

Automatic Quantification of Hip Osteoarthritis from Low-Quality X-ray Images

Giancarlos Gutiérrez C.^a, Eveling Castro G.^a, Javier Delgado O.^b, Paúl Pauca^c, Stefan Klein^d, and Hans Lamecker^e

^aUniversidad Nacional de San Agustín de Arequipa, Arequipa, Perú

^bUniversidad Austral de Chile, Valdivia, Chile

^cWake Forest University, North Carolina, USA

^dBiomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, the Netherlands

^e1000shapes GmbH, Berlin, Germany

ABSTRACT

Diagnosis of hip osteoarthritis is conventionally done through a manual measurement of the joint distance between the femoral head and the acetabular cup, a difficult and often error-prone process. Recently, Chen et al.¹ proposed a fully automated technique based on landmark displacement estimation from multiple image patches that is able to accurately segment bone structures around the pelvis. This technique was shown to be comparable or better than state-of-the-art random forest based methods. In this paper, we report on the implementation and evaluation of this method on low-resolution datasets typically available in parts of the developing world where high-resolution X-ray image technology is unavailable.

We employed a dataset of hip joint images collected at a local clinic and provided to us in JPG format and at 1/3 the resolution of typical DICOM X-ray images. In addition, we employed the Dice similarity coefficient, average Euclidean distance between corresponding landmarks, and Hausdorff distance to better evaluate the method relative to diagnosis of hip osteoarthritis. Our results show that the proposed method is robust with JPEG images at 1/3 the resolution of DICOM data. Additional preliminary results quantify the accuracy of the approach as a function of decreasing resolution. We believe these results have important significance for application in clinical settings where modern X-ray equipment is not available.

Keywords: Landmark Detection, Bone Structures Segmentation, Statistical Shape Models, Active Shape Models, X-ray Image, Supervised Learning, Hip Osteoarthritis

1. INTRODUCTION

Hip osteoarthritis is a common degenerative condition that occurs in people who are fifty years of age and older. In hip osteoarthritis, the cartilage in the hip joint wears away over time, reducing the protective joint space between the bones. X-ray imaging is the prevalent test for diagnosis of hip osteoarthritis in most parts of the developing world. The presence of osteophytes (bone spurs), cysts, subchondral sclerosis (hardening of the bone under the cartilage), and narrowing of the space between the femoral head and the acetabular cup are indicative of hip osteoarthritis. A trained radiologist typically carries out such evaluation from X-ray images, segmenting bone structures and measuring the joint space, when possible. This is often a challenging, subjective, and error-prone process.

Image segmentation is often one of the most critical tasks for image analysis using computer-aided systems and segmentation methods based on detection of reference points (or landmarks) have received considerable amount of interest in the literature. Information extracted during segmentation can lead to accurate object representation and measurement.² Recently, Chen et al.¹ proposed a new approach for landmark detection that jointly estimates displacement of multiple randomly sampled image patches to landmark positions and uses a

Further author information: (Send correspondence to Giancarlos G. C.)

Giancarlos G. C.: E-mail: ggutierrezce@unsa.edu.pe, Telephone: +51 959 037 270

shape composition model to generate segmented shape contours. This is a powerful approach that relies on local image features and the efficient solution of a convex optimization problem. Krishnakumari³ used the method proposed by Chen et al. to perform segmentation of the femur and pelvis and to subsequently measure the hip joint space. This method was tested on high-resolution X-ray data with pixel size of around 0.143 to 0.168 mm.

The main purpose of our work is to evaluate the feasibility of measuring the hip joint space using the landmark detection method of Chen et al. and the distance estimation of Krishnakumari on mid to low-resolution X-ray images. We use a small collection of X-ray images of the hip joint provided to us by a local clinic*. The images are stored in JPG format, at 1/3 the resolution of typical DICOM imagery. The ability to obtain accurate measurement of joint spaces from such data would be relevant to many clinical settings around the developing world.

The remainder of this paper is organized as follows. Relevant and related literature is briefly summarized in Section 2. In Section 3, we review the data-driven landmark displacement estimation and shape regularization approach proposed by Chen et al.¹ We also provide details regarding the use of gradient profiling proposed by Krishnakumari³ and the measurement of the hip joint space. We offer experimentation results and discussion in Section 4. Conclusions, and final remarks are found in Section 5.

2. RELATED WORK

The detection of landmarks has a long history in medical imaging and other applications and a number of methods have been proposed that exploit high frequency information, such as edges and contours along with topological constraints. More recently, the use of shapes and morphological parameters with machine learning based techniques have been shown to be robust to pose and variation changes in the segmentation of bony structures (femur, pelvis, vertebrae, skull, etc.). An important application for this type of work is in assisting or automating the diagnostic process of certain diseases, such as arthrosis of the hip or knee.^{4,5}

Landmark detection based on random forests has emerged as a leading approach in medical imaging segmentation tasks.⁶ In this approach, the displacement of local image patches relative to the landmarks is estimated via random forest regression, which presents favorable results compared to traditional methods of segmentation.²

Chen et al.¹ proposed a data-driven approach for the estimation of local image patch displacement, instead of random forests, and improve the displacement prediction accuracy. Their approach combines the comparison of a test patch with training patches, with a geometric constraint on displacements relative to a common landmark. The shape given by the positions of these landmarks is then regularized via a statistical shape model (SSM).⁷ In particular, they employ the sparse shape composition of Zhang et al.⁸ for shape regularization. Sarkalkan et al.⁹ affirm that the potential of SSM is well recognized in the research community for the study of osteoarthritis, osteoporosis, implant design, surgery planning and 2D/3D reconstruction.^{10,11}

Several computer-based methods have been proposed in the literature for the measurement of joint spaces.¹² Krishnakumari³ calculates the distances of the landmarks located along the femoral head and the acetabular cup after applying landmark based segmentation. These distances provide data that can be used to diagnose and assess osteoarthritis. Boniatis et al.¹³ performed osteoarthritis classification based on the measurement scale of Kellgren and Lawrence.¹⁴

3. METHODOLOGY

The approach proposed by Chen et al. for segmentation employs a two-step framework consisting of the detection of landmarks followed by shape regularization. Krishnakumari extends this approach by including a gradient profiling step with the objective of further improving detection accuracy. Measurement of the joint space is then calculated based on landmarks around the femoral head and the acetabular cup.

