Learning-based Landmarks Detection for Osteoporosis Analysis

Erkang Cheng, Ling Zhi, Jie Yang, Azhari Azhari, Suhardjo Sitam, Xin Liang, Vasileios Megaloikonomou, and Haibin Ling

Department of Computer and Information Sciences, Temple University, Philadelphia, USA
School of Medicine, Shanghai Jiao Tong University, Shanghai, China
Division of Oral and Maxillofacial Radiology, School of Dentistry, Temple University, Philadelphia, USA
Department of Dentomaxillofacial Radiology, Padjadjaran University, Bandung, Indonesia
School of Stomatostomy, Dalian Medical University, Dalian, China
Computer Engineering and Informatics Department, University of Patras, Patras, Greece

ABSTRACT

Osteoporosis is the common cause for a broken bone among senior citizens. Early diagnosis of osteoporosis requires routine examination which may be costly for patients. A potential low cost diagnosis is to identify a senior citizen at high risk of osteoporosis by pre-screening during routine dental examination. Therefore, osteoporosis analysis using dental radiographs servers as a key step in routine dental examination. The aim of this study is to localize landmarks in dental radiographs which are helpful to assess the evidence of osteoporosis. We collect eight landmarks which are critical in osteoporosis analysis. Our goal is to localize these landmarks automatically for a given dental radiographic image. To address the challenges such as large variations of appearances in subjects, in this paper, we formulate the task into a multi-class classification problem. A hybrid feature pool is used to represent the landmarks. For the discriminative classification problem, we use a random forest to fuse the hybrid feature representation. In the experiments, we also evaluate the performances of individual feature component and the hybrid fused feature. Our proposed method achieves an average detection error of 2.9 mm.

Keywords: Landmark detection, Random forest, Hybrid features

1. INTRODUCTION

Osteoporosis is the common cause for a broken bone among senior citizens. Patients may suffer pain from bone fractures such as spine or dental osteoporosis fractures. Early diagnosis of osteoporosis requires routine examination which may be costly for patients. A potential low cost diagnosis is to identify a senior citizen at high risk of osteoporosis by pre-screening during routine dental examination. Therefore, osteoporosis analysis using dental radiography serves as a key step in routine dental examination. Further analysis of dental radiographs would provide additional information. For example, bone mineral density (BMD) analysis is one of the most important means for osteoporosis fractures estimation. Texture analysis of region of interest (ROI) in dental screening image is a way for osteoporosis analysis, as many studies have shown that 2D texture analysis can indirectly evaluate 3D microarchitecture.

Recent studies include trabecular texture analysis in multi-ROI in dental cone beam computed tomography (CBCT) data. ROI in dental screening image can be used for osteoporosis analysis. The aim of this study is to localize landmarks in dental radiographic images which are helpful to assess the risk of osteoporosis. We collect eight landmarks which are useful to conduct osteoporosis analysis. Specifically, these landmarks provide six regions (ROIs) for further image-based osteoporosis analysis. Fig. 1 shows an example of eight landmarks on an X-ray image (detailed explanation is given in Sec. 1).

Our goal, in this paper, is to automatically localize these landmarks on a given dental radiographic image. Basically, we treat landmarks detection as a multi-class classification problem. There are two key ingredients in our framework: 1) a hybrid feature which brings diverse types of discriminative information in a dental
radiographic image. 2) a random forest which is a discriminative framework that automatically leverages and fuses the information to achieve multiple landmarks detection.

The rest of the paper is organized as follows. Sec. 2 presents the definitions of landmarks for osteoporosis analysis and describes the different parts of the proposed approach. Then experimental results are reported in Sec. 3. Finally, Sec. 4 concludes the paper.

2. METHODS

We first present a summary of our approach. For a dental radiographic image, the task is to localize the anatomical landmarks $L = \{L_i | i = 1, \ldots, 8\}$. These landmarks provide multiple ROIs which are helpful to assess the risk of osteoporosis. In this paper, we formulate the task into a multi-class classification problem. Given an image $I$ and a pixel $x$, we denote its feature representation by $\Phi(x)$. Then, a random forest classifier is applied to determine the probability that the pixel belongs to landmark $L_i$. The final localization of landmark $L_i$ is computed by finding the local maximum of the probabilistic map.

