Fully Automatic Vertebra Detection in X-Ray Images Based on Multi-Class SVM

Fabian Lecron, Mohammed Benjelloun, Said Mahmoudi
University of Mons, Faculty of Engineering, Computer Science Department
20, Place du Parc, 7000 Mons, Belgium

ABSTRACT
Automatically detecting vertebral bodies in X-Ray images is a very complex task, especially because of the noise and the low contrast resulting in that kind of medical imagery modality. Therefore, the contributions in the literature are mainly interested in only 2 medical imagery modalities: Computed Tomography (CT) and Magnetic Resonance (MR). Few works are dedicated to the conventional X-Ray radiography and propose mostly semi-automatic methods. However, vertebra detection is a key step in many medical applications such as vertebra segmentation, vertebral morphometry, etc. In this work, we develop a fully automatic approach for the vertebra detection, based on a learning method. The idea is to detect a vertebra by its anterior corners without human intervention. To this end, the points of interest in the radiograph are firstly detected by an edge polygonal approximation. Then, a SIFT descriptor is used to train an SVM-model. Therefore, each point of interest can be classified in order to detect if it belongs to a vertebra or not. Our approach has been assessed by the detection of 250 cervical vertebrae on radiographs. The results show a very high precision with a corner detection rate of 90.4% and a vertebra detection rate from 81.6% to 86.5%.

Keywords: Vertebra Detection, X-Ray Images, Support Vector Machine, SIFT Descriptor

1. INTRODUCTION
A wide range of people will be affected by back pain in their life. Sometimes, this affliction is depicted as the curse of the century. Therefore, we easily understand the importance to better prevent, study and analyze back problems.

One of the most common procedures in the literature to analyze the spine is the vertebra segmentation which gives a physician precious quantitative data he would have difficulties to get manually. One can find a myriad of vertebra segmentation applications and segmentation methods as well. Another interesting procedure is concerned by the 3D reconstruction of the spine from multi-planar radiographs. The common point between all these methods is that most of them are not entirely automatic. In fact, they require the intervention of a user on the image. However, given the growing number of patients, one can understand the interest to develop fully automatic tools. Some works related to the automatic vertebra detection can be found in the literature. Actually, these contributions are mainly interested in only 2 medical imagery modalities: Computed Tomography (CT) and Magnetic Resonance (MR). A few works are dedicated to the conventional X-Ray radiography. However, this modality is the cheapest and fastest one to obtain spine images. Klinder et al.\(^1\) deal with the detection, the identification and the segmentation of CT images. They propose an automatic detection method based on the Generalized Hough Transform (GHT) and announce a detection rate of 92%. Huang et al.\(^2\) have used a learning-based approach applied to the vertebra detection on MR images. To this end, features such as Harr wavelet are extracted on images to train an AdaBoost learning algorithm. Peng et al.\(^3\) detect the intervertebral disks on MR images by least-square pattern matching with a predefined model to locate the vertebra position. Concerning the X-Ray images, the literature is quite poor. Zamora et al.\(^4\) try to take advantage of the GHT on radiographs but they present a segmentation rate equal to 47% for lumbar vertebrae without providing information about the detection rate. Very recently, Dong and Zheng\(^5\) have proposed a method combining GHT and the minimal intervention of a user with only 2 clicks in the image, providing detection rate of 92.4%. A fully automatic
approach has been proposed by Casciaro and Massoptier. They use a shape constraint characterization by looking for every shape that could be an inter-vertebral disc. They showed a detection rate of 83.02%.

In this paper, we develop a fully automatic vertebra detection method in X-Ray Images. The approach is based on a learning method: Support Vector Machine. Actually, we propose to train an SVM (Support Vector Machine) classifier for detecting the vertebra anterior corners. To this end, a local descriptor is used to represent the corners, i.e. SIFT (Scale-Invariant Feature Transform). In section 2, we present the workflow and the theoretical background of our approach. The section 3 is dedicated to the experimental results and shows that the corner detection rate reaches 90.4%. Finally, we conclude and discuss about future works.

