Wavelet Appearance Pyramids for Landmark Detection and Pathology Classification: Application to Lumbar Spinal Stenosis

Qiang Zhang[1(✉)], Abhir Bhalerao[1], Caron Parsons[2,3], Emma Helm[2], and Charles Hutchinson[2,3]

[1] Department of Computer Science, University of Warwick, Coventry, UK
qiang@dcs.warwick.ac.uk
[2] Division of Health Sciences, University of Warwick, Coventry, UK
[3] Department of Radiology,
University Hospital Coventry and Warwickshire, Coventry, UK

Abstract. Appearance representation and feature extraction of anatomy or anatomical features is a key step for segmentation and classification tasks. We focus on an advanced appearance model in which an object is decomposed into pyramidal complementary channels, and each channel is represented by a part-based model. We apply it to landmark detection and pathology classification on the problem of lumbar spinal stenosis. The performance is evaluated on 200 routine clinical data with varied pathologies. Experimental results show an improvement on both tasks in comparison with other appearance models. We achieve a robust landmark detection performance with average point to boundary distances lower than 2 pixels, and image-level anatomical classification with accuracies around 85 %.

1 Introduction

Diagnosis and classification based on radiological images is one of the key tasks in medical image computing. A standard approach is to represent the anatomy with coherent appearance models or feature descriptors, and vectorise the representations as inputs for training a classifier (Fig. 1(A)). The training data usually consists of instances with landmarks annotated at consistent anatomical features. The appearance correspondence across the instances is built by aligning a deformable appearances, e.g., Active Appearance Model (AAM) [1], or extracting local features at the landmarks [2–4]. During testing, the landmarks are detected in new, unseen instances, and the features are extracted and sent to a classifier for the pathology classification. For a robust landmark detection, a prior model of the object class is learned by formulating the statistics of the vectorised representations, and the searching is conducted under the regularisation of the prior model. The deformable model is either holistic [1], which consists of the shape and aligned appearance, or part-based [2–5], which represents an object by locally rigid parts with a shape capturing the spatial relationships among

S. Ourselin et al. (Eds.): MICCAI 2016, Part II, LNCS 9901, pp. 274–282, 2016.
DOI: 10.1007/978-3-319-46723-8_32

parts. Part-based models have shown superior performance benefiting from the local feature detection [2,3,5] and shape optimisation methods [2,4]. However less attention has been paid to optimising the appearance representation and preserving the anatomical details.

We propose a new appearance model referred to as a Wavelet Appearance Pyramid (WAP) to improve the performance of landmark detection and pathology classification, see an overview in Fig. 1(B). The object is decomposed into multi-scale textures and each scale is further decomposed into simpler parts. To achieve an explicit scale decomposition, the filter banks are designed and arranged directly in Fourier domain. The logarithmic wavelets (loglets) [6] are adopted as the basis functions of the filter banks for their superior properties, such as uniform coverage of the spectrum (losslessness) and infinite number of vanishing moments (smoothness). The scales are complementary in the Fourier domain which enables the reconstruction of the appearance from a WAP. The variations in the population can be modelled and visualised, with the deformation approximated by local rigid translations of the multi-scale parts, and the appearance changes by linear modes of the assembly of the parts.

We apply the WAP to the problem of lumbar spinal stenosis (LSS) and present an approach for fitting the landmarks and grading the central and foremenal stenosis [7,8]. The Supervised Descent Method (SDM) [2] is integrated with the WAP for landmark detection. The performance is validated on MRI data from 200 patients with varied LSS symptoms. Experimental results show an improvement in both the landmark detection and pathology grading over other models such as Active Shape Models (ASMs), AAMs [1], and Constrained Local Models (CLMs) [9][1].

2 Method

To provide a more comprehensive description of an object, we decompose the appearance into pyramidal channels at *complementary* scale ranges with wavelets, and represent each channel with a part-based model. We refer to this form of appearance models as Wavelet Appearance Pyramids. We detail the method as follows.

