Application of computer-aided approaches to the PUMC classification of scoliosis

Junhua Zhang¹ · Hongjian Li² · Yufeng Zhang¹

Received: 23 January 2017 / Revised: 24 February 2017 / Accepted: 21 March 2017 / Published online: 3 April 2017
© Korean Society of Medical and Biological Engineering and Springer 2017

Abstract Surgical planning for scoliosis relies on the classification of the spinal curve pattern. To improve the reliability of the Peking Union Medical College (PUMC) classification system for scoliotic spinal curves, a computer-aided system is proposed and its reliability is evaluated. First, the reliability of curve measurements was improved by the computer-aided Cobb measurement approach. Second, judgmental errors were reduced by the computer program through the automation of the PUMC classification procedure. Four observers divided into an expert group and a resident group participated in the experiments. The kappa statistic was used to evaluate the variability. Classifications of 65 scoliotic cases by the four observers showed that with the computer's aid, the average intraobserver and interobserver kappa values were improved from 0.86 to 0.93 and from 0.75 to 0.86, respectively. The results indicate that the proposed computerized system can assist a surgeon in the PUMC classification of scoliosis and is especially useful for inexperienced surgeons.

Keywords Computer-aided diagnosis · Scoliosis · Classification · Radiograph

✉ Junhua Zhang
jhzhang@ynu.edu.cn; jhzhang_2006@163.com

Hongjian Li
samhongjian@126.com

Yufeng Zhang
yfengzhang@yahoo.com

¹ Department of Electronic Engineering, Yunnan University, No. 2 Cuihu Road, Kunming 650091, China

² Department of Orthopedics, the First People's Hospital of Yunnan Province, Kunming 650032, China

1 Introduction

Idiopathic scoliosis (IS) is a complex three-dimensional (3D) deformity of the spine for which there is no known cause [1]. Surgery is usually recommended if the spinal deformity is severe or progressive. Classification of the scoliotic curve pattern is very important for preoperative surgical planning to select fusion levels. Currently, the King classification [2], the Lenke classification [3], and the Peking Union Medical College (PUMC) classification [4] are the most widely used classification methods for that purpose. Qiu et al. [5] compared the reliability of these classification methods and reported that the PUMC classification had higher reliability (interobserver agreement of 84.3% with a kappa coefficient (κ) of 0.819, intraobserver agreement of 92.0% with a κ of 0.907) than the King classification (interobserver agreement of 81.2% with a κ of 0.773, intraobserver agreement of 91.5% with a κ of 0.897) or the Lenke classification (interobserver agreement of 60.5% with a κ of 0.560, intraobserver agreement of 81.8% with a κ of 0.796).

Many computerized systems have been developed to improve the accuracy, convenience, or repeatability of diagnostic classifications [6–9]. For idiopathic scoliosis, some studies have developed computer-aided methods to classify spinal deformity on radiographs. Stokes et al. [10] developed a computer program that applied a rule-based algorithm to automatically identified the King types based on the Cobb angles calculated from the coordinates of vertebral landmarks on the posteroanterior (PA) radiograph. Their approach required the manual identification of numerous landmarks (70 landmarks per radiograph). The inherent variability in landmark identification might result in unreliability of the measurement and classification. Phan

et al. [11] proposed a rule-based program for the Lenke-type classification.

All these classification methods heavily rely on the Cobb angle measured from radiographs. Recently, Anitha et al. [12] proposed an automated system to detect the vertebral endplates by using a customized filter that was a combination of anisotropic, sigmoid and differential filters. Automatic identification of these endplates by their system reduced the variability of the Cobb angle measurement. Utilizing curve fitting methods to determine the spinal curvature, Sardjono et al. [13] implemented the automatic Cobb measurement with the best results, namely, a mean absolute error of 3.91° . In our previous study [14], a computer-aided approach that automatically measured vertebral inclination was developed to improve the reliability of the Cobb measurement. We achieved average intraobserver and interobserver errors of $<2^\circ$ and 3° , respectively. Based on the computer-aided Cobb measurement, we have developed computer-aided systems to improve the reliability of the King classification [15] and the Lenke classification [16]. With the computer-aided King system, the average intraobserver and interobserver κ values were improved from 0.75 to 0.90 and from 0.66 to 0.86, respectively [15]. With the computer-aided Lenke system, the average intraobserver and interobserver κ values were improved from 0.77 to 0.89 and from 0.65 to 0.81, respectively [16]. In the present study, we propose a computer-aided PUMC classification system using the computerized Cobb measurement to improve the reliability of the PUMC classification.