2. METHOD OVERVIEW

The main idea of the proposed approach is to locate a vertebra by detecting its points of interest. Methods based on this idea have been proposed in the literature. In our case, the points of interest are the higher and lower anterior corners. Therefore, the first step of the method is to detect points of interest in the image. To this end, we have presented the benefits of the polygonal approximation dedicated to this task. The methodology is composed of 3 stages: a contrast limited adaptive histogram equalization, a Canny edge detection filter and an edge polygonal approximation. This way, the edges become segment lines whose intersections define the corners.

Once the points of interest in the radiograph are detected, we still need to classify them, to know if they belong to a vertebra or not. For that matter, we need to modelize the information characterizing the vertebra corners. In this paper, we propose to carry out this modelization by the Support Vector Machine learning approach. Such a method would be a modelization technique but also a classification one.

The Figure 1 gives the reader an overview of the framework we present for the vertebra detection. In the training phase, the points of interest are detected on sample images. Some features are then extracted to describe these points. The features are used to create a corner classifier thanks to the SVM-based learning. Finally, the corner classifier is part of the SVM classification in the detection procedure, to know if the detected pixels on an image belong the a vertebra or not. With the knowledge of the corners, the vertebra can finally be detected.

![Figure 1. Framework for the automatic vertebra detection](image)

The framework presented at the Figure 1 is quite used in the field of pattern recognition. The novelty of our work is in the way of conducting the steps of the framework. Our contributions are: to propose an original method for points of interest detection by edge polygonal approximation, the use of SIFT descriptor to describe the vertebra corners, the mix of these concepts in a learning system based on multi-class SVM. Finally, we present better results than those found in the literature.
2.1 Support Vector Machine

The SVM, introduced by Vapnik, is widely dedicated to classification and regression problems. Consider that we have a set of data \((x_1, y_1), \ldots, (x_m, y_m) \in \mathbb{R}^n \times \pm 1\). The goal of SVM is to determine the hyperplane separating at the best the data into 2 distinctive classes: +1 and −1.

In practice, the data are not linearly separable. In that case, the idea is to transpose them in a higher-dimensional space \(\mathcal{F}\), called feature space, to linearize the problem. Let \(\Phi(x) : \mathbb{R}^n \rightarrow \mathcal{F}\) be the mapping function. Let \(w \in \mathcal{F}\) and \(b \in \mathbb{R}\). The hyperplane \(\mathcal{H}\) is defined by the couple \((w, b)\) such as \(\langle w, \Phi(x) \rangle + b = 0\). Actually, \(w\) and \(b\) are obtained by solving the quadratic programming problem (1).

\[
\begin{align*}
\text{minimize}_{w,b} & \quad \frac{1}{2}||w||^2 \\
\text{subject to:} & \quad y_i \langle (w, \Phi(x_i)) + b \rangle \geq 1, \ i = 1, \ldots, m.
\end{align*}
\]

Generating the function \(\Phi(x)\) can turn out to be a very complex task. Usually, this is done by kernels introduced by the dual formulation of the problem (1). This problem is formalized by:

\[
\begin{align*}
\text{maximize}_\alpha & \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle \Phi(x_i), \Phi(x_j) \rangle \\
\text{subject to:} & \quad \alpha_i \geq 0, \ i = 1, \ldots, m, \\
& \quad \sum_{i=1}^{m} \alpha_i y_i = 0.
\end{align*}
\]  

If we focus on the dual problem (2), one can notice that actually, the function \(\Phi(x)\) is not needed explicitly. Only the inner products \(\langle \Phi(x_i), \Phi(x_j) \rangle\) are used in this dual formulation. Let \(K\) be a symmetric matrix consisting of the inner products \(\langle \Phi(x_i), \Phi(x_j) \rangle\). Any positive semidefinite matrix could be used as a kernel matrix \(K\), since the positive semidefiniteness ensures the existence of the function \(\Phi(x)\). Let \((\alpha^*_1, \alpha^*_2, \ldots, \alpha^*_m)\) be an optimal solution of the problem (2). It can be shown that \(\langle w^*, \Phi(x) \rangle = \sum_{i=1}^{m} \alpha^*_i y_i k(x, x_i)\).

From this point, we know all the parameters to compute the hyperplane \(\mathcal{H}\) separating our data. Therefore, any new point \(x \in \mathbb{R}^n\) can be classified by determining which side of the hyperplane \(x\) belongs to. This can be done by verifying the equation (3).