*<http://orthodoc.aaos.org/javierdelgado/>

3.1 Landmark Detection

The general idea of landmark detection is to detect or approximate the position of each landmark in the shape corresponding to a bone structure (femur or pelvis). A shape is defined as an ordered set of landmarks along a bone boundary. Given a set of manually labeled ground-truth landmarks, during training, a rectangular image patch is sampled randomly around each landmark. Figure 1 shows sample images from our dataset with manually obtained ground-truth landmarks.

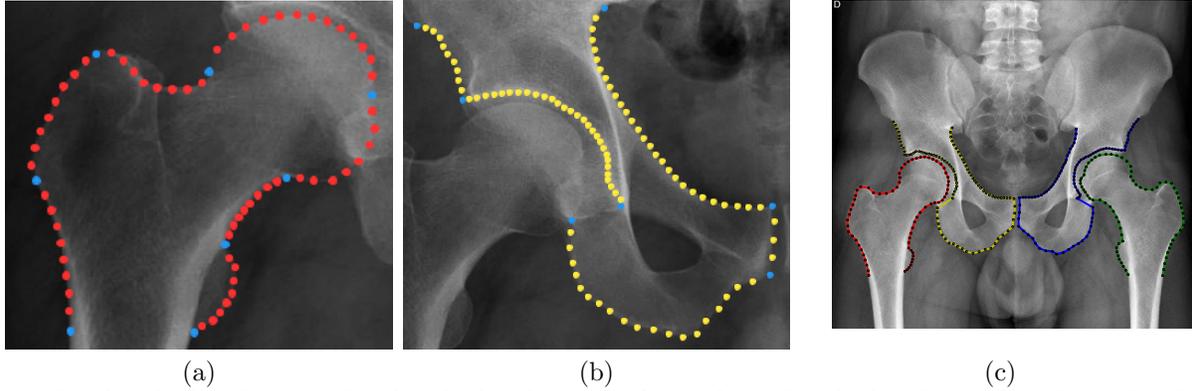


Figure 1. Landmarks distribution. **a** Landmarks distribution for femur. **b** Landmarks distribution for pelvis. **c** Landmarks distribution of the femur and pelvis for both sides of the hip.

Figure 2(b) shows the patch random sampling around a specific landmark in a sample image. Notice that in practice, a shape (e.g., Figure 2(a)) is subdivided into subshapes, that is, into subsets of landmarks of a specific length. See Section 4 for more details. For each patch, a visual feature, $\tilde{\mathbf{f}}$, and the patch displacement from

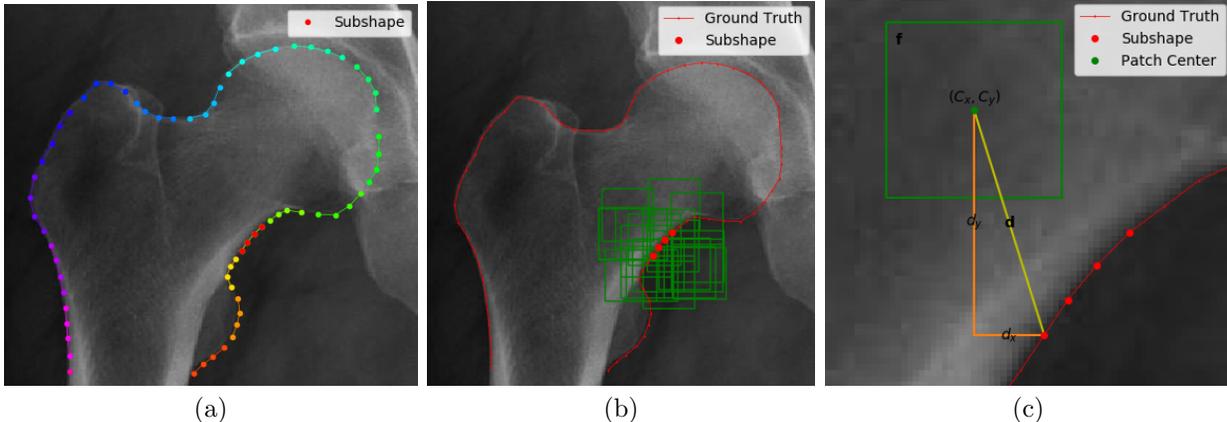


Figure 2. Training pipeline for a subshape. **a** Subshape with L landmarks. **b** Sampling patches around subshape. **c** Training data for each patch for a given landmark in the subshape. Feature vector \mathbf{f} , Patch centre (C_x, C_y) and Displacement Vector (d_x, d_y) .

the ground-truth landmark, $\tilde{\mathbf{d}}$, are obtained. The training set is then given by $\{(\tilde{\mathbf{f}}_k, \tilde{\mathbf{d}}_k)\}$. The area randomly sampled around each landmark is controlled by two parameters: *patch size* and *sampling radius*. The patch size defines the patch dimension and the sampling radius defines how far the patches can be sampled away from a landmark. The visual features $\tilde{\mathbf{f}}$ are obtained using multi-level histogram of oriented gradients (HoG).¹⁵

During testing, a new image with unknown landmark locations is given. Rectangular patches are sampled randomly everywhere in the image (see Figure 3(a)) and HoG image features are calculated for each test patch. These HoG features, $\{\mathbf{f}_k\}$, as well as the test patch centers, $\{\mathbf{c}_k\}$, are then used along with the training data to calculate the displacement $\mathbf{d}_{k,\ell}$ of the k th test patch relative to the ℓ th (unknown) landmark.

The calculation of the displacements $\mathbf{D} = [\mathbf{d}_{k,\ell}]$ is done for all test patches via minimization of a cleverly designed objective function:

$$E(\hat{\mathbf{D}}) = E_g(\hat{\mathbf{D}}) + \alpha E_f(\hat{\mathbf{D}}) + \beta E_p(\hat{\mathbf{D}}), \quad (1)$$

where $\hat{\mathbf{D}} = [\tilde{\mathbf{D}} \mathbf{D}]$ is the matrix composed of \tilde{K} training displacements, $\tilde{\mathbf{D}}$, and K test displacements, \mathbf{D} , to be estimated. Here, $E_g(\hat{\mathbf{D}})$ measures the discrepancy between computed training displacements and the ground truth,

$$E_g(\hat{\mathbf{D}}) = \frac{1}{2LK} \left\| \hat{\mathbf{D}}\mathbf{P} - \tilde{\mathbf{D}}_{CT} \right\|_F^2, \quad (2)$$

where $\tilde{\mathbf{D}} = \hat{\mathbf{D}}\mathbf{P}$. The term $E_f(\hat{\mathbf{D}})$ is given by,

$$E_f(\hat{\mathbf{D}}) = \frac{1}{2L \sum_{i \neq j} s_{ij}} \sum_{i \neq j} s_{ij} \left\| \text{col}_i(\hat{\mathbf{D}}) - \text{col}_j(\hat{\mathbf{D}}) \right\|_{\ell_2}^2. \quad (3)$$

This term penalizes patches whose HoG features are similar in the feature space but whose corresponding displacements are large. Similarity in the feature space is determined from all pairwise distances between features. That is, $s_{ij} = 1$ if and only if the i th and the j th patches are mutually ρ nearest neighbors in the feature space,¹ and $s_{ij} = 0$ otherwise.