2 Methods

In this study, radiographs obtained from 65 patients (51 girls, 14 boys; 12.5 ± 3.5 years of age) with idiopathic scoliosis were used. The selection criteria were (1) diagnosis of idiopathic scoliosis, (2) age between 9 and 18 years, (3) no prior spine surgery, (4) Cobb angle $<90^\circ$, and (5) visibility of the pelvis and T1–L5 vertebral levels on radiographs. The exclusion criteria were other musculoskeletal or neurological disorders. Informed consent was obtained from all patients/parents. The institutional review board of Yunnan University approved the study.

In the PUMC classification, the most important measurement is the Cobb angle. The Cobb angle can be manually measured by calculating the angle between the lines drawn along the upper endplate of the superior end-vertebra and the lower endplate of the inferior end-vertebra. Studies of intra- and interobserver variability of manual Cobb measurements have revealed that the errors are typically $\pm 5^\circ$ [17], which introduces variability into the classification method. In our previous study [14], an

approach based on the fuzzy Hough transform (FHT) was developed for Cobb measurements of a spinal curve. We used this approach in this study to identify the inclination of each vertebral endplate for the automatic PUMC classification. This technique is described in the following subsection.

2.1 Measurement of vertebral endplate inclination

We normalized each radiograph that contained vertebrae from T1 to L5 to a standard height of 1000 pixels. After image enhancement by histogram equalization, the user successively selected vertebrae from T1 to L5 by clicking the mouse on each vertebra. Once the user clicked on a vertebra, an initial rectangle of 100×80 pixels was created and displayed. The user could adjust the rectangle to fit the vertebra by clockwise or anticlockwise rotation, magnification or minification as well as up, down, left, or right movement of the rectangle. As an example, Fig. 1 shows the selected rectangles on a PA radiograph. The user-assigned rectangle defined the region of interest (ROI) for a vertebra.

For each ROI, the Canny edge detector was used to obtain the required edge image for the FHT. Figure 2 shows an ROI (Fig. 2a) and the edge image of the ROI (Fig. 2b). It can be seen in Fig. 2b that the vertebral radiograph contains a complex background due to various types of artifacts. Therefore, it was necessary to delete the artifacts and noise. According to the user-assigned rectangle an inner rectangle and an outer rectangle were automatically defined in the algorithm, as the rectangles shown in Fig. 2a. The distance between the inner rectangle and the user-assigned rectangle was $1/6$ of the width of the user-assigned rectangle, and the distance between the outer rectangle and the user-assigned rectangle was $1/8$ of the width of the user-assigned rectangle. The ROI just covered the outer rectangle. The noise and artifacts inside the inner rectangle and outside the outer rectangle were then deleted from the ROI region. Figure 2c shows the edge image with the noise and artifacts deleted. The FHT was then utilized on each edge image of the ROI. In the Hough space, the vertebral shape priors were used to select the candidate peaks corresponding to the lines that best fitted to a vertebra (i.e., a pair of endplates and a pair of vertical edges): (1) the distance between two endplates of a vertebra was in the range of 30–60 pixels, and the distance between two vertical edges was in the range of 40–80 pixels; (2) the average angle of two endplates was tilted $<45^\circ$, and the average angle of the vertical edges was between 45° and 90° ; (3) the angle difference between two endplates or two vertical edges of a vertebra was $<10^\circ$; and (4) the endplates and the vertical edges were close to being perpendicular to each other. Using these shape priors, four lines fitted to a

◀**Fig. 1** A spinal radiograph with user-assigned *rectangles* and central sacral *vertical line*

vertebra were identified by selecting peaks in the Hough space. Figure 2d shows the detected lines for the edge image shown in Fig. 2c. More details of this technique were described by Zhang et al. [14]. The inclination of each line that was fitted to a vertebral endplate was recorded automatically. For each vertebra, the rectangle that consisted of the four detected lines, i.e., the vertebral rectangle, was also recorded.