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{m} \alpha^*_i y_i k(x, x_i) + b \right)
\]  

Now that we have detailed the classification approach, we need to find a way to describe the data. To this aim, we define in the next section a local descriptor to characterize the corners.

2.2 SIFT Descriptor

The SIFT (Scale-Invariant Feature Transform) descriptor has been proposed by Lowe. Actually, this notion of descriptor is in keeping with an approach of invariant feature extraction in images and object matching between different views. In our case, we are not interested by the extraction since our goal is to classify points whose coordinates are already known. The advantage of this descriptor is the scale, translation and rotation invariance and the partial invariance to illumination changes.

The first step of the descriptor extraction is to compute the point orientation to assure a rotation invariance. To this end, Lowe proposes to compute the module and the orientation of the gradient in the neighborhood of the considered point. This operation is conducted on the image \(L(x,y)\) defined at the equation (4), where \(G(x, y, \sigma)\) is a Gaussian kernel and \(I(x, y)\) the original image. Due to reasons of simplicity, we replace \(L(x,y)\) by the notation \(L_{x,y}\).

\[
L_{x,y} = G_{x,y,\sigma} * I_{x,y}
\]  

The notation \(L_{x,y}\) of the considered point. This operation is conducted on the image \(L(x,y)\) defined at the equation (4), where \(G(x, y, \sigma)\) is a Gaussian kernel and \(I(x, y)\) the original image. Due to reasons of simplicity, we replace \(L(x,y)\) by the notation \(L_{x,y}\).
For a pixel \((x, y)\), the module and the direction of the gradient are given, respectively by the equation (5) and (6).

\[
m_{x,y} = \sqrt{(L_{x+1,y} - L_{x-1,y})^2 + (L_{x,y+1} - L_{x,y-1})^2}
\] (5)

\[
\theta_{x,y} = \tan^{-1}\left(\frac{L_{x,y+1} - L_{x,y-1}}{L_{x+1,y} - L_{x-1,y}}\right)
\] (6)

An orientation histogram is then built from the orientations of the points close to the pixel of interest. The highest peak detected in the histogram stands for the orientation of the point of interest.

Afterwards, we can rotate the coordinates of the point with an angle equal to the orientation. Then, a 4 x 4 window is created around the pixel. The 16 resulting regions are also divided in smaller ones. Let \(bSize\) be the size of these regions. Inside them, the module and the gradient of every points are computed and an orientation histogram discretized in 8 classes is built. The quantity added in the histogram corresponds to the module of the gradient at the current point. Nevertheless, pixels far from the point of interest must have a smaller influence than close ones. For this reason, a Gaussian function, with a \(\sigma\) equal to 1.5 times the descriptor window size, is used to weight the histogram. A graphical example is represented at the Fig. 2.

![Figure 2. Gaussian window around the descriptor](image)

3. EXPERIMENTAL RESULTS

3.1 Data and Methods
To conduct our experiments, we used 50 X-Ray radiographs coming from the NHANES II database*. They were chosen randomly but they all are focused on the cervical vertebrae C3 to C7. An example of a radiograph is given at the Figure 3.

The training phase requires to get the 128 coordinates of the SIFT descriptor from the points of interest detected in the sample images. All these points have to be divided into 3 classes: the higher anterior corners, the lower anterior corners and the bad points.

The results presented in the next section depend on an execution of our algorithm on 50 images. In order to have a complete independence between learning and testing data, we have performed a leave-one-out cross-validation. In this way, at every run of the algorithm, one image is tested and the rest of the sample (49 images)

is used for the learning phase. The classification precision is then computed and averaged among the 50 images. A 100% rate indicates that all the points in the radiograph have been correctly sorted in their associated class, i.e. higher anterior corner, lower anterior corner and bad point.

![Example of input image where the cervical vertebrae are visible](image.png)

Figure 3. Example of input image where the cervical vertebrae are visible (the resolution is $1763 \times 1755$ pixels for all the images)

### 3.2 Results and Discussion

Several parameters can influence the classification results and therefore the vertebra recognition rate. Nevertheless, our studies showed that only one parameter dictates the results in a significative way, i.e. $b\text{Size}$. As reminder, this variable stands for the size of the 16 regions composing the descriptor window. A prevalidation permits us to estimate the value of $b\text{Size}$ around 10.

