Medical Image Analysis 43 (2018) 23–36

Contents lists available at ScienceDirect

Medical Image Analysis

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

Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization

Martin Urschler[a,b,c], Thomas Ebner[a], Darko Štern[a,b,*]

[a] Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria
[b] Institute for Computer Graphics and Vision, Graz University of Technology, Austria
[c] BioTechMed-Graz, Austria

ARTICLE INFO

Article history:
Received 29 March 2017
Revised 27 July 2017
Accepted 11 September 2017
Available online 21 September 2017

Keywords:
Anatomical landmarks
Localization
Random regression forest
Coordinate descent

ABSTRACT

In approaches for automatic localization of multiple anatomical landmarks, disambiguation of locally similar structures as obtained by locally accurate candidate generation is often performed by solely including high level knowledge about geometric landmark configuration. In our novel localization approach, we propose to combine both image appearance information and geometric landmark configuration into a unified random forest framework integrated into an optimization procedure that iteratively refines joint landmark predictions by using the coordinate descent algorithm. Depending on how strong multiple landmarks are correlated in a specific localization task, this integration has the benefit that it remains flexible in deciding whether appearance information or the geometric configuration of multiple landmarks is the stronger cue for solving a localization problem both accurately and robustly. Furthermore, no preliminary choice on how to encode a graphical model describing landmark configuration has to be made. In an extensive evaluation on five challenging datasets involving different 2D and 3D imaging modalities, we show that our proposed method is widely applicable and delivers state-of-the-art results when compared to various other related methods.

1. Introduction

A large number of medical image analysis applications rely on automatic anatomical landmark localization algorithms as a preliminary step for, e.g. segmentation based on deformable and statistical shape models (Heimann and Meinzer, 2009; Zhang et al., 2012; Lay et al., 2013), registration of images using rigid (Hajnal et al., 2001) and deformable transformations (Johnson and Christensen, 2002; Urschler et al., 2006), construction of anatomical atlases for population studies (Toews et al., 2010), or to focus on anatomical structures of interest for computer-aided diagnosis (Doi, 2007) and regression tasks like skeletal age estimation (Thodberg et al., 2009; Štern et al., 2014). However, due to anatomical variation and especially in the presence of potentially ambiguous (i.e. locally similar) landmarks, the task of both accurate and robust anatomical landmark localization becomes challenging.

The majority of recent machine learning approaches for multiple landmark localization either solely makes use of appearance features from the whole input image, or combines appearance features derived from the local vicinities of landmarks with a parametric or graphical model fitting step on top. Criminisi et al. (2013) have shown that global anatomical configuration can be captured with a regression-based prediction strategy, when the scale of exclusively used appearance features is allowed to vary up to the image size. By learning from the appearance of all anatomical structures present in the training dataset, their regression-based random forest (RF) for organ bounding box localization demonstrates robustness in the presence of ambiguities, however, their approach lacks in accuracy. On the other hand, restricting appearance features to a local neighborhood during training enables accurate prediction of landmark positions, however, on its own it inevitably leads to false positives due to locally similar anatomical structures not being distinguishable. Therefore, to select a likely global landmark configuration, sophisticated geometric configuration models like graphical (Donner et al., 2013) or constrained local models (Lindner et al., 2015) are required on top for disambiguation.

* Corresponding author at: Institute for Computer Graphics and Vision, Graz University of Technology, Austria.
E-mail address: stern@icg.tugraz.at (D. Štern).

