The following scientific article was officially published in the Proceedings of the International Conference on Image Analysis and Recognition in 2010, in book form in the *Lecture Notes in Computer Science* series published by Springer Verlag. This article's citation is as follows:

The final publication is available at Springer via **http://dx.doi.org/10.1007/978-3-642-13775-4_2**

The manuscript, as accepted by the publisher, is reproduced in the following pages.

Towards Non Invasive Diagnosis of Scoliosis using Semi-Supervised Learning Approach

Lama Seoud[1], Mathias M. Adankon[1], Hubert Labelle[2], Jean Dansereau[1], and Farida Cheriet[1]

[1] École Polytechnique de Montréal,
lama.seoud@polymtl.ca, mathias-mahouzonsou.adankon@polymtl.ca,
jean.dansereau@polymtl.ca, farida.cheriet@polymtl.ca
[2] Sainte Justine Hospital Research Center, Montréal, Québec, Canada
hubert.labelle@recherche.ste-justine.qc.ca

Abstract. In this paper, a new methodology for the prediction of scoliosis curve types from non invasive acquisitions of the back surface of the trunk is proposed. One hundred and fifty-nine scoliosis patients had their back surface acquired in 3D using an optical digitizer. Each surface is then characterized by 45 local measurements of the back surface rotation. Using a semi-supervised algorithm, the classifier is trained with only 32 labeled and 58 unlabeled data. Tested on 69 new samples, the classifier succeeded in classifying correctly 87.0% of the data. After reducing the number of labeled training samples to 12, the behavior of the resulting classifier tends to be similar to the reference case where the classifier is trained only with the maximum number of available labeled data. Moreover, the addition of unlabeled data guided the classifier towards more generalizable boundaries between the classes. Those results provide a proof of feasibility for using a semi-supervised learning algorithm to train a classifier for the prediction of a scoliosis curve type, when only a few training data are labeled. This constitutes a promising clinical finding since it will allow the diagnosis and the follow-up of scoliotic deformities without exposing the patient to X-ray radiations.

1 Introduction

Scoliosis is a three-dimensional deformity of the spine and the ribcage that affects the general appearance of the trunk. In general, one of the first symptoms of scoliosis is the manifestation of a hump on the back, called the rib hump. It constitutes one of the most disturbing aspects of the deformity for the patients.

The management of scoliosis depends essentially on the severity, the type and the risk of progression of the curve. Those parameters are commonly evaluated on standard frontal and lateral X-rays of the patient's trunk in upright position. However, there are several limitations attributed to the radiographic evaluation of scoliosis. First of all, it provides only bi-dimensional information that is not sufficient to fully evaluate a complex three-dimensional pathology like scoliosis. Second, only the internal deformities can be evaluated in the radiographs while the patients' main concern is their external appearance. And last but not least, as X-ray acquisition is invasive, and considering the risks associated with radiation exposure from repeated radiographs, its frequency

is limited to every 6 months which represents a long interval for the follow-up of a progressive scoliosis.

In order to evaluate the scoliosis on a more frequent basis, non invasive imaging techniques that provide a three-dimensional reconstruction of the trunk surface have been proposed in the literature. The main challenge currently is to relate the topographic measurements with the radiographic ones. In this context, several authors have tried to predict the severity of scoliosis [1–4] or the 3D shape of the spine [5] from metric evaluations on the surface of the back or of the trunk, using statistical methods [4] or machine learning techniques like neural networks [2] and supervised support vector machines [3, 5]. To build such classifiers, a large set of labeled samples, called the training data, is necessary. As the labeling is based on radiographic measurements, the size of the training database is thus limited by the X-rays acquisition frequency. This affects negatively the performance of the classifiers.

