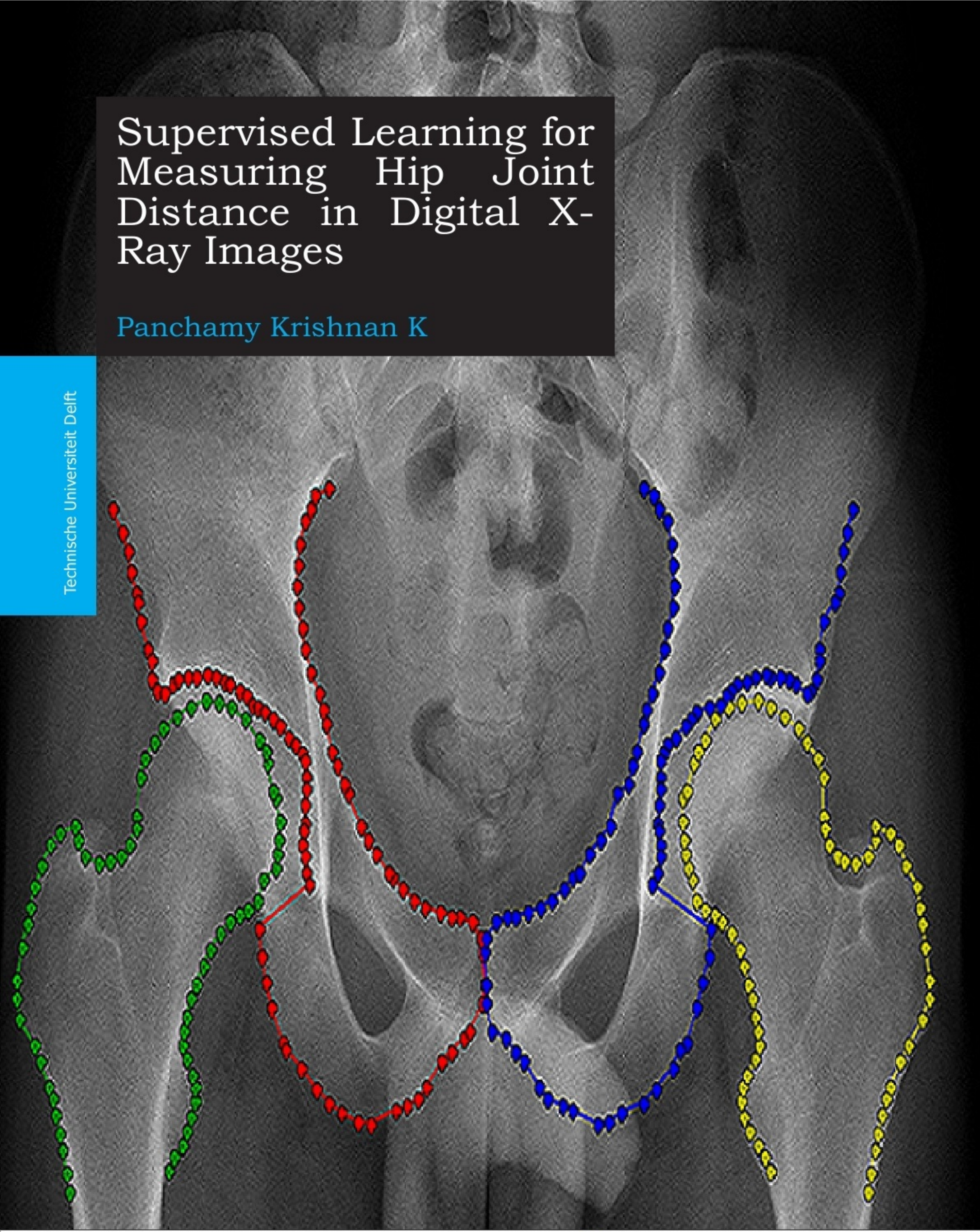
Supervised Learning for Measuring Hip Joint Distance in Digital X-Ray Images

Panchamy Krishnan K

Technische Universiteit Delft



TU Delft Delft University of Technology

Clinical Graphics See Beyond

Challenge the future

Supervised Learning for Measuring Hip Joint Distance in Digital X-Ray Images

by

Panchamy Krishnan K

in partial fulfillment of the requirements for the degree of

Master of Science
in Computer Science

at the Delft University of Technology,
to be defended publicly on Monday August 31, 2015 at 12:00 AM.

Student Number:	4411617	
Thesis Committee:		
Chair:	Prof. Dr. Elmar Eisemann,	TU Delft
University Supervisor:	Dr. Anna Vilanova,	TU Delft
Company Supervisor:	Ivo Flipse,	Clinical Graphics B.V
Committee Member:	Prof. Erik Jansen,	TU Delft

This thesis is confidential and cannot be made public

TU Delft Delft University of Technology

Clinical Graphics See Beyond

Contents

LIST OF FIGURES

LIST OF TABLES

Supervised Learning for Measuring Hip Joint Distance in Digital X-Ray Images

Panchamy Krishnan Krishnakumari

Computer Graphics & Visualization Department,
Delft University of Technology, Netherlands
E-mail: kpanchamy@gmail.com

Abstract: Osteoarthritis is a degenerative joint disease which is hard to diagnose objectively and may vary based on the surgeon. This disease is usually diagnosed by measuring several characteristic features of Hip X-rays mainly the joint distance between the femoral head and acetabular cup. Hip joint distance reduction is a clear symptom of Osteoarthritis as it suggest cartilage disappearance. Hip joint distance metric involves segmentation of the femur and pelvis in X-rays, which is a challenging task because of contrast variations as well as external factors like anatomical and pose-variation.