At the Table 1, we present the anterior corner and vertebra detection rates for a $b\text{Size}$ value of, respectively, 8, 10 and 12 pixels. For each vertebra level, we distinguish the type of the corner: Higher and Lower. The percentage indicates the ratio between the number of correctly detected corners by our approach and the total number of corners. The first particular value concerns the C7 vertebra. We notice that, whatever the value of $b\text{Size}$ is, the lower corner detection rate is very weak, i.e. 64%. Such a limitation can be explained by the nature of the human body. The C7 vertebra is the less visible one on the radiographs of our database. Actually, this vertebra is often partially hidden by other parts of the body (e.g. shoulder). As a result, the lower anterior corner is less detected by the initial points of interest detector. The SVM classification method is not to blame here because corners which are not detected can’t be classified.

Globally, the detection rate for the other vertebra levels are very satisfying. For each vertebra, the lower anterior corner is characterized by a better success rate than the higher one (except for the C6 vertebra). We think that there is no obvious reason to explain this fact. However, this will have no impact on the vertebra detection results since we can extract an information about the column curvature and the mean vertebra height thanks to the successfully detected corners, and recover the missing ones.
Regarding the evolution of the bSize parameter, one can observe that the detection rates are improved as the value of bSize is increased. The enlargement of the descriptor window seems, in our case, to bring more information in the anterior corners detection. This observation is reinforced by the results presented at the Table 2. The reader can find the mean precision of the classification (Mean.), the associated standard deviation (S. D.) and the global percentage of misclassified points (Mis. Pts.). We notice that these indicators are the best for bSize = 12. Beyond this value, the results are not improved anymore.

<table>
<thead>
<tr>
<th>Vert.</th>
<th>Corn.</th>
<th>bSize = 8</th>
<th>bSize = 10</th>
<th>bSize = 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>High.</td>
<td>82%</td>
<td>74%</td>
<td>86%</td>
</tr>
<tr>
<td>Low.</td>
<td>90%</td>
<td></td>
<td>96%</td>
<td>82%</td>
</tr>
<tr>
<td>C4</td>
<td>High.</td>
<td>86%</td>
<td>80%</td>
<td>86%</td>
</tr>
<tr>
<td>Low.</td>
<td>92%</td>
<td></td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>C5</td>
<td>High.</td>
<td>76%</td>
<td>70%</td>
<td>78%</td>
</tr>
<tr>
<td>Low.</td>
<td>94%</td>
<td></td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>C6</td>
<td>High.</td>
<td>96%</td>
<td>94%</td>
<td>98%</td>
</tr>
<tr>
<td>Low.</td>
<td>98%</td>
<td></td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>C7</td>
<td>High.</td>
<td>96%</td>
<td>62%</td>
<td>96%</td>
</tr>
<tr>
<td>Low.</td>
<td>64%</td>
<td></td>
<td>64%</td>
<td>64%</td>
</tr>
</tbody>
</table>

Table 1. Anterior corners and vertebra detection rates

<table>
<thead>
<tr>
<th>bSize</th>
<th>Mean</th>
<th>S. D.</th>
<th>Mis. Pts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>97.85%</td>
<td>1.65%</td>
<td>2.14%</td>
</tr>
<tr>
<td>10</td>
<td>98.48%</td>
<td>1.60%</td>
<td>1.59%</td>
</tr>
<tr>
<td>12</td>
<td>98.84%</td>
<td>1.54%</td>
<td>1.15%</td>
</tr>
</tbody>
</table>

Table 2. Evolution of the classification precision with the bSize parameter (3-class SVM)

The Table 3 compares the detection rate of our approach with previous methods found in the literature. Among all the methods detailed in the section 1, we present at the Table 3 only those where the vertebra detection rate is known. The semi-automatic approach developed by Dong et al.\textsuperscript{5} is obviously the one with the highest detection rate. The real interest is to compare our approach with a fully automatic one. We show at the Table 3 two results related to our method. The first takes into account all the studied vertebrae, i.e. C3 to C7. Nevertheless, we noticed that, for the majority of images, the C7 vertebra was not clearly visible, even to the naked eye. Therefore, we also present the detection rate by not considering the C7 vertebra. In that case, we observe that our approach outperforms the Casciaro et al.’s one.\textsuperscript{6}