2.2 Computer-aided PUMC classification

The PUMC classification recognizes three main types according to the number of apexes: type I for one, type II for two, and type III for three apexes. There are a number of subtypes for each curve type, with a total of 13 subtypes [4], depending on the characteristics of 3D deformities and the flexibility of the curvature. For type I, subtype Ia has a thoracic curve; Ib has a thoracolumbar curve; and Ic has a lumbar curve. For type II, subtype IIa has double thoracic curves; IIb has a thoracic curve plus a thoracolumbar/lumbar curve, with the former being at least 10° higher than the latter; IIc has a thoracic curve plus a thoracolumbar/lumbar curve, with the curve magnitude difference being $<10^\circ$; and IId has a thoracic curve plus a thoracolumbar/lumbar curve, with the former being at least 10° smaller than the latter. For type III, there are two subtypes. The distal curve of subtype IIIa must meet all 4 criteria: (1) without lumbar kyphosis, (2) a Cobb angle of the lumbar curve $\leq 45^\circ$, (3) rotation of the lumbar curve being $<2^\circ$, and (4) flexibility of the lumbar curve $\geq 70\%$. According to the curve flexibility, rotation of the apical vertebra, Cobb angle on the convex bending radiograph, and thoracolumbar/lumbar kyphosis on the sagittal radiograph, subtypes IIb, IIc, and IId can be further divided into 2, 3, and 2 subtypes, respectively: IIb1 and IIb2 of IIb; IIc1, IIc2, and IIc3 of IIc; and IId1 and IId2 of IId.

To implement the PUMC classification, the user assigned two symmetrical landmarks on the sacrum of the PA radiograph by mouse clicks. According to the accepted definitions [18], the lateral edge of the radiograph was used to define the vertical direction. The central sacral vertical line (CSVL) was determined as the vertical line passing through the midpoint of the two sacral landmarks, as shown in Fig. 1. The apical vertebrae were then identified based on the detected endplate inclinations and vertebral positions. The apical region that was first located included the two vertebrae above and the two vertebrae below a disc separating two vertebrae tilting in opposite directions to the

Fig. 2 Vertebral endplates detection: **a** region of interest (ROI) with the *inner* and *outer* rectangles, **b** edge image of the ROI, **c** edge image with noises deleted, **d** detected endplates

horizontal. The vertebra in this region that had the greatest horizontal distance from the CSVL was then identified as the apical vertebra. Based on the detected apical vertebra, the upper vertebra with the greatest inclination of the superior endplate and the lower vertebra with the greatest inclination of the inferior endplate were selected as the two end-vertebrae for a curve. The angle between the superior endplate of the upper end-vertebra and the inferior endplate of the lower end-vertebra was automatically calculated as the Cobb angle.

According to the PUMC classification criteria, the computerized classification algorithm was implemented, as shown in Fig. 3, to perform the automatic PUMC classification. After the Cobb angles on the standing radiograph were obtained, the system automatically differentiated Ia, Ib, Ic, IIa, IIb, IIc, IId, and III. To further classify the subtypes of IIb (IIb1 and IIb2), IIc (IIc1, IIc2, and IIc3), IId (IId1 and IId2), and III (IIIa and IIIb), the rotation of the apical vertebra together with the parameters of flexibility and kyphosis that were calculated from the Cobb angle in the convex bending and sagittal radiographs could be determined. After the Cobb measurement, when these parameters were input into the computer, it automatically further classified the subtypes of IIb, IIc, IId, and III.

2.3 Evaluation

In this study, two groups of observers participated in the experiments. The first group included two pediatric orthopedic surgeons with 12 and 10 years of experience in a scoliosis clinic. The second group included two orthopedic residents. With the aid of the computer, each observer performed the PUMC classification twice over a period of three weeks. Without aid from the computer, the four observers also classified spinal curve patterns twice using the traditional PUMC method [4]. The kappa statistic in SPSS software (SPSS Inc., Chicago, IL) was used to assess the reliability of the PUMC classification under the conditions with and without the aid of the computer. Under each of these two conditions, the κ values were calculated for paired sets of classifications by each observer (intraobserver repeatability) and between observers (interobserver reliability) using all combinations of paired observations. The resulting values were averaged over combinations of pairs (intraobserver or interobserver) to provide an overall assessment of reliability.

Cases that were consistently classified by the two surgeons twice without the computer's aid were used as the ground truth. The surgeons' first trials with and without the

Fig. 3 Flowchart of the PUMC classification rule

computer's aid were compared to evaluate the accuracy of the proposed system.

3 Results

In the experiments, we only required the observers to differentiate cases into subtypes Ia, Ib, Ic, IIa, IIb, IIc, IId, IIIa, and IIIb because for some cases, only the standing radiographs were available. Although the proposed Cobb measurement can be applied to the convex bending and sagittal radiographs, and the functions of further classification were implemented in the system, further classification of these subtypes was not tested in the experiments. In

addition, no cases were classified as type III by any of the observers.