Explicit Scale Selection in the Fourier Domain. We start by decomposing an image I into multi-scale channels directly in the Fourier domain. When considered in polar coordinates, the Fourier spectrum $\mathcal{I}$ actually spans a scale space with larger scales at lower frequency and smaller scales spreading outwards. Therefore a multi-scale decomposition of image textures can be achieved explicitly by dividing the spectrum into subbands, see Fig. 2(b). In practice, filtering the spectrum with sharp windows will introduce discontinuities therefore causing aliasing. To design a bank of window functions which are smooth in

[1] Supplementary videos of the paper can be found at https://sites.google.com/site/waveletappearancepyramids.

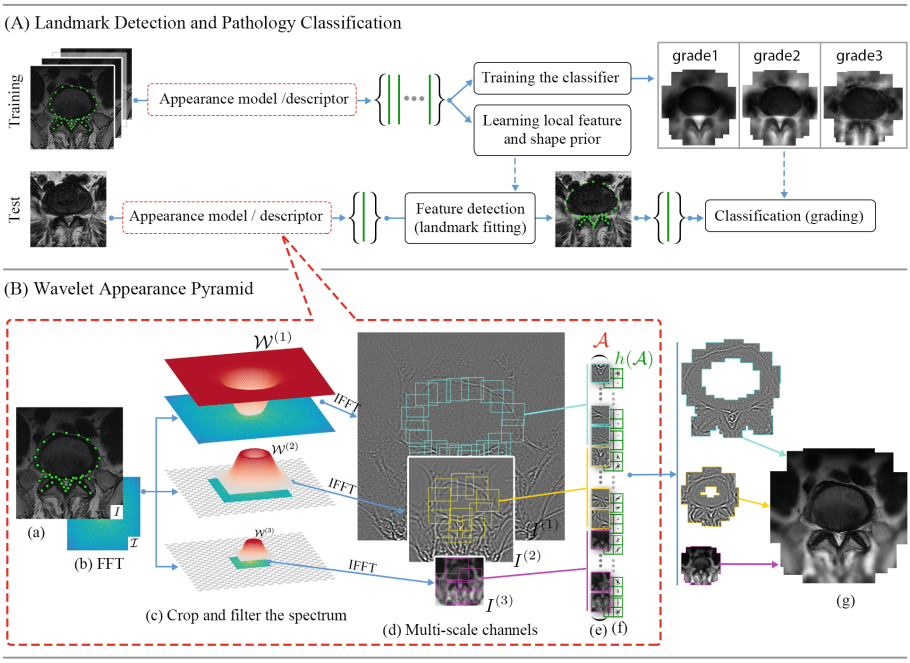

Fig. 1. (A) A standard approach of landmark detection and pathology grading. (B) The proposed appearance model ($\mathcal{A}$) and feature descriptor ($h(\mathcal{A})$).

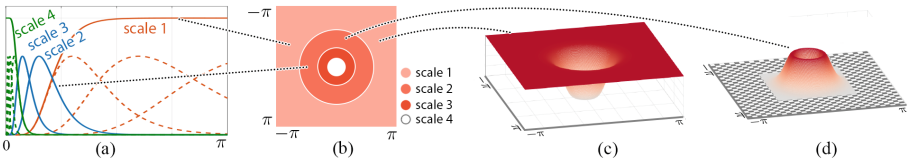

Fig. 2. (a) Radial profiles of the filters. (b) Scale selection in the Fourier domain. (c) The high pass filter in the Fourier domain. (d) The first bandpass filter.

shape while uniformly cover the spectrum, we use loglets as the basis functions because they possess a number of useful properties [6].

Denoting the frequency vector by $\boldsymbol{u}$ and its length by ρ, a bandpass window with a loglets basis can be designed in the Fourier domain as,

$$\mathcal{W}(\boldsymbol{u}; s) = \mathrm{erf}\left(\alpha \log\left(\beta^{s+\frac{1}{2}} \frac{\rho}{\rho_0}\right)\right) - \mathrm{erf}\left(\alpha \log\left(\beta^{s-\frac{1}{2}} \frac{\rho}{\rho_0}\right)\right) \tag{1}$$

where α controls the radial bandwidth, s is an integer defining the scale of the filter, and $\beta > 1$ sets the relative ratio of adjacent scales – set to two for one octave intervals. ρ_0 is the peak radial frequency of the window with scale $s = 0$.