Finally, $E_p(\hat{\mathbf{D}})$ is given by,

$$E_p(\hat{\mathbf{D}}) = \frac{1}{2LK} \left\| \hat{\mathbf{D}}\mathbf{Q}\mathbf{U} - \bar{\mathbf{C}} \right\|_F^2, \quad (4)$$

where $\hat{\mathbf{D}}\mathbf{Q}\mathbf{U}$ calculates the difference vectors between displacements of different patches and $\bar{\mathbf{C}}$ calculates the difference vectors between different patch centers. $E_p(\hat{\mathbf{D}})$ exploits the geometric triangular constraint imposed by the displacements of two different patches relative to the same landmark. Inclusion of this term in the minimization of (1), which is shown to be convex, is central to the method proposed by Chen et al.

For each ℓ th landmark, $\{\mathbf{v}_k = \mathbf{c}_k + \mathbf{d}_{k,\ell}\}$ for all k patches, constitutes the set of votes towards the position of the ℓ th landmark. Figure 3(b) shows the positions voted by all test patches for the landmarks in the left femur. The final landmark positions, giving the shape of a bone structure, can be calculated using some voting scheme. Cheng et al. calculate a response function $I(x, y)$ over the test image giving the probability of a landmark being found at position (x, y) , using a Gaussian distribution with mean \mathbf{v}_k and variance σ_k obtained from the training displacements associated with the k th patch. Figure 3(c) shows the response function $I(x, y)$ with high density positions highlighted.

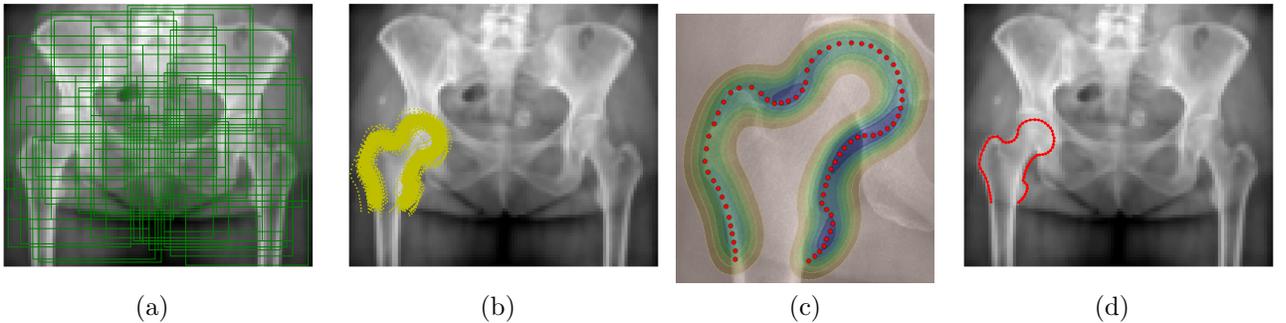


Figure 3. Testing Pipeline. **a** Randomized patch sampling around the entire image. **b** Landmark positions contributed by each patch. **c** Image of the response functions for landmarks in the femur, with high density positions highlighted for each landmark. **d** Result of landmark detection at the first scale.

In practice a multi-scale strategy is used to improve computational efficiency, where the output of one scale is the input of the next. The first scale is the *initialization* scale (see Fig. 4 (a)) and the aim is to find the global position of the bone structure in the X-ray image. The concept of subshapes is not used in this scale, instead, the complete shape of the bone structure is treated as a single subshape and the patches are randomly

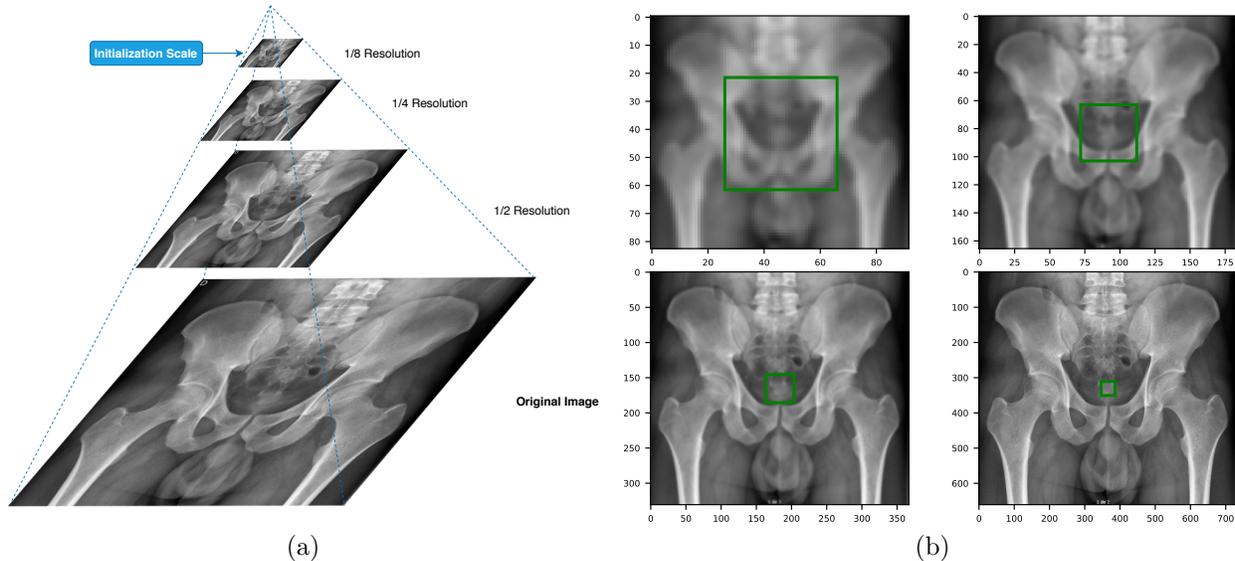


Figure 4. Multi-scale approach. **a** Gaussian pyramid. **b** Variation of the resolution of the X-ray image with a patch of constant size.

sampled throughout the X-ray image, as shown in Figure 3(a). In each subsequent scale, the landmark detection approach is performed over each subshape keeping the patch size constant, as illustrated in Figure 4(b).