The intraobserver reliabilities are shown in Table 1. With the computer's aid, the κ values were in the range from 0.90 to 0.94, which were in the excellent range (≥ 0.90). For each observer, the κ value was improved by using the computerized method. The average intraobserver κ value and classification consistency were improved to 0.93 and 95%, respectively. Table 1 also shows that larger improvements were obtained for the resident group than for the expert group.

The interobserver reliabilities are shown in Table 2. With the computer's aid, the average interobserver κ value

Table 1 Intraobserver reliability

Reliability	With aid		Without aid	
	Consistency (%)	κ	Consistency (%)	κ
Expert group				
Expert 1	96	0.94	93	0.91
Expert 2	96	0.94	92	0.90
Average	96	0.94	93	0.91
Resident group				
Resident 1	95	0.92	86	0.83
Resident 2	93	0.90	81	0.80
Average	94	0.91	84	0.82
All observers				
Average	95	0.93	88	0.86

Table 2 Interobserver reliability

Reliability	With aid		Without aid	
	Consistency (%)	κ	Consistency (%)	κ
Expert group				
Trial 1	91	0.89	85	0.83
Trial 2	92	0.90	84	0.82
Average	92	0.90	85	0.83
Resident group				
Trial 1	86	0.83	79	0.75
Trial 2	88	0.86	81	0.78
Average	87	0.85	80	0.77
All observers				
Trial 1	88	0.85	78	0.74
Trial 2	88	0.86	80	0.76
Average	88	0.86	79	0.75

was improved from 0.83 to 0.90 for the expert group and from 0.77 to 0.85 for the resident group. For all observers, the average interobserver consistency was improved from 79 to 88% and the average κ value was improved from 0.75 to 0.86. The results also demonstrated that the use of the computer reduced the difference between experts and residents in the classification.

Without the computer's aid, two surgeons consistently classified 53 cases in two trials. For the 53 cases, the surgeons' first trials with and without the aid of the computer were compared. The results showed that under the two conditions, one surgeon's classifications were consistent in all cases (100% accuracy) and the other surgeon's classifications were consistent in 52 cases (98% accuracy). The one inconsistent case was due to inconsistent classification between types IIc and IIc.

To evaluate the importance of the repeatability of the Cobb measurement and objective judgments in the classification, we used the Cobb angles manually measured by the one surgeon (with 12 years of experience) as the standard value. With the standard values of the Cobb angles, the system resulted in consistent classification in all cases. Without the computer's aid, the two residents were asked to make classifications based on the standard Cobb angles. Two residents' classifications were consistent in 63 cases (97% consistency). The average judgment time of residents for one case was 35 s. The results demonstrated the importance of the Cobb measurement in the classification. In addition, the system can aid in the decision. If given the parameters, the system could perform consistent classifications in real time.

4 Discussion

The reliability of spinal deformity classification is an important topic in the orthopedic community. Although the PUMC classification has higher reliability than the King or Lenke classifications [5], improving reliability in preoperative surgical planning is still necessary. The variability is mainly due to human technical and judgment errors. This article proposes a computer-aided system to reduce variability in the PUMC classification. The results of this study indicated that the tasks of measuring scoliosis curves on radiographs and subsequently classifying curve types were made more reliable with the aid of the proposed system. First, the reliability of identifying the radiographic features that provided the measurements used for classification was improved by the computer-aided Cobb measurement method. In our previous study [14], the accuracy and reliability of the proposed Cobb measurement had been demonstrated ($<3^\circ$, comparable to the 3.91° that was

obtained with the system from Sardjono et al. [13]). The higher repeatability of the Cobb measurement results in the higher reliability of the PUMC classification. Second, judgment errors were reduced by using a computerized classification algorithm. With the computer's aid, the variability produced by the interpretation and memory of the classification criteria was avoided. Although the proposed method still requires user interaction, only limited user skills are required. The results show that the system can be used equally well by individuals with less clinical experience. Because the only hardware requirement was a personal computer and the software environment was Visual C++, the developed software could be easily implemented in a clinic.

Scoliosis classification is used to facilitate the objective assessment of scoliosis, i.e., to make it possible to "speak the same language" in the assessment. There is no gold standard for the classification results before surgery. Therefore, the accuracy of the system was evaluated by comparing the surgeons' classifications under the conditions with and without the aid of the computer using cases that were consistently classified by two surgeons twice without the computer's aid as the gold standard. The comparative results demonstrated the accuracy of the proposed method (accuracy of more than 98%).