Since the back surface acquisition is totally non-invasive, it would be advantageous to complement the training database with topographic data that is not necessarily labeled, which means that no X-ray information is available for this data. Training using both labeled and unlabeled data is called semi-supervised learning. Recently, one such approach has been proposed in the literature [6] and its effectiveness has been tested on artificial data as well as on real problems such as the character recognition. The results have shown the usefulness and the good performance of this method even when the number of labeled data is too small.

The aim of the current study is to prove the feasibility of using a semi-supervised learning approach for the prediction of scoliosis curve types by analysing the back surface of the trunk.

2 3D Back surface analysis

2.1 Data acquisition

Currently, at Sainte-Justine Hospital Research Centre (SJHRC) in Montreal (Canada), the back surface of the trunk is acquired using an optical digitizer (InSpeck Inc., Montreal, Canada), comprised of a color CCD camera and a structured light projector. The acquisition process consists of projecting successively four fringe patterns, obtained by phase-shifting a set of light fringes, onto the surface. Based on the four resulting images, the system computes, by interferometry and triangulation, the depth of each surface point relative to the reference plane of the digitizer. A fifth image, with no fringes, acquires the texture of the surface which is then mapped onto the 3D model (figure 1).

The resulting mesh consists of more than 10,000 nodes, depending on the size of the patient.The accuracy of this system was evaluated in [7], using markers placed on a mannequin whose coordinates were previously recorded by a computer measuring machine. The results showed a reconstruction accuracy of 0.56 mm over the back.

During the acquisition, the patient is in upright position, with his arms slightly abducted on the sides. Prior to the acquisition, a nurse locates by palpation and places markers over several anatomic landmarks on the trunk, such as the center of the posterior superior iliac spines (CPSIS) and the vertebral prominence (VP). These markers are used for clinical measurements and registration.
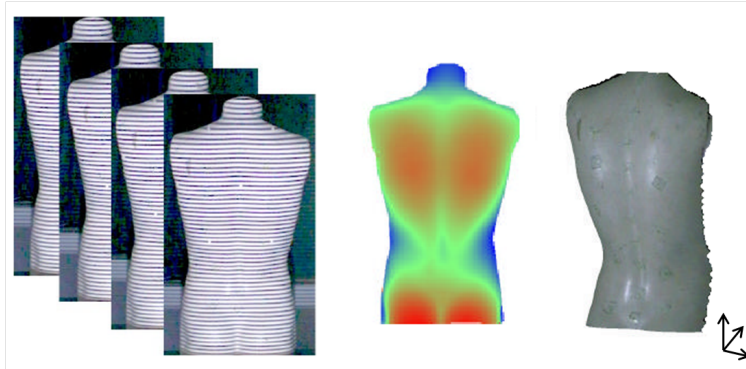
Fig. 1. On the left: four fringe patterns are projected on the back surface of a mannequin. In the center: color representation of depth. On the right: 3D reconstruction of the surface with the mapped texture.

2.2 Features extraction

To evaluate the scoliosis deformities on the surface of the trunk, local measurements are computed on horizontal cross-sections of the trunk. More specifically, 50 sections, equally spaced along the vertical axis of the trunk, are automatically extracted starting from the CPSIS and going up to the VP (Figure 2). On each of the 50 sections, an automatic algorithm computes the back surface rotation (BSR). This clinical measurement is related to the amplitude of the rib hump and is defined as the angle, projected onto the axial plane, between the dual tangent to the section and the X-axis of the acquisition reference frame. The BSR can be negative or positive depending on the side of the hump. The accuracy of this measurement computed on a 3D reconstruction of the trunk surface was previously evaluated at 1.4mm [8].

Thus, each back is characterized by 50 BSR values. In order to filter outliers and obtain smoother value sets, an averaging window was applied to each set. Moreover, for each patient, the angle values were normalized between -100 and 100 degrees to compensate for differences between the patients in term of severity. Finally, the BSR values corresponding to the upper 5 sections were not considered because they were too noisy and the values of the BSR are not relevant in this area in the context of scoliosis assessment.