In our implementation, the patch size (40×40 px) and the sampling radius (30 px) remain constant for all scales. The only parameter that changes is the resolution of the X-ray image (see Figure 4(b)). Thus, the information contained in the sampled patches and the computational efficiency depends on the scale.

3.2 Shape Regularization

Shape regularization is employed to further optimize the shape contour captured in the response image of each landmark by considering the shape prior information encoded in the training dataset. Chen et al. employ a shape model where a shape to be regularized is approximated by a linear combination of a small subset of training shapes. This is often posed as an L1-regularized least squares problem that has advantages over L2-regularization when the desired solutions are also sparse. In this study, we apply the traditional active shape model based on principal component analysis (PCA)¹⁶ to generate a statistical shape model (SSM).

Figure 5 shows the alignment of the shapes described by the set of landmarks and the mean shape of the SSM of both the femur and the pelvis, which make up the ASM training data.

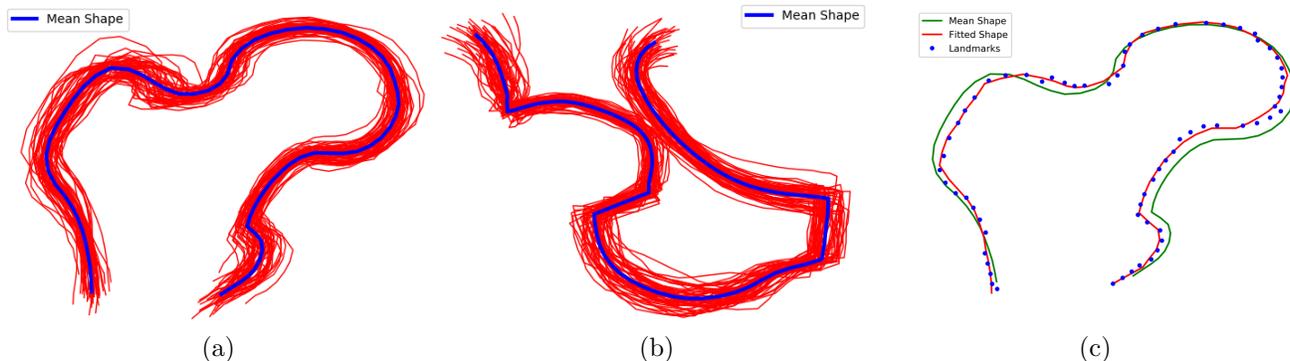


Figure 5. Creation of statistical shape model stage training data. **a** SSM for the femur. **b** SSM for the pelvis. **c** Sample result of shape regularization via SSM.

3.3 Gradient Profiling

The use of gradient profiling is based on the assumption that, for each shape contour, each landmark is on a distinct edge from other landmarks in that shape.³ Edges are typically found using gradient approximations and it is well-known that Gaussian derivatives are less sensitive to image noise. Krishnakumari³ suggests taking rectangular patches of the gradient image, centered around each landmark, as training data. The arrangement of the patches is appreciated in Figure 6 for the femur and pelvis. During testing, the process of gradient profiling is

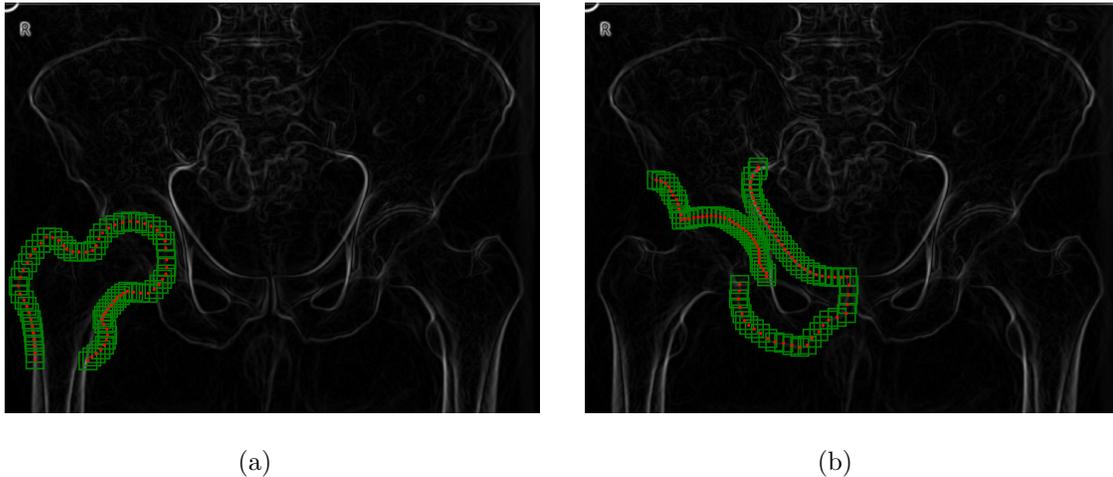


Figure 6. Creation of the training data of the gradient profiling stage. **a** Patch sampling for the femur. **b** Patch sampling for the pelvis.

applied to the test image and the resulting profile is compared with the training profiles via template matching, based on normalized cross-correlation. The center of a sub-patch in the test patch that provides the highest value is considered to be new updated landmark position voted by the training patch. This process is repeated at each scale. The parameters used for gradient profiling are given in Table 1. Note that the initial scale it is not considered.

Table 1. Gradient Profiling Parameters

Parameter	Scale 0.25	Scale 0.5	Scale 1
Training Patch size	10	20	30
Testing Patch size	20	30	40
σ	0.5	1	2

3.4 Joint Space Measurement

After segmentation and refinement of the femur and pelvis at all scales, the final step is the measurement of the distance between the femur and pelvis. This distance is also referred to as the joint space width in the literature and a number of metrics can be considered.¹² For the present work, the joint space width is calculated considering landmarks 28 through 40 in the femur and landmarks 10 through 32 in the pelvis. In particular, for each femur landmark, the closest point is found in the contour of the pelvis and the distance to that point is considered as the joint distance for that landmark. Depending on the number of landmarks considered, distances are calculated as shown in Figure 7(b). The set of distances from each femur landmarks towards the pelvis contour, the average distance, the mean, the minimum, and the maximum of these distances have been considered as metrics for the measurement of the joint space.