A multiscale approach based on Machine Learning is presented in this work for the segmentation of multiple bone structures. This technique uses landmark detection via data-driven joint estimation of image displacements and introduces a unique refinement step for improving the accuracy of detection. The detection is based on supervised learning using manually annotated landmarks. Therefore, the landmark placement along the edge of the bone has been covered in detail. The detected landmarks are then used to determine the joint distance in several locations along the hip joint. Aside from the segmentation technique, this work also introduces novel joint distance metrics which can be used to detect joint space narrowing. A detailed quantitative evaluation proved this work to be superior to the current state-of-the-art segmentation that handles multiple bone structures and is the first in evaluating the joint space width metric. We have also considered and discussed in brief the impact of such a system for diagnostic purposes.

Keywords:
Joint space, Landmark detection, Active Shape Models, 2D gradient profiling, X-ray image, Automatic segmentation, Supervised learning, Osteoarthritis

1. INTRODUCTION

Conventional radiographs remain the primary examination for detecting signs of degenerative disease in hip and knee joints, although MRI is a superior technique for revealing degenerative changes in smaller areas. Osteoarthritis(OA) in hip is commonly diagnosed using Hip Anter-Posterior(AP) X-rays. The radiological hallmarks of OA in Hip-AP X-rays are osteophyte formation, joint space narrowing, sclerosis and cyst formation as shown in Fig. 1. The severity of the disease can vary and there are different grading system to score the severity, all based on these hallmarks and other clinical symptoms. Joint space width (JSW) measurement remains the major criterion in the diagnosis of OA from radiographs and for monitoring progression of the disease.

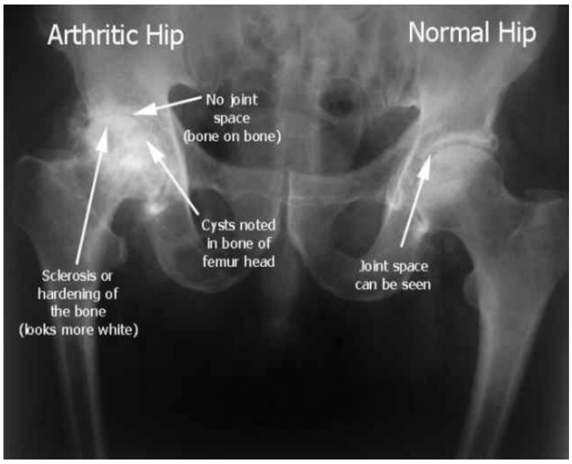

Arthritic Hip
Normal Hip
No joint space (bone on bone)
Cysts noted in bone of femur head
Sclerosis or hardening of the bone (looks more white)
Joint space can be seen

Fig. 1. Radiological Hallmarks of Osteoarthritis [1]

The JSW is generally measured by a trained physician using a graduated magnifying lens and is prone to the subjectivity and variation associated with observer measurement as well as being time consuming. There are computerized methods for measuring the JSW but requires manual pre-processing like cropping, centering, etc. of the X-rays [12]. They also require standardized radiographs and is usually constrained by the pose and shape variation of the bone which requires more interactions from the user leading to a non-reproducible subjective metric.

X-rays also make it more challenging to diagnose OA. This is mainly due to the characteristics of the X-ray images like overlapping of the bones with organs, tissues, etc. and the noise in the image due to the discrete nature of the X-ray photon source. There are also the factors like the

[1] http://www.drwolgin.com/Pages/Osteoarthritis.aspx

inhomogeneity of the X-ray due to varying bone density among different patients and absence of definitive edge between organs as the neighboring organs have similar X-ray absorption rates. The shape of the bone differs among patients as well especially the pelvic bone varies between men and women. Also the patient pose may vary which makes the bone to be located at different parts in different images.

These reasons lead to the need for an advanced automatic technique for extracting the joint distance from the X-rays for diagnosing OA. Most of the available research have been focused on automatic segmentation of bones in radiographs, mainly single bone [10][14][1], or classification of OA based on manually classified X-rays [12]. This paper tries to bridge the gap between these two as a first step to developing a fully automatic diagnostic tool for detecting OA. The paper is based on state-of-the-art method to segment multiple bones - femur and pelvis - in radiographs and shows an improved accuracy and robustness in segmentation than the current methods along with providing a joint distance metric. The method discussed here is robust to work with shape, pose and contrast variations in X-rays. The method has been refined to provide improved segmentation accuracy along the joint gap rather than in segmentation of the bones itself.

