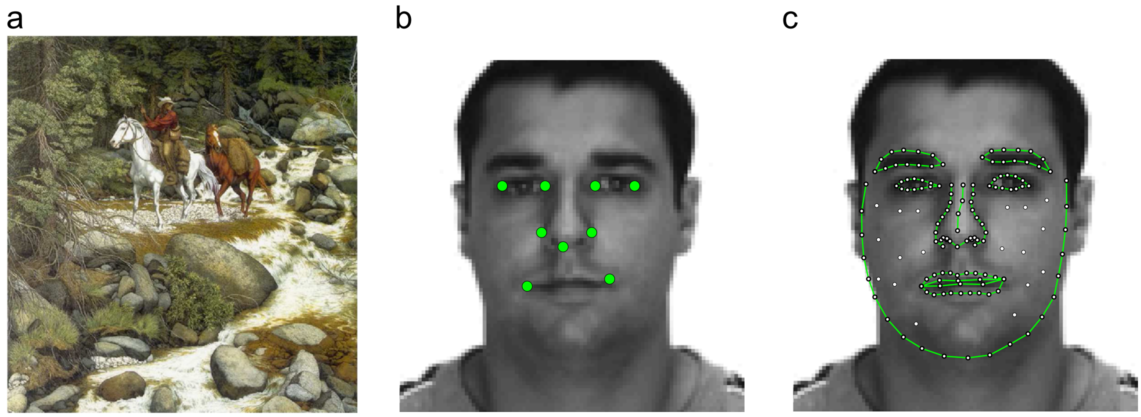

Fig. 1. Can you find the 10 faces in (a) these faces are difficult to see until a face feature is detected (*e.g.*, a nose); then the entire face becomes salient. (b) The output of a standard face landmark detector is typically restricted to a few salient points. (c) Our novel method provides dense detections that include both salient and non-salient landmarks.

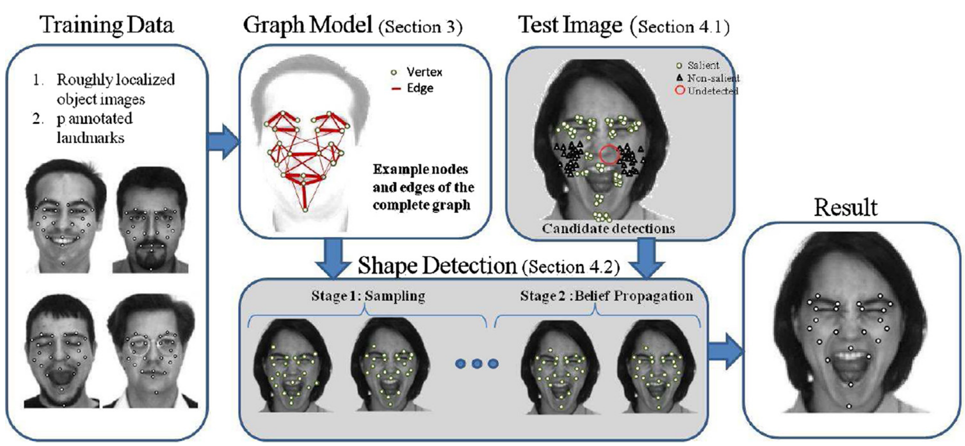

Fig. 2. Illustration of proposed methodology: the 'Graph Model' block shows the graph learning phase of the method (Section 3), where we model the relationship between salient and non-salient landmarks as a probabilistic graph. Thicker edges represent larger weights between fiducials. Only a subset of edges of the fully connected graph are shown. The 'Test Image' block highlights cases of misdetections and multiple detections for a particular fiducial. The 'Shape Detection' block (Section 4) shows that the graph model is used along with the local feature detections to determine the most probable shape configuration.

in an image. To accomplish this, we propose a sampling algorithm which selects the most likely fiducial positions given a set of candidate detections (Section 4), which resolves the classical computational complexity problem of pattern recognition algorithms that employ graphical models. This algorithm can deal with missing detections, false positives, and occlusions. In addition, we show how to augment the graph and use the original low-level detections to infer many more landmark coordinates in an incremental fashion. Our experimental results demonstrate accurate, dense landmark detections within and across different databases (Section 5).

2. Related work

Algorithms such as active appearance models (AAM) [9] and 3DMM [7] use a probabilistic shape and texture model to interpret an object image. These models are learned from a set of annotated training samples so they cannot detect variations beyond what is specified in the training set. In addition, the global shape model often favors a configuration similar to the mean shape and fails to capture subtle important changes such as eye blinks or single eyebrow motion. Furthermore, these algorithms are best suited to subject-specific modeling. On the other hand, algorithms that rely on fiducial detections are able to fit salient key points reliably without being overconstrained by a global model. However, these approaches can yield unrealistic shape estimates when the global shape is not constrained. These methods have advanced considerably in recent years, with algorithms that even rival human manual annotations [10–15]. However, they provide a very limited number of fiducials around salient features such as the eyes, nose, and mouth, and most require high-quality images.

Our new pattern recognition method overcomes the above shortcomings by utilizing the positive aspects of fiducial detection and probabilistic shape and texture model approaches. We take advantage of advances on local feature detection to ensure that subtle shape changes are not missed by our method. Each landmark position takes into account the position of *all* other locally detected fiducials to generate a plausible configuration for the whole set of detected points. If some landmarks cannot be detected or are misdetected by the local feature detector, the other fiducials will constrain estimation of their positions.

Graphical models have previously been used to guide the fiducial position estimation, although using different approaches which are limited to a very small number of landmarks. Felzenszwalb and Huttenlocher [4] use a tree structured graph to infer the location of five face fiducials. For tree based graphs, a poorly localized root node negatively influences all daughter nodes. In contrast, our model represents a dense interconnection of landmarks. Every node has influence from all other shape landmarks so the effect of a few poorly localized nodes is circumvented by information from other nodes in the graph.