3   Semi-supervised learning

Pattern recognition problems are solved with classifiers which are designed using prototypes of the data to be recognized. This data, called the training set, consists of the patterns and their labels (the category of the pattern). This is the supervised learning where the features extracted from the patterns and their labels are used for modeling the classifier parameters [9, 10]. However, the labeling process can become extremely expensive and cumbersome. For instance, the labeling of handwritten documents, images, or web pages requires both human expertise and insight, whereas in the field of
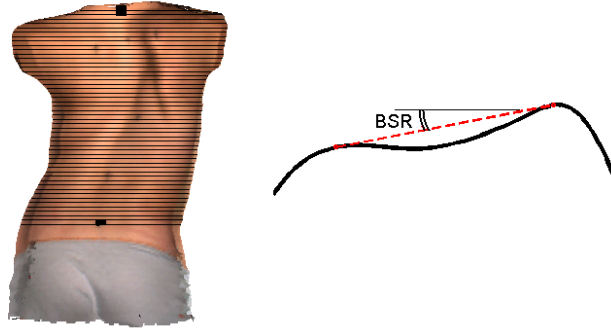
Fig. 2. On the left: 50 horizontal sections (black lines) of the trunk are extracted between the CPSIS (black squared marker on the bottom) and the VP (black squared marker on the top). On the right: the BSR is the angle defined by the dual tangent (red dashed line) to the back section (black curve) and the projection of the lateral axis onto the axial plane (black line).

medicine or biology, the labeling process may require some data acquisitions that are limited for ethical reasons. Thus, it may be very difficult, or even impossible, to label all the available data. The alternative is the semi-supervised learning [11–13], where both labeled and unlabeled data are used to train the classifier. Hence, it is not necessary to label all the data collected in order to build the classifier.

In this work, we propose using the least squares support vector machine (LS-SVM) in order to classify scoliosis curve types. To this end, we have collected labeled and unlabeled data in order to train our machine in semi-supervised mode.

The LS-SVM is an interesting variant of the SVM proposed by Suykens et al.[14, 15]. The standard Vapnik SVM classifier [16] is modified to transform the QP problem into a linear one. These modifications are formulated in the LS-SVM definition as follows, when we consider a binary classification problem with $\{(x_1, y_1), ..., (x_\ell, y_\ell)\}$ the training dataset, $x_i \in \mathbb{R}$ and $y_i \in \{-1, 1\}$.

$$\min_{w,b,\xi} \frac{1}{2}w'w + \frac{1}{2}C \sum_{i=1}^{\ell} \xi_i^2 \qquad (1)$$

$$\text{s.t.} \quad \xi_i = y_i - [w'\phi(x_i) + b] \quad \forall i = 1, ..., \ell \qquad (2)$$

where w' denotes the transpose of w, ϕ is the mapping function used implicitly via the kernel function $k(x_i, x_j) = \phi(x_i).\phi(x_j)$ for non linear problems, C is used to balance the trade-off between maximizing the margin and minimizing the training error quantified by the variable ξ.

The original SVM formulation is modified at two points. First, the inequality constraints with the slack variable ξ_i expressed in SVM formulation are replaced by equality constraints. Second, a squared loss function is considered in the objective function. These two essential modifications simplify the problem, which becomes linear.

The Lagrangian of problem (2) is expressed by :

$$\mathcal{L}(w, b, \xi, \alpha) = \frac{1}{2}w'w$$

$$+\frac{1}{2}C\sum_{i=1}^{\ell}\xi_i^2 - \sum_{i=1}^{\ell}\alpha_i\{y_i - [w'\phi(x) + b] - \xi_i\}$$

where α_i are Lagrange multipliers.