To extract the sharp textures of an image, the first scale channel should cover the highest frequency components. Noting the uniform property of the loglets,

we accumulate a group of loglets successively having one-octave higher central frequencies as the first scale window, i.e., $\mathcal{W}^{(1)} = \sum_s \mathcal{W}(\boldsymbol{u}; s), s = \{0, -1, ...\}$, which achieves an even coverage towards the highest frequency, see the 1D profile in Fig. 2(a) shown as a red curve, and the 2D window in Fig. 2(c). The second and larger scale features can be selected by windows covering lower frequencies, $\mathcal{W}^{(s)}(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}; s-1)$. Profiles of two adjacent larger scale windows are shown in Fig. 2(a) as blue curves, and a 2D window shown in Fig. 2(d). For a lossless decomposition, the largest scale window should uniformly cover the lowest frequencies, so it is designed as a accumulation of the remaining loglets functions, $\mathcal{W}^{(L)} = \sum_s \mathcal{W}(\boldsymbol{u}; s), s = \{L-1, L, ...\}$, see the green curve in Fig. 2(a). L is the total number of scales in the filter banks.

As the image filtering can be implemented in the Fourier domain by multiplication, the filters can be efficiently applied by windowing them on the image spectrum $\mathcal{I}$, and the image channels obtained by the inverse Fourier transform of the windowed spectrum, $I^{(s)} = \mathcal{F}^{-1}(\mathcal{I} \cdot \mathcal{W}^{(s)}), s = \{1, 2, ..., L\}$. The image is thus decomposed into complementary channels $\{I^{(s)}\}$.

Wavelet Image Pyramid. It is evident that larger scale textures can be described sufficiently at a lower resolution. Note in Fig. 2(a) that the magnitude of the two larger scale windows beyond $\pi/2$ and $\pi/4$ is almost zero. Therefore we can discard these areas of the spectrum, which results in an efficient downsampling without information loss or aliasing effect[2]. As a result, the resolution is reduced by 2^s at scale s and a subband pyramid is obtained, see Fig. 1(c, d).

Wavelet Appearance Pyramid (WAP). Given a landmark $\boldsymbol{x}$, we extract an image patch A_s at each scale s of the pyramid. All patches $\{A_s\}_{s=1}^{L}$ have the same size in pixels, which describe the local features at octave larger scales, domain sizes and lower resolutions, see Fig. 1(d, e). A WAP, denoted by $\Phi = [\mathcal{A}, \boldsymbol{s}]$, consists of an assembly of feature patches $\mathcal{A} = \{\{A_{s,i}\}_{s=1}^{L}\}_{i=1}^{N}$ extracted at all the landmarks $\{\boldsymbol{x}_i\}_1^N$, and a shape $\boldsymbol{s} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_N]$ designating the locations of the patches. At larger scales fewer patches are manually chosen at key landmarks to reduce the overlapping. Φ is then flattened into a 1D vector serving as the profile of the anatomy. A further feature extraction function such as histogram of oriented gradients (HOG) can be readily applied on the patches to reduce the dimensionality and enhance its robustness, i.e., $h(\mathcal{A}) = \{\{h(A_{s,i})\}_{s=1}^{L}\}_{i=1}^{N}$, see Fig. 1(f). To reconstruct the original appearance from the profile, we first pad the patches at each scale with the geometry configured by $\boldsymbol{s}$ to recover the individual channels. As the scales are complementary, all channels are then accumulated to recover the object appearance, see Fig. 1(g).

Landmark Detection. We integrate our WAP representation with the SDM algorithm [2] for robust landmark detection. To deduce the true landmark location $\boldsymbol{s}^*$ given an initial estimation $\hat{\boldsymbol{s}}$, we extract the descriptor $h(\mathcal{A}(\hat{\boldsymbol{s}}))$ at $\hat{\boldsymbol{s}}$ and learn the mapping $h(\mathcal{A}(\hat{\boldsymbol{s}})) \rightarrow \Delta\boldsymbol{s}^*$, in which $\Delta\boldsymbol{s}^* = \boldsymbol{s}^* - \hat{\boldsymbol{s}}$. The direct mapping function satisfying all the cases in the dataset is non-linear in nature and can

[2] Spectrum cropping as image downsampling is explained at https://goo.gl/ApJJeL.

be over-fitted. So we adopt the SDM and approximate the non-linear mapping with a sequence of linear mapping $\{R^{(i)}, \boldsymbol{b}^{(i)}\}$ and landmark updating steps,

$$\begin{cases} \text{Mapping:} & \Delta \boldsymbol{s}^{(i)} = R^{(i)} h(\mathcal{A}(\hat{\boldsymbol{s}}^{(i)})) + \boldsymbol{b}^{(i)}, \\ \text{Updating:} & \hat{\boldsymbol{s}}^{(i+1)} = \hat{\boldsymbol{s}}^{(i)} + \Delta \boldsymbol{s}^{(i)}. \end{cases} \tag{2}$$

The descriptor $h(\mathcal{A})$ is extracted and updated at each iteration. More details on the SDM can be found at [2].