In this section, we first introduce the definition of the eight landmarks. Then, the components of the hybrid features are given followed by the discriminative learning framework.

2.1 Landmarks definition

As illustrated in Fig. 1, the eight landmarks $L$ are used to define six ROIs $P = \{P_i | i = 1, \ldots, 6\}$. These six ROIs $P$ are studied for texture analysis in the assistance of osteoporosis analysis.

Formally, the six patches are defined as:

- Patch 1 and 4: inferior and internal to horizontal line from internal margin of mental foramen;
- Patch 2 and 5: external and superior to the intersecting point of the vertical line between internal margin of rami and the internal margin of inferior alveolar canal;
- Patch 3 and 6: superior to the external one-fourth point of the horizontal line between bottom of sigmoid notch and external margin of condyle neck.
The spatial relation between landmarks \( L_i \) and ROIs \( P_j \) is defined by:

\[
\begin{align*}
  x_{P_i} &= x_{L_i} + \Delta x, \\
  y_{P_i} &= y_{L_i} + \Delta y, \\
  x_{P_j} &= x_{L_j} - \Delta x, \\
  y_{P_j} &= y_{L_j} - \Delta y, \\
  x_{P_k} &= x_{L_k} + \frac{(x_{L_3} - x_{L_4})}{4}, \\
  y_{P_k} &= y_{L_k} = y_{L_4}, \\
  x_{P_l} &= x_{L_l} - \Delta x, \\
  y_{P_l} &= y_{L_l} + \Delta y, \\
  x_{P_m} &= x_{L_m} + \Delta x, \\
  y_{P_m} &= y_{L_m} - \Delta y, \\
  x_{P_n} &= x_{L_n} - \frac{(x_{L_5} - x_{L_6})}{4}, \\
  y_{P_n} &= y_{L_n} = y_{L_6}.
\end{align*}
\]  

(1)

where \((x_{L_i}, y_{L_i})\) is the location of landmark \( L_i \) and \((x_{P_j}, y_{P_j})\) represents the center of ROI \( P_j \). Parameters \( \Delta x \) and \( \Delta y \) are constant offset values.

### 2.2 Feature Representation

To address the diverse appearances of the eight landmarks, we use hybrid features to incorporate short and long range context information. Hybrid features\(^6\) have been studied in retinal vessel segmentation. Here, we present the feature representations suitable for landmark detection in dental radiographic images. These rich feature representations include short range context feature (Haar feature and Local feature) and the long range context feature.

#### 2.2.1 Haar Feature

The Haar feature in 2D image was originally introduced for face detection\(^6\) and is very popular in object detection. Taking benefit of the integral images, a Haar feature can be calculated efficiently in constant time. The computation of Haar features is based on Haar-like filters. Fig. 2 (a) provides seven Haar-like filters which are used in our method. Given a local patch size of \( m \times m \), Haar feature exhaustively explores the local contrast cues across local regions. The combinations of local contrast would provide discriminative context information for landmarks detection.

#### 2.2.2 Local Feature

Another type of short range context feature is the local feature which depends on local intensity or gradient values around one pixel \( x \). The local feature was studied in a single landmark detection in 3D CT data.\(^7\) In our method, we apply local feature for multiple landmarks detection for 2D dental radiographic images.

Basically, in local feature, we sample a few points from the image around a given pixel \( x \). Then, a few local features are extracted for each sampling point (e.g. intensity and gradient). Fig. 2 (b) shows an example of local feature extraction used in our work. For a given pixel \( x \) (in white), \( K \) points (red dots) are sampled in a local region of size \( n \times n \). At each sampling point \( x_k \), we use values such as intensity \( I(x_k) \) and gradient \( |g(x_k)| \) to be the local features. Transformations of \( I(x_k) \) and \( |g(x_k)| \) are also included to boost the feature pool.
2.2.3 Long Range Context Feature

Long range context feature\(^3\) was proposed for 3D organ localization. The problem with identifying landmarks in image is that different landmarks may share similar appearance. For example, in Fig. 1, landmarks \(L_3\) and \(L_7\) share very similar local appearance due to the spatial symmetry. Landmark pair \((L_2, L_4)\) and \((L_1, L_3)\) also have similar local context. Thus, local context information is not sufficient to differentiate these landmarks. Long range spatial context and topological cues must be used to enrich the feature representation. Ref. 8 presents an efficient way to capture the long range context information. We take the similar approach and generalize it for 2D images.