The Karush$-$Kuhn$-$Tucker (KKT)[3] conditions for optimality yield

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{\ell}\alpha_i\phi(x_i) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^{\ell}\alpha_i = 0 \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \Rightarrow \alpha_i = C\xi_i, \quad \forall i = 1, ..., \ell \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \Rightarrow \xi_i = y_i - [w'\phi(x_i) + b] \quad \forall i = 1, ..., \ell \end{cases}$$

We note that the system coming from the KKT conditions is linear, and that its solution is found by solving the system of linear equations expressed in the following matrix form :

$$\begin{pmatrix} K + C^{-1}I & \mathbf{1}' \\ \mathbf{1} & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{3}$$

where :
$K_{ij} = k(x_i, x_j)$
$Y = (y_1, ..., y_\ell)'$
$\alpha = (\alpha_1, ..., \alpha_\ell)'$
$\mathbf{1} = (1, ..., 1)$

In our previous work [6], we have proposed the semi-supervised LS-SVM ($S^2LS-SVM$) using the following expressions:

$$\min_{w, b, \xi, \xi^*, y_1^*, ..., y_n^*} \frac{1}{2}w'w + \frac{1}{2}C\sum_{i=1}^{\ell}\xi_i^2 + \frac{1}{2}C^*\sum_{j=1}^{n}(\xi_j^*)^2 \tag{4}$$

$$\text{s.t.} \quad \xi_i = y_i - [w'\phi(x_i) + b], \quad \forall i = 1, ..., \ell \tag{5}$$

$$\xi_j^* = y_j^* - [w'\phi(x_j^*) + b], \quad \forall j = 1, ..., n \tag{6}$$

$$y_j^* \in \{-1, 1\} \quad \forall j = 1, ..., n \tag{7}$$

In this equation, the parameters C and C^* balance the error between the labeled $\{(x_1, y_1), ..., (x_\ell, y_\ell)\}$ and unlabeled data $\{x_1^*, ..., x_n^*\}$.

Considering the combinatorial view of the optimization problem (4), the variables w, b, ξ, ξ^* and y^* are optimized at different levels. Then, for a given fixed set $y_1^*, \ldots, y_n^*$,

---

[3] KKT conditions are necessary conditions for optimality obtained from first derivative.

6

the optimization over (w, b) is standard LS-SVM training, and we obtain a linear system in dual space expressed in matrix form by:

$$\begin{pmatrix} K + \Gamma & \mathbf{1}' \\ \mathbf{1} & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{8}$$

where:
$K_{ij} = k(x_i, x_j)$
$Y = (y_1, ..., y_\ell, y_1^*, ..., y_n^*)'$
$\alpha = (\alpha_1, ..., \alpha_\ell, \alpha_1^*, ..., \alpha_n^*)'$
$\mathbf{1} = (1, ..., 1)$
Γ is a diagonal matrix with $\Gamma_{ii} = 1/C$ for $i = 1, ..., \ell$ and $\Gamma_{ii} = 1/C^*$ for $i = \ell + 1, ..., \ell + n$

Two methods are proposed in [6] for solving the semi-supervised problem expressed in (4). In this paper, we used the second approach which is described as follows.

The unlabeled examples are labeled gradually during the learning process: one sample is labeled and added to the labeled set. The added sample is chosen in order to obtain the smallest increase in the objective function. The criterion we use to select this point is based on the value of α_j^*; because considering the equation $\alpha_j^* = C^* \xi_j^*$, it is clear that the error is proportional to the value of α^*.

First, we identify the label of the point to be labeled according to the objective function. Next, for each remaining unlabeled sample, we compute $a_j^{(1)} = \alpha_j^*$ if the identified label is 1 and $a_j^{(-1)}$ for the opposite label. As the goal is to find the unlabeled sample x^* with the smallest increase in the objective function, we select, at each step, the unlabeled sample, the corresponding α_j^* of which will be the smallest if it is added to the previous solution. We repeat this procedure until all unlabeled samples are labeled.