Anatomical Classification. For the classification tasks, the correspondence of anatomical features should be built such that the differences among the descriptors account for the true variations rather than the miss-alignment. In a WAP the appearance correspondence is built by extracting local features at corresponding landmarks. A classifier predicts the label ℓ given an anatomical observation Φ, i.e., $\ell = \arg\max p(\ell|\Phi)$. The most significant variations in the training data $\{\Phi\}$ can be learned by principal components analysis and the dimensionality reduced by preserving the first t significant components, which span a feature space $P \in \mathbb{R}^{M \times t}$ with M being the dimensionality of Φ. A WAP therefore can be represented in the feature space by a compact set of parameters $\boldsymbol{b}_\Phi$, i.e., $\boldsymbol{b}_\Phi = P^T(\Phi - \bar{\Phi})$, in which $\bar{\Phi}$ is the mean of $\{\Phi\}$. Using $\boldsymbol{b}_\Phi$ as inputs the classifier now predicts $\ell = \arg\max p(\ell|\boldsymbol{b}_\Phi)$.

3 Results and Discussion

Clinical Background. Lumbar spinal stenosis (LSS) is a common disorder of the spine. The important function of radiological studies is to evaluate the morphological abnormalities and make the anatomical classification. Disc-level axial images in MRI scans can provide rich information for the diagnosis. In paired sagittal-axial scans, the disc-level planes (red line in Fig. 3(a)) are localised in sagittal scans, and the geometry is mapped to the registered axial scans (dashed lines in Fig. 3(a)) to extract the disc-level images. On the disc-level image shown in Fig. 3(b), conditions of the posterior margins of the disc (red line), posterior

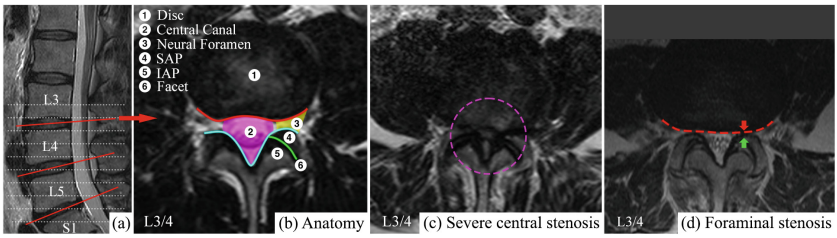

Fig. 3. (a) Mid-sagittal scan of a lumbar spine. (b) Anatomy of a L3/4 disc-level axial image. (c) Severe central stenosis. (d) Foraminal stenosis. The neural foramen are suppressed by the thickening of the facet (green) and the disc (red).

spinal canal (cyan line) and the facet between the superior and inferior articular processes (green line) are typically inspected for diagnosis and grading. Degeneration of these structures can constrict the spinal canal and the neural foramen causing central and foraminal stenosis. Pathological examples are given in Fig. 3(c, d). In clinical practice, parameters such as antero-posterior diameter, cross-sectional area of spinal canal are typically measured [7]. However there is a lack of consensus in these parameters and no diagnostic criteria are generally applicable [8]. A more detailed appearance model of the anatomy, followed by a classification therefore could contribute to reliable diagnoses.

Data and Settings. The dataset consists of T2-weighted MRI scans of 200 patients with varied LSS symptoms. Each patient has routine paired sagittal-axial scans. The L3/4, L4/5, L5/S1 disc-level axial planes are localised in the sagittal scans and the images sampled from the axial scans. We obtain 3 subset of 200 disc-level images from the three intervertebral planes, 600 images in total. Each image is inspected and graded, and the anatomy annotated with 37 landmarks outlining the disc, central canal and facet.