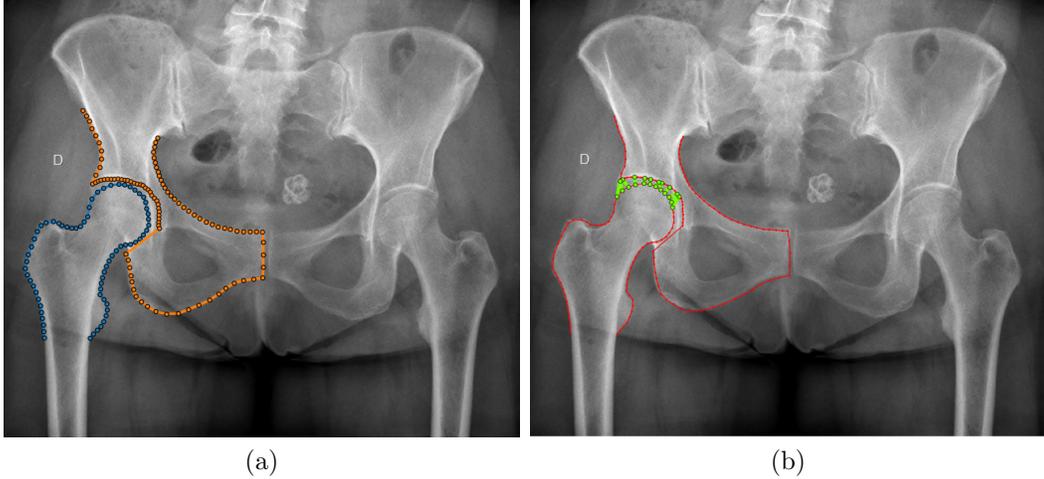


Figure 7. Joint space measurement. **a** Result of the final segmentation of the femur and pelvis. **b** Distances of joint space.

4. EXPERIMENTATION AND RESULTS

4.1 Data

The dataset consists of 60 JPG images of hip X-rays in anterior-posterior view of an average size of 800×700 pixels. A total of 50 were used for training and 10 during testing. To generate the ground-truth data, a clinician manually identified the positions of the *dominant landmarks* for each bone structure (those that are in positions of interest). Once the positions of the dominant landmarks have been identified, a certain number of points were distributed in a uniform manner to generate the entire set of landmarks for each bone structure. For the present work we identified 8 dominant landmarks and a total of 69 landmarks for the femur (see Figure 1(a)). In the same way for the pelvis, 7 dominant landmarks were identified with a total of 95 landmarks (see Figure 1 (b)). The dominant landmarks are included within the total number of landmarks. The set of manually segmented images is referred to as the *gold standard* in the remainder of this paper. The gold standard is useful for assessing accuracy of the segmentation method in the testing phase.

4.2 Parameters

Our implementation of the multi-scale strategy employed four scale levels $[1/8, 1/4, 1/2, 1]$ and the original images were resized accordingly. Following the work of Chen et al., all the shapes were divided into subshapes of $L = 4$ consecutive landmarks, resulting in 17 subshapes for the femur and 24 subshapes for the pelvis. Similarly, for the multi-level HoG descriptor we used cell sizes of $(2, 2)$ and $(4, 4)$ and 18 orientations for each HoG feature. The patch size and sampling radius were fixed for all scale levels at 40×40 pixels and 30 pixels, respectively. We sampled $\tilde{K} = K = 200$ training and test patches for each subshape. Finally, after some experimentation, we selected $\alpha = 0.05$ and $\beta = 0.005$ for the solution of the objective function (1).

4.3 Segmentation Results

We first evaluate the performance of our segmentation approach, consisting of landmark detection followed by shape regularization via SSM and PCA, on our entire dataset. Qualitative results for segmentation of the right femur and right pelvis at each of our 4 scale levels are shown in Figure 8. Notice the progression in accuracy of the segmented shapes from the initial scale at $1/8$ resolution (Figure 8(a)) to the last scale (Figure 8(d)). In the same way, segmentation results corresponding to the left femur and left pelvis are shown in Figure 9. These results are consistent with those found in the literature.

When gradient profiling was included as the last step in the process, the segmentation results were not satisfactory. These results are not included in this paper. The lower quality of our data (JPG) might be a possible reason for this issue. Khrisnakumari employs high resolution data stored in DICOM format. Another possible cause might be the relatively smaller amount of X-ray images available to us for training during gradient profiling.

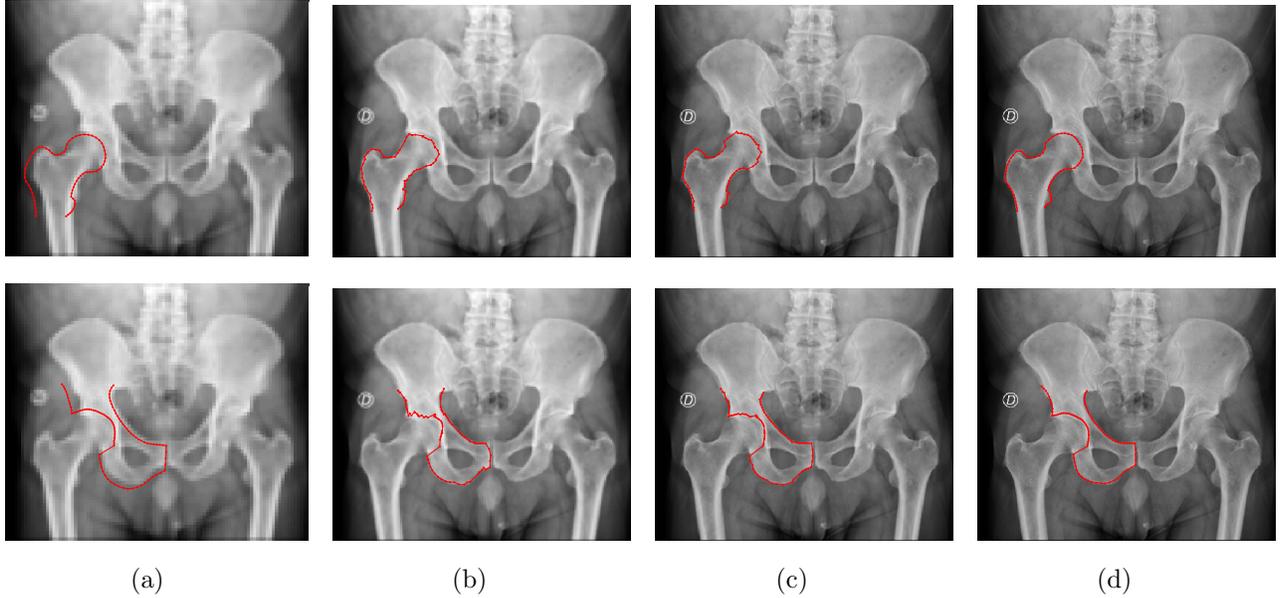


Figure 8. Results of the segmentation of the femur and pelvis on the right side for all scales. **a** Scale 1/8 **b** Scale 1/4 **c** Scale 1/2 **d** Scale 1