4   Dataset and experimental setup

This study was conducted on a cohort of 159 adolescents with scoliosis who were candidates for surgery. Among the cohort, 101 patients had their topographic and radiographic acquisitions done at the same visit. Based on the radiographic measurements of each of those patient, the scoliosis curve type was determined according to the common clinical classification that distinguishes between 4 types of curves: thoracic major curves, thoracolumbar major curves, lumbar major curves and double major curves. The number of lumbar major curves being too small (4/101), we mixed them with the thoracolumbar major curves (18/101), being quite similar. The distribution of the patients among the three considered classes is illustrated in Table 1. For the remaining 58 patients, their radiographs were not acquired at the same date as the trunk topography, thus they were considered unlabeled.

For all the patients, the 45 features are automatically extracted according to subsection 2.2, and four classifiers are trained distinctly. The classifiers' performance is evaluated on the same testing dataset composed of 69 labeled samples chosen quasi-randomly among the database, under the condition of having a class distribution as similar as possible to the one of the whole cohort.

Table 1. Distribution of the patients among the classes

Classes	Curve type	Number of patients
Class 1	Thoracic major curve	45
Class 2	Double major curve	34
Class 3	Lumbar or thoracolumbar major curve	22
Unlabeled	Unknown	58
Total		159

A first classifier (C-SSL1) is built using the semi-supervised learning algorithm as described in section 3 and using a total of 90 training data: the 58 unlabeled samples and the remaining 32 labeled ones. A second classifier (C-SL1) is trained using a supervised LS-SVM and a training database made of the same 32 labeled data as for the C-SSL1. C-SL1 constitutes the reference case since it is trained with all the available labeled data.

In order to compare the supervised and the semi-supervised learning when only a small amount of training data are labeled, a third classifier (C-SSL2) is built using the semi-supervised learning as described in section 3 and using, for training, only 12 labeled data (5 of class 1, 4 of class 2, 3 of class 3) chosen pseudo-randomly (under the condition of having a class distribution among the retained data as similar as possible to the one of the whole cohort) and the remaining 78 training samples considered unlabeled. A fourth classifier (C-SL2) is trained using a supervised LS-SVM algorithm and the same 12 labeled data as in C-SSL2 are considered for training.

Since, we have a multi-class problem, each classifier is built by training three machines using the *one-against-all* strategy. We used a radial basis function (RBF) kernel and performed model selection with cross validation procedure [17].

5   Results and Discussion

Table 2 presents the prediction rates obtained in testing each of the four classifiers. These results show first that the performance of the classifiers trained using the semi-supervised algorithm (C-SSL1 and C-SSL2), compared respectively to C-SL1 and C-SL2, is significantly improved by the addition of unlabeled data. Second, the performance of C-SSL2, trained with only 12 labeled samples tends to be equal to the one of the ideal classifier C-SL1, that is built using the maximum number of available labeled samples. The latter outcome answers to the main goal of the semi-supervised learning: even with a few labeled data (12 among 90 training samples), the generalisation capacity of the classifier C-SSL2 is similar to the reference case.

Table 3 illustrates the confusion matrices obtained in testing the four classifiers. It shows first that all the classifiers clearly distinguish between patterns of class 3 and class 1. The major confusion is between classes 1 and 2 and between classes 2 and 3. This is also illustrated by the plot of the mean BSR values for each class (figure 3). In fact, for some double major curves (class 2) the thoracic hump is more accentuated than the lumbar hump which results in a pattern that is quiet similar to the thoracic major

Table 2. Learning and testing databases of the four classifiers and the prediction rates in testing

Classifiers	Training data		Testing data	Prediction rate
	Lab.	Unlab.		
CSSL1	32	58	69	87.0%
CSL1	32	-	69	82.6%
CSSL2	12	78	69	79.7%
CSL2	12	-	69	75.4%

curves (class 1). The same logic follows in the case of some double major curves (class 2) where the lumbar hump is more prominent than the thoracic hump which results in a pattern that is quiet similar to the lumbar major curves (class 3). Nevertheless, it seems that, with the semi-supervised learning, the unlabeled samples guide the classifier towards more generalizable boundaries for class 2 (prediction rate of 65.2% for C-SSL2 versus 56.5% for C-SL2 and 73.9% for C-SSL1 versus 65.2% for C-SL1) and for class 3 (prediction rate of 80.0% for C-SSL2 versus 66.7% for C-SL2 and 100% for C-SSL1 versus 93.3% for C-SL1). Furthermore, table 3 illustrates once again how the behavior of C-SSL2, trained with only 12 labeled samples out of the 90 training samples, tends to be similar to the reference classifier (C-SL1).