Results of Landmark Detection. To cover richer pathological variations, we perform the landmark detection on the mixed dataset containing all 600 images. We randomly choose 300 images for training and detect the landmarks on the remaining 300. Two metrics are used for the evaluation: the Point to Boundary Distance (PtoBD) and the Dice Similarity Coefficients (DSC) of the canal and disc contours. DSC is defined as the amount of the intersection between a fitted shape and the ground truth, $DSC = 2 \cdot tp/(2 \cdot tp + fp + fn)$, with tp, fp, fn denoting the true positive, false positive and false negative values respectively. We compare the proposed WAP with AAMs, ASMs and CLMs. To validate the improvement of the loglets pyramid decomposition, we also report the performance of an alternative model by replacing our pyramids with the original images but using the same HOG features and SDM algorithm. We refer to this control model as WAP$^-$. The mean results of landmark detection are shown in Table.1. We can see that the WAP outperforms the other methods by a favourable margin. Several qualitative results by WAP are shown in Fig. 4(Top). Generating a WAP is also efficient as the filtering is conducted directly in the Fourier domain. The most expensive computation is extracting the HOG descriptors, which takes only 54 ms on an image of size 496×496 pixels.

Table 1. Performance of landmark detection

Metrics	AAM	ASM	CLM	WAP$^-$	WAP
PtoBD (in pixels)	3.10 ± 1.29	2.51 ± 1.32	2.34 ± 1.15	1.95 ± 0.92	$\mathbf{1.87 \pm 0.73}$
DSC (%)	90.6 ± 4.9	92.1 ± 5.2	92.4 ± 5.2	93.9 ± 3.3	$\mathbf{94.7 \pm 2.6}$

Results of Anatomical Classification. For central stenosis, in each of the three subsets, the morphology of the central canal is inspected and labelled with

three grades: normal, moderate and severe. The average appearances of these classes delineated by WAPs are shown in Fig. 5(a). We randomly pick 100 samples to train the classifier, and test on the remaining 100, and repeat for 100 times for an unbiased result. The WAP extracted from the detected landmarks are projected onto the feature space and represented by a compact set of parameters $\boldsymbol{b}_\Phi$ (Fig. 4 Bottom), which are used as inputs of the classifier. The performance of normal/abnormal classification is measured with accuracy, which is calculated by $(tp + tn)/(tp + tn + fp + fn)$. The grading errors are measured with Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE). We compare the performance of our method against approaches using other models as inputs to the same classifier. The agreements of the results with manual inspection are reported in Table 2. Similarly we perform another normal/abnormal classification on the morphology of the neural foremen. The average appearances are given in Fig. 5(b). The classification accuracy of methods compared is reported in Table 3. We can see that in both tasks, our WAP appearance models enable a significant improvement.

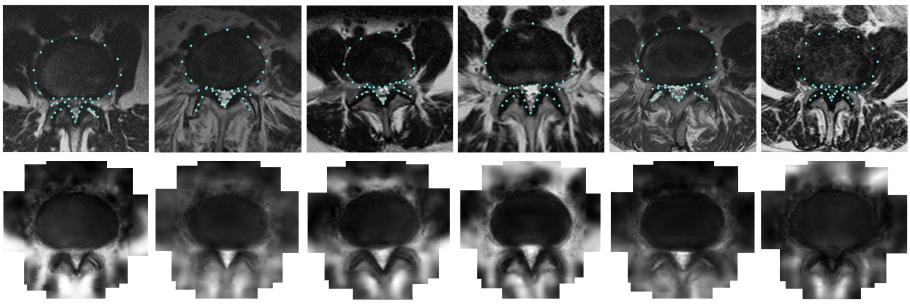

Fig. 4. Top: Qualitative results of landmark detection. Bottom: Appearance fitted by WAP. The appearances shown are represented by $\boldsymbol{b}_\Phi$ in the feature space which are used as inputs for classification.

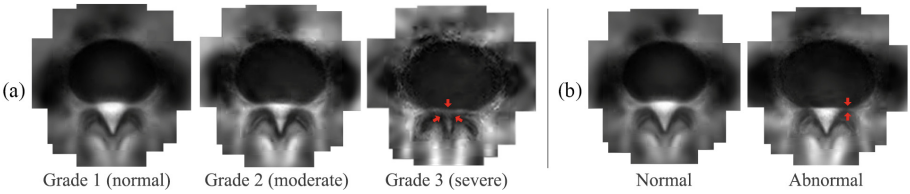

Fig. 5. Average appearance of classes represented by WAP. (a) Three grades of central stenosis. (b) Normal and abnormal in terms of foreminal stenosis.