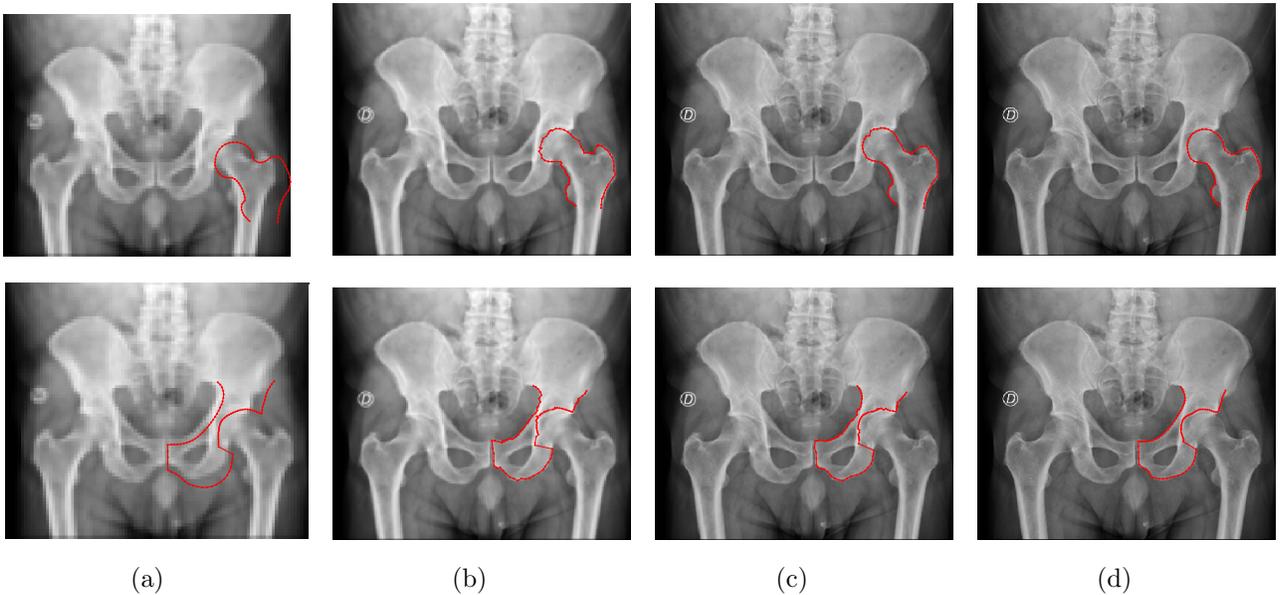


Figure 9. Results of the segmentation of the femur and pelvis on the left side for all scales. **a** Scale 1/8 **b** Scale 1/4 **c** Scale 1/2 **d** Scale 1

JPG is a method for lossy compression of digital imagery. As the amount of JPG compression in our dataset is unknown, it is hard to estimate the actual pixel resolution in our images. Though ideally, images should be exported in DICOM format, their size can be large and inconvenient for data transfer in many places around the world. The question of whether segmentation is feasible in low resolution imagery can be partially answered from observation of the segmentation results in Figures 8 and 9 in columns (a) - (c). To further explore this question, we applied our segmentation approach to the set of images at the first scale level only (1/8 the resolution) and mapped the results to the original images. A comparison of these results relative to the ground-truth data (gold standard) can be seen in Figure 10(a) for the femur and pelvis. Results obtained using the original images and the multi-scale approach with $L = 4$ appear in Figure 10(a). There is clearly a significant discrepancy between the

estimated and the ground-truth shapes in (a), which are compounded by an apparent shift along the horizontal direction. However, the shape is fairly well preserved and the result is surprising given that the resolution was decreased by a factor of 8.

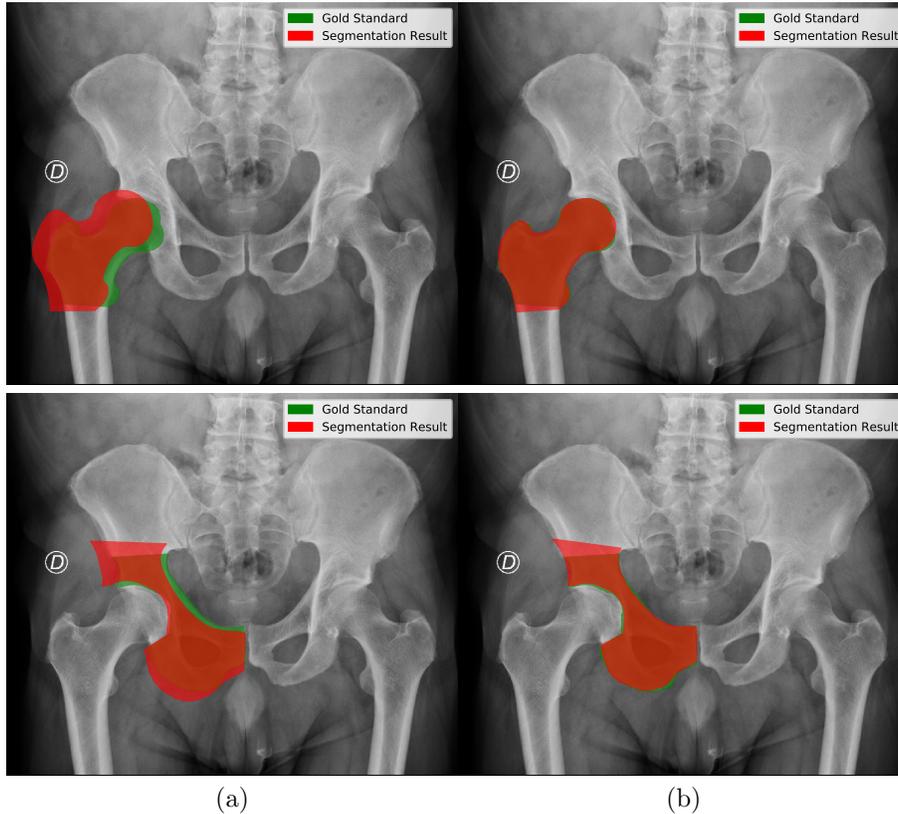


Figure 10. Results of the segmentation of the femur and pelvis with and without a multi-scale approach. **a** Segmentation result at the initial scale (1/8 resolution). **b** Segmentation result with $L = 4$ scales.