Table 3. Confusion matrices in testing the four classifiers

Predicted class	C-SSL1			C-SL1			C-SSL2			C-SL2		
	Target Class			Target Class			Target Class			Target Class		
	1	2	3	1	2	3	1	2	3	1	2	3
1	**28**	4	0	**28**	6	0	**29**	8	0	**29**	9	0
2	3	**17**	0	3	**15**	1	2	**14**	3	2	**13**	5
3	0	2	**15**	0	2	**14**	0	1	**12**	0	1	**10**
Prediction rate per class (%)	90.3	73.9	100	90.3	65.2	93.3	93.6	65.2	80.0	93.6	56.5	66.7

In this study, we considered the BSR as the only clinical index to describe the back surface deformity. This choice is based on the clinical observation that the rib hump generally appears on the convex side of each spinal curve. Our results show that the BSR, computed on 45 cross-sections of the back, is a good discriminant feature. However, in future works, other characteristics of the scoliosis deformity will be considered in order to reduce even more the misclassification rate.

Furthermore, in this work, only the back surface of the trunk is acquired. As demonstrated in the literature, measurements made on the back surface are sensitive to the patient's posture during the acquisition. To overcome this imprecision, the entire trunk's surface can be reconstructed using four optical digitizers placed all around the patient and a registration process [7]. The measurements can thus be computed in a patient specific coordinates system. Moreover, as the whole shape of the trunk is deformed in 3D,
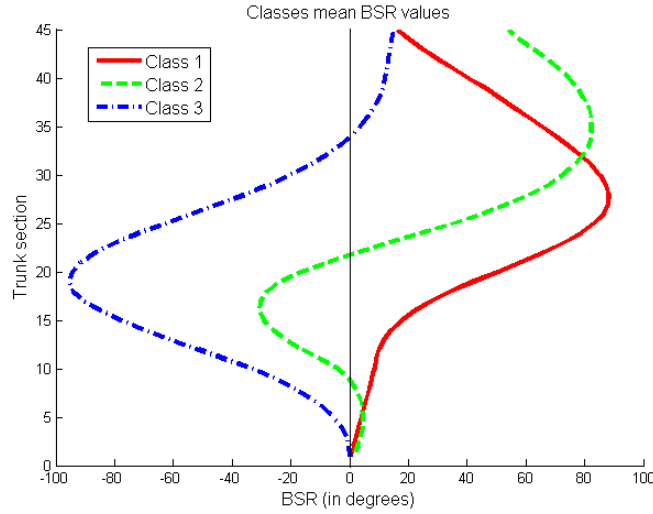
Fig. 3. Mean BSR values for each of the 45 sections of the trunk, computed for each class

it could be more interesting to consider inclined cross-sections that follows the general shape of the trunk to compute local measurements [8].

Finally, in the present paper, the back surfaces are classified in three different classes. Due to the small number of lumbar curves in the cohort, no distinction is made between lumbar and thoracolumbar major curves. However, in the presence of more lumbar curves, we could consider a fourth class in our classification. Furthermore, the current classification can be considered a high level classification and in future works, we can refine the classification's resolution by identifying features on the trunk's surface that could define distinct clusters within each class. This would be valuable in clinic since the actual scoliosis classification systems relies only on the spinal deformity and do not take into account the general appearance of the trunk which constitutes the patient's major preoccupation.