<table>
<thead>
<tr>
<th>Method</th>
<th>Mode</th>
<th>Detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dong et al.\textsuperscript{5}</td>
<td>Semi-automatic</td>
<td>92.40%</td>
</tr>
<tr>
<td>Casciaro et al.\textsuperscript{6}</td>
<td>Fully automatic</td>
<td>83.02%</td>
</tr>
<tr>
<td>Proposed method (with C7)</td>
<td>Fully automatic</td>
<td>81.60%</td>
</tr>
<tr>
<td>Proposed method (without C7)</td>
<td>Fully automatic</td>
<td>\textbf{86.50%}</td>
</tr>
</tbody>
</table>

Table 3. Comparison of the vertebra detection rate with previous methods

The confusion matrix related to our experiments is given at the Table 4. The results show that the 3 classes are well distinguished. One can notice that the lowest prediction rate characterizes the class representing the higher vertebra corners. This confirms the similar observation we made based on the analysis of the Table 1.

This matrix allows to determine the prediction errors made by the classification algorithm (SVM). Most of the confusions are between the class of the bad points and the class of the higher corner. Thereby, 12 bad points have been classified as higher vertebra corners while 14 higher vertebra corners have been classified as bad points. This observation also appears for the lower vertebra corners but to a lesser extent. This confirms that lower corners are better classified than higher corners. We think that this fact can be explained by the nature of the images.
### Table 4. Confusion matrix for the parameter bSize = 12

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Target class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High.</td>
</tr>
<tr>
<td>High.</td>
<td>227</td>
</tr>
<tr>
<td>Low.</td>
<td>1</td>
</tr>
<tr>
<td>Bad.</td>
<td>14</td>
</tr>
<tr>
<td>Prediction rate per class (%)</td>
<td>93.80</td>
</tr>
</tbody>
</table>

Finally, we propose at the Figure 4 an illustration of vertebrae detected on a radiograph. For better visualization, the image is focused on the region of interest. The higher corners have been correctly detected and are painted in blue while the lower corners are represented in green. The *bad* points that could have been detected, are obviously eliminated since they are considered as non relevant.

From the result showed at the Figure 4, several applications can be imagined. By knowing the exact position of the vertebra anterior corners, it is possible to compute angles between the vertebrae. Therefore, our approach of vertebra detection is ideal for the study of the vertebral mobility. Furthermore, numerous segmentation methods based on a model rely on an initialization. A user is frequently asked to point out the approximate position of the objects of interest on the image. This tedious part could be replaced by our fully automatic vertebra detection approach.

![Figure 4. Visual detection results for the vertebrae C3 to C7 (for better visualization, the image is focused on the region of interest)](image)

### 4. CONCLUSION

In this paper, we detailed an original procedure to automatically detect vertebra in X-Ray images. To this aim, we have developed a learning-based method combining the features of a local descriptor, *i.e.* SIFT, with a classification method, *i.e.* SVM. The experimentations were conducted on 50 radiographs focused on the cervical vertebrae C3 to C7. The obtained results were very promising and show a corner detection rate of 90.4% and a vertebra detection rate varying from 81.6% to 86.5%. One advantage of our method is that it is extensible to other detection problems. The workflow of the Figure 1 could be applied to a new object identifiable by points of interest.
Furthermore, several applications can be imagined. By knowing the exact position of the vertebra anterior corners, it is possible to compute angles between the vertebrae. Therefore, our approach of vertebra detection is ideal for the study of the vertebral mobility. Moreover, numerous segmentation methods based on a model rely on an initialization. A user is frequently asked to point out the approximate position of the objects of interest on the image. This tedious part could be replaced by our fully automatic vertebra detection approach.

In future works, we aim to improve our results by considering an a priori on the spine. For now, our approach does not allow to detect 100% of the vertebrae. A few corners are indeed not detected. Nevertheless, these corners can be retrieved by considering specific characteristics of the spine. Other descriptors could be used as well as other classification methods. For instance, the SURF (Speeded-Up Robust Features)\textsuperscript{13} has been proposed as an alternative faster than SIFT.

REFERENCES