To quantitatively evaluate the overall performance of our segmentation approach (without grading profiling), we employed three well-known metrics in medical imaging. Specifically, for all test images and the ground-truth data, we calculate the Dice similarity coefficient (DSC), the average Euclidean distance (AED) between corresponding landmarks, and the bidirectional Hausdorff distance (HD).¹⁷⁻¹⁹ The average values of these metrics over all test images are provided in Table 2. DSC values of 0.94 and 0.96 reflect high precision in the segmentation of the femur and pelvis, respectively, which occur when the multi-scale approach is used. In comparison, the DSC values decrease significantly when segmentation is applied to the images at 1/8 resolution. As discussed previously, the apparent horizontal shift might negatively impact the segmentation quality. In terms of AED

Table 2. Average of the Dice similarity coefficient, the average Euclidean distance, and the Hausdorff distance.

Bone Structure	DSC	AED (px)	HD (px)
Femur (at 1/8 resolution)	0.74	24	29
Pelvis (1/8 resolution)	0.81	17.4	34
Femur (with $L = 4$ levels)	0.96	6.8	14.3
Pelvis (with $L = 4$ levels)	0.94	7.1	18.2

and HD, we also observe an increase in the distance values when segmentation is applied only at the lowest

resolution. Interestingly, the increase factor is higher for the pelvis than for the femur. The distance is provided in pixels as the resolution of the original images is unknown.

4.4 Joint Space Measurement Results

From the final segmentation of the femur and the pelvis, the measurement of the joint space is made by obtaining the distances from 12 landmarks in the femur to the acetabular cup. This step was applied, both to the gold standard and to the result obtained during segmentation of the X-ray images. Figure 11 shows the distances obtained in one image of our dataset. The absolute error between corresponding landmarks of the gold standard and the segmentation result distances is calculated for each image. The distributions of these errors are shown as a box and whisker plot in Figure 12(a). Notice that the distribution of error for the first landmark (to the inside of the joint) is significantly larger than the others.

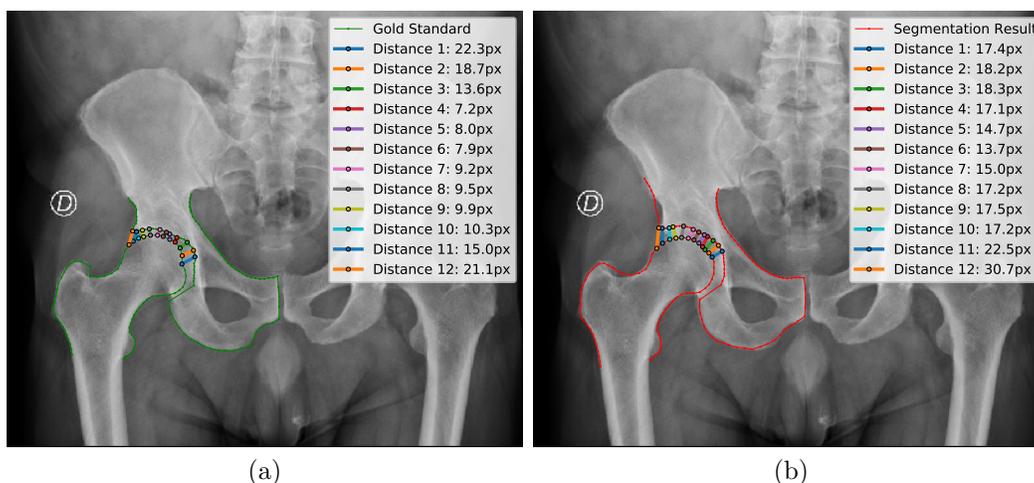


Figure 11. The results of the hip joint (right side). **a** Segmentation’s Gold Standard. **b** Result of the automatic segmentation.

To further evaluate the overall performance of the joint space measurement step, for each image, we average the distances of the 12 landmarks for both the gold standard and the segmentation result and then subtract these values to obtain one error value for each image. The distributions of the mean, median, minimum, and maximum errors found are reported in Figure 12(b).

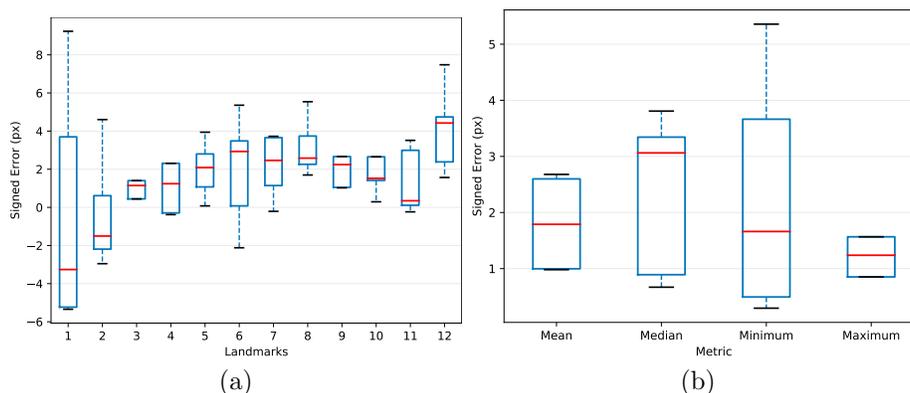


Figure 12. Joint Space Width Error Rates. **a** Landmark Distances as Metric. **b** Other Metrics.

In addition, the average error rates are extracted and reported in Table 3.

Table 3. Average error rates of the joint space width measurement.