Table 2. Agreement of classification and grading of central stenosis

Method	Accuracy (%) of classification			MAE of grading			RMSE of grading		
	L3/4	L4/5	L5/S1	L3/4	L4/5	L5/S1	L3/4	L4/5	L5/S1
ASM	79.1 ± 4.8	77.4 ± 4.3	81.7 ± 4.5	0.25	0.31	0.20	0.55	0.67	0.48
AAM	70.1 ± 7.1	69.7 ± 7.3	71.3 ± 8.8	0.41	0.44	0.32	0.72	0.79	0.58
CLM	81.0 ± 4.9	82.4 ± 4.5	82.7 ± 4.4	0.23	0.25	0.23	0.53	0.56	0.52
WAP$^-$	80.7 ± 4.9	82.1 ± 4.6	84.7 ± 4.2	0.23	0.25	0.18	0.53	0.58	0.47
WAP	$\mathbf{84.7 \pm 4.6}$	$\mathbf{84.5 \pm 4.3}$	$\mathbf{85.9 \pm 4.2}$	**0.19**	**0.21**	**0.16**	**0.48**	**0.54**	**0.44**

Table 3. Accuracy (%) of classification of foreminal stenosis

Anatomy	ASM	AAM	CLM	WAP$^-$	WAP
L3/4	83.3 ± 3.8	73.3 ± 5.5	83.1 ± 4.7	84.3 ± 4.1	$\mathbf{85.0 \pm 3.9}$
L4/5	82.4 ± 4.6	76.2 ± 5.8	83.3 ± 4.3	86.9 ± 3.9	$\mathbf{87.8 \pm 3.5}$
L5/S1	81.8 ± 4.7	74.5 ± 5.7	82.9 ± 4.5	85.2 ± 4.3	$\mathbf{85.7 \pm 4.3}$

4 Conclusions

We have presented a novel appearance model and demonstrated its applications to the problem of LSS for variability modelling, landmark detection and pathology classification. The improvement in the diagnosis and grading lies in its ability to capture detailed appearances, better appearance correspondence by scale decomposition, and more precise landmark detection. The model can be readily applied to other anatomical areas for clinical tasks requiring segmentation and classification. The source code will be released for research purposes. For the task of LSS, our future work is aimed towards patient-level diagnosis by utilising the image-level anatomical classification together with etiological information.

References

1. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)
2. Xiong, X., De la Torre, F.: Supervised descent method and its applications to face alignment. In: 2013 IEEE Conference on CVPR, pp. 532–539. IEEE (2013)
3. Lindner, C., Thiagarajah, S., Wilkinson, J., Consortium, T., Wallis, G., Cootes, T.F.: Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE Trans. Med. Imaging **32**(8), 1462–1472 (2013)
4. Antonakos, E., Alabort-i Medina, J., Zafeiriou, S.: Active pictorial structures. In: Proceedings of the IEEE Conference on CVPR, pp. 5435–5444 (2015)
5. Saragih, J.M., Lucey, S., Cohn, J.F.: Face alignment through subspace constrained mean-shifts. In: IEEE 12th International Conference on Computer Vision, pp. 1034–1041. IEEE (2009)

6. Knutsson, H., Andersson, M.: Loglets: generalized quadrature and phase for local spatio-temporal structure estimation. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 741–748. Springer, Heidelberg (2003). doi:10.1007/3-540-45103-X_98
7. Steurer, J., Roner, S., Gnannt, R., Hodler, J.: Quantitative radiologic criteria for the diagnosis of lumbar spinal stenosis: a systematic literature review. BMC Musculoskelet. Disorders **12**(1), 175 (2011)
8. Ericksen, S.: Lumbar spinal stenosis: imaging and non-operative management. Semin. Spine Surg. **25**, 234–245 (2013). Elsevier
9. Cristinacce, D., Cootes, T.: Automatic feature localisation with constrained local models. Pattern Recogn. **41**(10), 3054–3067 (2008)