As explained in Fig. 2 (c), for a pixel \(x\), long range context feature incorporates information from multiple regions which are offset by a quantity \(\Delta L\) in a given direction. Therefore, long range context features capture both the appearance of anatomical structures as well as their relative context. Mathematically, a feature \(f(x)\) is defined with two offset regions \(R_1, R_2\) shooting from pixel \(x\):

\[
f(x) = \sum_{x_p \in R_1} I(x_p) - \sum_{x_q \in R_2} I(x_q)
\]  

(2)

To summarize, for a pixel \(x\), a hybrid feature vector \(\Phi(x)\) is computed to fuse Haar feature, local feature and long range context feature.

2.3 Discriminative Framework

The proposed hybrid features form a large feature pool \(\Phi(x)\). To handle the feature pool with large dimension, we choose the random forest framework\(^8\) due to its automatic suitability for multi-class classification and robustness in the medical image domain.\(^9\)

A random forest is an ensemble classifier that consists of decision trees and each tree is constructed via some randomized configuration. The randomization allows flexibility to explore a large feature space effectively because it only considers a subset of features in each tree node. As illustrated in Fig. 3, a leaf node encodes the class distribution of landmarks \(L\) for samples that reach that node. An internal node instead performs a binary test to split the samples to its children nodes. The splitting terminates when a leaf node is reached. The posterior probability at each leaf node is learned as the proportion of the training samples at the given leaf node. More details can be found in Ref. 8, 9.
Table 1. Detection of results of landmarks compared to manual annotation (in mm).

<table>
<thead>
<tr>
<th>Landmark</th>
<th>Local</th>
<th>Long range</th>
<th>Local</th>
<th>Long range</th>
<th>Haar</th>
<th>Local</th>
<th>Haar</th>
<th>Local</th>
<th>Haar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td></td>
<td>mean</td>
<td></td>
<td></td>
<td>mean</td>
<td></td>
<td>mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>std</td>
<td></td>
<td>std</td>
<td></td>
<td></td>
<td>std</td>
<td></td>
<td>std</td>
<td></td>
</tr>
<tr>
<td>Landmark 1</td>
<td>5.7750</td>
<td>4.5857</td>
<td>4.4480</td>
<td>6.2808</td>
<td>4.5843</td>
<td>3.5938</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.5960</td>
<td>2.9303</td>
<td>2.9602</td>
<td>4.5934</td>
<td>3.5581</td>
<td>3.1747</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 2</td>
<td>5.0671</td>
<td>4.0347</td>
<td>3.7715</td>
<td>5.6875</td>
<td>3.7185</td>
<td>3.3208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.7102</td>
<td>2.8247</td>
<td>3.0097</td>
<td>5.2972</td>
<td>2.6662</td>
<td>2.7912</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 3</td>
<td>3.1374</td>
<td>3.6186</td>
<td>2.8627</td>
<td>3.4535</td>
<td>2.3237</td>
<td>2.4059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.1806</td>
<td>2.0407</td>
<td>2.0661</td>
<td>5.2019</td>
<td>2.9475</td>
<td>2.0354</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 4</td>
<td>2.7591</td>
<td>5.0984</td>
<td>3.6208</td>
<td>4.8343</td>
<td>3.6792</td>
<td>3.1940</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.0265</td>
<td>3.3065</td>
<td>3.3789</td>
<td>5.6928</td>
<td>3.9082</td>
<td>2.3544</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 5</td>
<td>4.5707</td>
<td>4.1906</td>
<td>4.2055</td>
<td>5.9198</td>
<td>4.1117</td>
<td>3.3965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.8585</td>
<td>3.4817</td>
<td>3.2916</td>
<td>4.4255</td>
<td>2.9604</td>
<td>2.9489</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 6</td>
<td>4.6599</td>
<td>3.6882</td>
<td>3.4827</td>
<td>5.2048</td>
<td>3.6339</td>
<td>2.9455</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.3526</td>
<td>2.4540</td>
<td>2.3221</td>
<td>4.2059</td>
<td>2.4312</td>
<td>2.1191</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 7</td>
<td>2.0877</td>
<td>4.2010</td>
<td>2.4107</td>
<td>3.1629</td>
<td>2.1932</td>
<td>2.3247</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5453</td>
<td>4.3895</td>
<td>1.5814</td>
<td>5.0811</td>
<td>2.7612</td>
<td>1.7350</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmark 8</td>
<td>2.7693</td>
<td>4.0882</td>
<td>3.1336</td>
<td>3.0413</td>
<td>3.7652</td>
<td>2.7548</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.3321</td>
<td>2.5934</td>
<td>2.1617</td>
<td>2.8873</td>
<td>4.7359</td>
<td>1.7971</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>mean</td>
<td>4.8508</td>
<td>4.1893</td>
<td>3.9291</td>
<td>4.6998</td>
<td>3.5400</td>
<td>2.9931</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>std</td>
<td>4.0119</td>
<td>3.0915</td>
<td>2.7187</td>
<td>4.8572</td>
<td>3.6882</td>
<td>2.4420</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the testing phase, the feature \( \Phi(x) \) of a pixel \( x \) is first fed into the root of each tree and then it follows the splitting rule till it reaches a leaf (red paths in Fig. 3). Each tree returns a posterior probability that \( x \) belongs to each landmark \( L_i \). The mean of leaf distributions from all trees is used for final decision. Specifically, the probability that \( x \) is a landmark \( L_i \) is estimated by:

\[
\Pr(L_i | x) = \frac{1}{T} \sum_{i=1}^{T} p_i (L_i | \Phi(x)),
\]

where \( p_i (.) \) is the output from the \( t \)-th tree and \( T \) is the number of trees in the forest. The final localization of landmark \( L_i \) is computed by finding the local maximum of the probabilistic map.

3. EXPERIMENTS

3.1 Experiment Data

In the experiments, we used a dataset of 150 images with size of \( 2340 \times 1532 \). The resolution of these images is 400 dpi. Each landmark \( L_i \) is annotated by an expert. The parameters which define the spatial relations of landmarks \( L \) and ROIs \( P \) are set to \( \Delta x = 50 \) and \( \Delta y = 15 \). Due to issues of computation, we study the down-sampled images (\( 710 \times 383 \)) in the experiments.

3.2 Experimental Results

In the experiments, 100 images are used in the training of random forest classifier and 50 are used for evaluation. The details of each feature component for a pixel \( x \) are summarized as follows:

- Haar feature: The local patch size for Haar feature is \( 50 \times 50 \). These seven types of Haar-like filters give a Haar feature with dimension of 45,228 for each pixel \( x \);
- Local feature: For a pixel \( x \), the local region is set to \( 100 \times 100 \) and the sampling stride is set to 2 pixels. We extract 20 values for the local feature based on intensity and gradient values.
• Long range context feature: The long range context feature is computed from a large window of $400 \times 300$ around pixel $x$. The size of offset regions $R_1$ and $R_2$ is set to $30 \times 30$.

A random forest with 10 decision trees is learned from these 100 training images. In a splitting node of the forest, 100 random features from each feature component of feature $\Phi(x)$ are chosen.

In order to study the effects of individual feature components in our method, we also conduct the experiments using each individual feature component. Instead of using the hybrid features, we only apply a specific type of feature in the evaluations. Similarly, in a splitting node of the forest, 100 random features are chosen in the specified feature component.

The evaluation metric is the distance to the manual annotation of these landmarks. The experimental results are listed in Table 1. These studies confirm that the proposed features fused by the random forest framework boost the performance by 25% in average.

4. CONCLUSION

To summarize, in this study, we propose a framework for automatic landmark detection in dental radiographic image. A discriminative random forest is applied to localize each landmark in a dental radiographic image. A fused rich feature representation is proposed to address the large variations of appearance of these landmarks. In the experiments, we evaluate the ability of different individual features and the hybrid one. The proposed approach achieves an average detection error of 2.9 mm. In the future, we plan to study image-based osteoporosis analysis from ROIs.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grants IIS-1407156 and IIS-1350521. The funding agency specifically disclaims responsibility for any analyses, interpretations and conclusions.

REFERENCES