6 Conclusion

In conclusion, this preliminary study constitutes a proof of feasibility of the semi-supervised learning in the clinical context of classifying the scoliosis curve types based on the analysis of the back surface of the trunk. With only 12 labeled samples out of 90 training data, it is possible to predict the scoliosis curve type with a success rate similar to the reference case where 32 labeled data are used for supervised learning. Moreover, the unlabeled samples improve significantly the definition of the boundaries between the classes. Unlike supervised learning, there is no need to consider a large set of labeled data to build a consistent classifier with high generalization performance.

On a more clinical aspect, these results are valuable since it demonstrates that it is possible to identify the scoliosis curve type without exposing patients to ionizing

radiation. This is an important finding since it could reduce the frequency of X-ray acquisitions during the scoliosis progression follow-up.

Acknowledgments

References

1. Ajemba, P., Durdle, N., Hill, D., Raso, J.: Classifying torso deformity in scoliosis using orthogonal maps of the torso. Med. Biol. Eng. Comput. 45, 575–584 (2007)
2. Jaremko, J., Poncet, P., Ronsky, J., Harder, J., Dansereau, J., Labelle, H., Zernicke, R.: Genetic algorithm-neural network estimation of Cobb angle from torso asymmetry in scoliosis J. Biomech. Eng. 124(5), 496-503 (2002)
3. Ramirez, L., Durdle, N., Raso, J., Hill, G.: A support vector machines classifier to assess the severity of idiopathic scoliosis from surface topography. IEEE Trans. Inf. Technol. Biomed. 10, 84-91 (2006)
4. Stokes, I., Moreland, M.: Concordance of back surface asymmetry and spine shape in idiopathic scoliosis. Spine 14, 73-78 (1989)
5. Bergeron, C., Cheriet, F., Ronsky, J., Zernicke, R., Labelle, H.: Prediction of anterior scoliotic spinal curve from trunk surface using support vector regression. Eng. Appl. Artificial Intell. 18, 973-983 (2005)
6. Adankon, M., Cheriet, M., Biem, A.: Semisupervised least squares support vector machine. IEEE Trans. Neural Net. 20(12), 1858-1870 (2009)
7. Pazos, V., Cheriet, F., Song, L., Labelle, H., Dansereau, J.: Accuracy assessment of human trunk surface 3D reconstructions from an optical digitising system. Med Biol Eng Comput 43, 11-15 (2005)
8. Pazos, V., Miled, F., Debanne, P., Cheriet, F.: Analysis of trunk external asymmetry in side-bending. IRSSD, Liverpool (2008)
9. Duda, R. O., Hart, P., E., and Stork, D., G.: Pattern Classification 2nd edition, New York, John Wiley and Sons (2001)
10. Bishop, C., M. : Pattern Recognition and Machine Learning. Springer (2006)
11. Seeger, M.: Learning with labeled and unlabeled data. Inst. Adapt. Neural Comput., Univ. Edinburgh, Edinburgh, U.K., Tech. Rep. (2001)
12. Zhu, X.: Semi-supervised learning literature survey. Comput. Sci.,Univ. Wisconsin-Madison, Madison, WI, Tech. Rep. 1530, (2007) Available: http://www.cs.wisc.edu/ jerryzhu/pub/ssl survey.pdf
13. Adankon, M., Cheriet, M.: Help-Training for semi-supervised discrimininative classifier. Application to SVM. 19th Internationale Conference on Pattern Recognition, Tampa, 1-4 (2008)
14. Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. Singapore: World Scientific (2002)
15. Suykens, J. A. K., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett., vol. 9, no. 3, pp. 293-300 (1999)
16. Vapnik, V. N.: Statistical Learning Theory. New York: Wiley (1998)
17. Adankon, M., Cheriet, M.: Model selection for LS-SVM. Application to handwriting recognition. Pattern Recognit. 42(12), 3264-3270 (2009)