The results of the consistent classifications under the given standard Cobb angles demonstrated the importance of the repeatability of the Cobb measurement in the classification. In the experiments, the further classification of the subtypes of IIb, IIc, and IIc was not tested, which might be one of the reasons for the good performance of two residents even without the computer's aid (97% consistency), although it took more than half a minute for judgment. Another reason might be due to the given Cobb angles which reduced the effort required in the PUMC classification. When the Cobb measurement and further classifications are required, the system would play a more important role in facilitating the classification.

In conclusion, the proposed computerized system may be of value to assist a surgeon in the PUMC classification of scoliosis. It can reduce the Cobb measurement variability and human judgment errors in the PUMC classification. The computer's aid was especially useful for inexperienced observers.

Acknowledgements The research work was supported by the grants from the Natural Science Foundation of China under Grants 61002049 and 61361010, and from the Key Program of the Yunnan Natural Science Foundation under Grant 2013FA008.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Stokes IAF. Three-dimensional terminology of spinal deformity. A report presented to the scoliosis research society by the scoliosis research society working group on 3-D terminology of spinal deformity. *Spine*. 1994;19(2):236–48.
2. King HA, Moe JH, Bradford DS, Winter RB. The selection of fusion levels in thoracic idiopathic scoliosis. *J Bone Joint Surg Am*. 1983;65(9):1302–13.
3. Lenke LG, Betz RR, Harms J, Bridwell KH, Clements DH, Lowe TG, Blanke K. Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. *J Bone Joint Surg Am*. 2001;83-A(8):1169–81.
4. Qiu G, Zhang J, Wang Y, Xu H. A new operative classification of idiopathic scoliosis a Peking Union Medical College method. *Spine*. 2005;30(12):1419–26.
5. Qiu Y, Zhu Z, Zhu F, Li W, Wang W, Wang B, Yu Y, Qian B. Comparison of interobserver and intraobserver reliability between the King, Lenke and PUMC classification systems for adolescent idiopathic scoliosis. *Chin J Orthop*. 2007;27(10):748–52.
6. Sahu O, Anand V, Kanhangad V, Pachori RB. Classification of magnetic resonance brain images using bi-dimensional empirical mode decomposition and autoregressive model. *Biomed Eng Lett*. 2015;5(4):311–20.
7. Choi JY. A generalized multiple classifier system for improving computer-aided classification of breast masses in mammography. *Biomed Eng Lett*. 2015;5(4):251–62.
8. Mohammadi MR, Khaleghi A, Nasrabadi AM, Rafieivand S, Begol M, Zarafshan H. EEG classification of ADHD and normal children using non-linear features and neural network. *Biomed Eng Lett*. 2016;6(2):66–73.
9. Raj S, Maurya K, Ray KC. A knowledge-based real time embedded platform for arrhythmia beat classification. *Biomed Eng Lett*. 2015;5(4):271–80.
10. Stokes IAF, Aronsson DD. Computer-assisted algorithms improve reliability of King classification and Cobb angle measurement of scoliosis. *Spine*. 2006;31(6):665–70.
11. Phan P, Mezghani N, Nault ML, Aubin CE, Parent S, de Guise J, Labelle H. A decision tree can increase accuracy when assessing curve types according to Lenke classification of adolescent idiopathic scoliosis. *Spine*. 2010;35(10):1054–9.
12. Anitha H, Karunakar AK, Dinesh KVN. Automatic extraction of vertebral endplates from scoliotic radiographs using customized filter. *Biomed Eng Lett*. 2014;4(2):158–65.
13. Sardjono TA, Wilkinson MH, Veldhuizen AG, van Ooijen PM, Purnama KE, Verkerke GJ. Automatic Cobb angle determination from radiographic images. *Spine*. 2013;38(20):E1256–62.
14. Zhang J, Lou E, Shi X, Raso JV, Wang Y, Le LH, Shi X. Computer-aided assessment of scoliosis on posteroanterior radiographs. *Med Biol Eng Comput*. 2010;48(2):185–95.
15. Zhang J, Li H, Lv L, Shi X, Zhang Y. Computer-aided King classification of scoliosis. *Technol Health Care*. 2015;23(s2):s411–7.
16. Zhang J, Li H, Lv L, Shi X, Guo F, Zhang Y. A computer-aided method for improving the reliability of Lenke classification for scoliosis. *J Healthc Eng*. 2015;6(2):145–58.
17. Pruijs JEH, Hageman MAPE, Keessen W, Meer R, Wieringen JC. Variation in Cobb angle measurements in scoliosis. *Skelet Radiol*. 1994;23(7):517–20.
18. Stokes IAF, Aronsson DD. Identifying sources of variability in scoliosis classification using a rule-based automated algorithm. *Spine*. 2002;27(24):2801–5.