https://doi.org/10.1016/j.media.2017.09.003

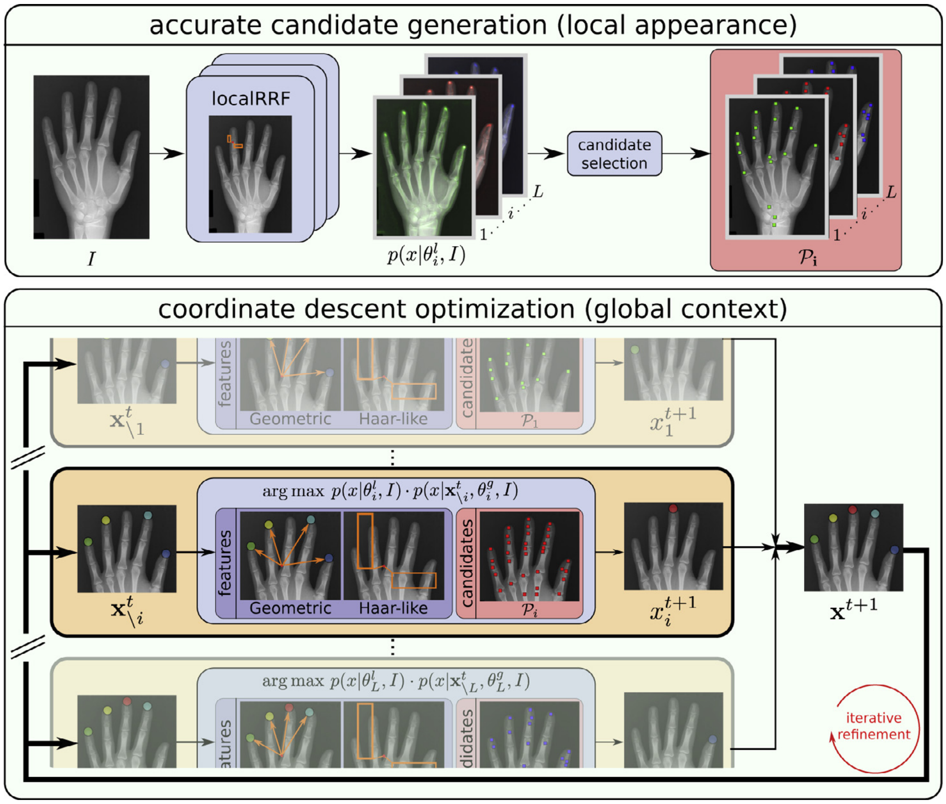

Fig. 1. Our proposed method for multiple landmark localization integrates appearance information and geometric configuration by first accurately generating landmark candidates (top), followed by iteratively refining landmark configuration based on coordinate descent optimization to achieve robustness in the presence of locally similar structures (bottom).

In this work we propose to integrate image appearance information and geometric landmark configuration into a unified framework. Thus, we aim for structured output prediction, where the landmark configuration is *implicitly* modeled by simultaneously taking into account both appearance features derived from the whole image and geometric constraints that landmarks impose onto each other. We pose our novel approach as an optimization problem that maximizes the joint probability for locating all landmarks simultaneously given an input image. This problem is solved by the coordinate descent algorithm (Wright, 2015), which decouples the NP hard problem of simultaneously searching for all landmarks by successively optimizing for individual ones. Thus, as illustrated in Fig. 1, by passing information to other landmarks and receiving information required to update its own position, each landmark location iteratively achieves a self-improvement according to both image appearance features and the current best beliefs of where the other landmarks are located.

1.1. Related work

While localization of multiple anatomical landmarks from medical images is in principle possible by solely relying on heuristic intensity-based image processing techniques (Wörz and Rohr, 2006; Donner et al., 2007), nowadays machine learning approaches capturing prior knowledge on appearance and shape from training data are predominantly used due to better generalizability in the presence of anatomical variation. These learning methods allow landmark localization either by using predictions for each landmark based on appearance information extracted from the whole image, or by combining potentially ambiguous predictions based on locally restricted appearance information with an *explicit* regularization model encoding geometric landmark configuration.

Using a regression-based approach implemented in an RF framework, Pauly et al. (2011) and Criminisi et al. (2013) proposed to learn the relative positions between the organs of interest and all the anatomy available in the training data solely with arbitrary scale Haar-like appearance features, i.e. the size of the appearance features and their distance from the training voxels are arbitrary. In a test image, the position of organs of interest is then obtained from the recognized anatomy, implemented by accumulating the relative positions of each voxel of the image to the organ of interest. Mainly aiming for image retrieval applications, the focus of Criminisi et al. (2013) lay on fast and robust but approximate and inaccurate bounding box localization of larger organ structures. The stratified decision forest method (Oktay et al., 2017) extended the work of Criminisi et al. (2013) for cases when there is significant variation of pose and size in a dataset. Gauriau et al. (2015) combined cascaded regression with a simple statistical shape prior derived from segmentation masks. For a multi-object segmentation task, Glocker et al. (2012b) developed an RF based method combining a classification and regression objective function. Ebner et al. (2014) adapted the multi-output regression approach of Criminisi et al. (2013) to the landmark localization task. They improved localization accuracy by putting more trust into the surrounding anatomy of the predicted landmark, since the relative position of closer anatomy shows less variation regarding the landmark location. This was achieved by introducing a distance weight at testing time that reduces influence of regions farther away from the anatomical landmark. Additionally, they used a two-stage regression RF cascade, with the scale range of Haar-like appearance features in the second stage restricted on capturing only the appearance of the structures locally surrounding a landmark. Applied to a segmentation task, Peter et al. (2015) developed this idea further by automatically selecting the most informative scale range of the appearance features defining surrounding