Error	Landmark Distances (px)	Mean (px)	Median (px)	Minimum (px)	Maximum (px)
Error Rate	2.4	1.8	3.1	1.7	1.6

5. CONCLUSIONS AND FINAL REMARKS

In this paper, we implemented and evaluated the data-driven, multi-scale, landmark displacement estimation approach proposed by Chen et al.¹ on a small dataset of JPG X-ray images. Our shape regularization step employs a statistical shape model based on PCA, rather than on a sparse representation approach. We also explored the use of gradient profiling, proposed by Khrisnakumari,³ to further improve segmentation accuracy.

Our experiments show the viability of using landmark-based segmentation on JPG compressed images. The multi-scale approach contributed to better segmentation results, as evaluated by the Dice similarity coefficient, whose values exceeded 94% in the femur and pelvis bone structures. We further explored the question of performance on low resolution data by applying our segmentation approach to images at 1/8 the original resolution. The quality of segmentation was lower, as expected, but the Dice similarity coefficients still exceeded 74%.

The purpose of performing the segmentation of the bone structures was to obtain the measurement of the hip joint space. The results show an average signed error rate of 2.4 px using the landmarks distances as a metric. The average signed error using the other metrics are 1.8 px for the mean measurement, 3.1 px for the median, 1.7 px for the minimum, and 1.6 px for the maximum. These values are deemed to be acceptable for assistance in the diagnosis of hip osteoarthritis.

The measurement of the hip joint space depends heavily on the quality of the segmentation performed on the bone structures of the hip. If the segmentation is poor, particularly around the femur and the acetabular cup, then the joint space measurement is also poor. In future work, we would like to explore the integration of the joint space distance as part of the optimization step. This could be done, for example, by modeling the shape of the joint space itself. We are also interested in testing our algorithms on larger and more diverse datasets.

6. ACKNOWLEDGMENTS

This research was supported in part by Universidad Nacional de San Agustín de Arequipa (UNSA), with Contract: IBA-0012-2016, as part of the project “Construction of a 3D model to assist in hip joint replacement surgery from x-ray images, for the treatment of osteoarthritis and osteoporosis, at the UNSA Hospital of Arequipa”. The authors thank the CiTeSoft team at UNSA for their effort under Contract: EC-0003-2017-UNSA and for the equipment and the resources allocated to this research project.

REFERENCES

- [1] Chen, C., Xie, W., Franke, J., Grutzner, P. A., Nolte, L. P., and Zheng, G., “Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements,” *Medical Image Analysis* **18**(3), 487–499 (2014).
- [2] Patil, D. D., Deore, S. G., and Bhusawal, S., “Medical Image Segmentation: A Review,” *Ijcsmc* **2**(1), 22–27 (2013).
- [3] Krishnakumari, P., *Supervised Learning for Measuring Hip Joint Distance in Digital X-Ray Images*, Master’s thesis, Delft University of Technology (2015).
- [4] S., S., U., P., and R., R., “Detection of Osteoarthritis using Knee X-Ray Image Analyses: A Machine Vision based Approach,” *International Journal of Computer Applications* **145**(1), 20–26 (2016).
- [5] Kawathekar, P. P. and Karande, K. J., “Severity analysis of Osteoarthritis of knee joint from X-ray images: A Literature review,” in [*2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014)*], 648–652, IEEE (jul 2014).

- [6] Lindner, C., Thiagarajah, S., Wilkinson, J. M., Consortium, T., Wallis, G. A., and Cootes, T. F., “Fully Automatic Segmentation of the Proximal Femur Using Random Forest Regression Voting,” *Medical Image Analysis* **32**(8), 1462–1472 (2013).
- [7] Cootes, T., Baldock, E., and Graham, J., “An introduction to active shape models,” *Image Processing and Analysis* , 223–248 (2000).
- [8] Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D. N., and Zhou, X. S., “Sparse shape composition: A new framework for shape prior modeling,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , 1025–1032 (2011).
- [9] Sarkalkan, N., Weinans, H., and Zadpoor, A. A., “Statistical shape and appearance models of bones,” *Bone* **60**, 129–140 (2014).
- [10] Baka, N., Kaptein, B. L., de Bruijne, M., van Walsum, T., Giphart, J. E., Niessen, W. J., and Lelieveldt, B. P. F., “2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models,” *Medical Image Analysis* **15**(6), 840–850 (2011).
- [11] Xie, W., Chen, C., Grützner, P. A., Schumann, S., Nolte, L. P., and Zheng, G., “A complete-pelvis segmentation framework for image-free total hip arthroplasty (THA): methodology and clinical study,” *The international journal of medical robotics + computer assisted surgery : MRCAS* **7**(April), 375–392 (2014).
- [12] Sharp, J. T., Angwin, J., Boers, M., Duryea, J., Von Ingersleben, G., Hall, J. R., Kauffman, J. A., Landewé, R., Langs, G., Lukas, C., Maillefert, J. F., Bernelot Moens, H. J., Peloschek, P., Strand, V., and Van Der Heijde, D., “Computer based methods for measurement of joint space width: Update of an ongoing OMERACT project,” *Journal of Rheumatology* **34**(4), 874–883 (2007).
- [13] Boniatis, I., Panagiotopoulos, E., Lymberopoulos, D., and Panayiotakis, G., “Assessment of osteoarthritis severity by wavelet analysis of the hip joint space radial distance signature,” in [*BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*], 1–6, IEEE (2008).
- [14] Kellgren, J. H. and Lawrence, J. S., “Radiological Assessment of Osteo-Arthrosis,” *Annals of the Rheumatic Diseases* **16**(4), 494–502 (1957).
- [15] Dalal, N. and Triggs, B., “Histograms of oriented gradients for human detection,” *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* **I**, 886–893 (2005).
- [16] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J., “Active shape models-their training and application,” *Comput. Vis. Image Underst.* **61**, 38–59 (Jan. 1995).
- [17] Chen, C., Belavy, D., Yu, W., Chu, C., Armbrrecht, G., Bansmann, M., Felsenberg, D., and Zheng, G., “Localization and segmentation of 3d intervertebral discs in mr images by data driven estimation,” *IEEE Transactions on Medical Imaging* **34**(8), 1719–1729 (2015).
- [18] Yeghiazaryan, V. and Voiculescu, I., “An overview of current evaluation methods used in medical image segmentation,” Tech. Rep. RR-15-08, Department of Computer Science, Oxford, UK (2015).
- [19] Sadat Fasihi, M. and Mikhael, W. B., “Overview of current biomedical image segmentation methods,” *2016 International Conference on Computational Science and Computational Intelligence* , 803–808 (2016).