This thesis has been structured to briefly introduce the current front runners in bone segmentation which is mainly landmark detection followed by shape regularization. The paper proposes a multi-scale machine learning approach for this purpose [1]. The necessary background to understand the paper is described in the Background section. The training data set plays a major role in machine learning and hence the process in obtaining this is explained in detail. The pipeline, landmark detection based on supervised learning and shape refinement technique which constitutes a classic ASM followed by two dimensional gradient profiling in each scale has been described in Method section. Implementation details including the hardware specifications, the parameters and software requirements have been added to aid in analyzing the performance. The method is evaluated for its accuracy in segmentation as well as in obtaining the joint space width(JSW) metric using leave one out cross validation with the manually segmented contours as the ground truth. Even though the aim of the method is to provide an accurate JSW metric, the method has also been evaluated for its accuracy in the segmentation of femur and pelvis. The possible tuning and improvements of the method are discussed in the Conclusion and Future Work.

2. RELATED WORK

There is no article that addresses the automatic extraction of JSW metric from X-rays but there are several research focused on segmentation of bones. This section briefly explains the state-of-the-art segmentation techniques for bones in X-rays and other modalities. The section gives an overview of the conventional segmentation methods as well and explains the constraints of these methods.

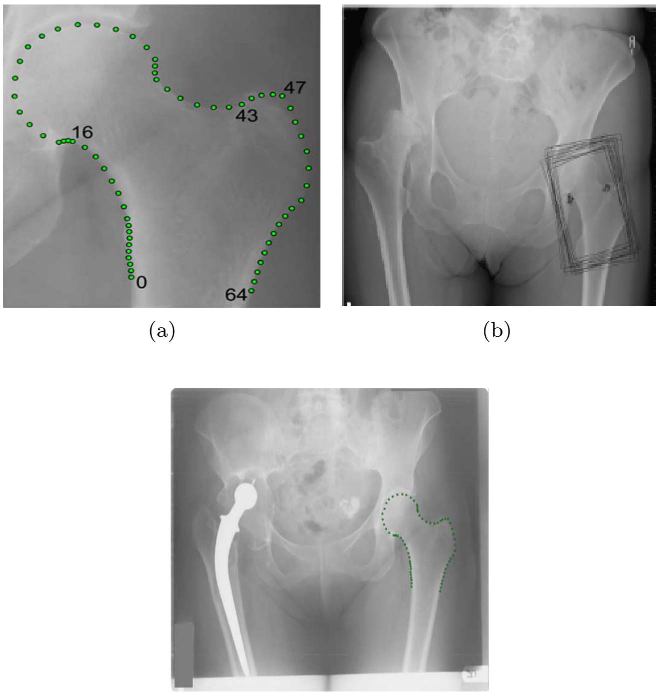
Classic image segmentation algorithms are mainly based on edge detection and deformable model-based techniques. There have been a lot of research done on medical image segmentation, but the accuracy and robustness of most of these algorithms do not extend to pose, contrast and shape variation of the X-rays usually. And the error from these methods is not acceptable for the JSW metric as the joint distance for a normal hip is usually in the range of 11 to 13mm and these methods generates segmentation error in the range of 4mm or more making it highly unreliable [11]. Several research has been done in extracting bone structures from the images where the success rate is based on whether the bone has been extracted or not but the segmentation accuracy has not been studied in detail. Recent research has focused segmentation of bones in medical images on landmark detection with improved accuracy and this is the basis of our work.

2.1 Segmentation based on Landmark Detection

Most of the existing methods make assumptions about the femur pose and are tested on X-ray images with similar quality. They do not correct the orientation variation which occurs due to the varying patient postures as described in [11]. A new approach based on Random Forest(RF) regression voting in a sliding window was proposed to handle wide range of image quality and femur poses [10]. The approach falls into the category of deformable models although the Random Forest is an ensemble learning method. The method is based on using multiple decision trees trained on random subset of features to predict the position of a point(landmark) relative to the sampled region. The features can be any descriptive features of the images, Lindner *et al.* uses Haar features. Since multiple decision trees are used, the prediction of the position of the landmark is a weighted average from all the trees. This in general is more accurate than using a single tree and this is the main advantage of Random Forests.

Lindner *et al.* presents a two stage process - a global search followed by a local search. Since both stages are based on machine learning, they both require training and testing. In the global search, the aim is to find the global position and alignment of the femur by predicting the center of a reference frame as shown in Fig. 2b. The reference frame is a patch enclosing the 16 and 43 landmark since they are relatively stable and remain constant with respect to each other in most femur shapes as shown in Fig. 2a. For detecting the reference frame center, the training is done by sampling the Haar features of reference frame in all the training data and the decision trees are trained on this to detect the frame center. During testing, multiple random patches are sampled in the test image and each of these patches is tested on the decision trees to vote for the frame center.

After these two landmarks are detected in the global search, the local search uses Constrained Local Model(CLM) to detect the rest of the landmarks. In the CLM framework, the aim is to generate a response image or a point cloud for each landmark independently. For this, the training is done by sampling random patches around each landmark and the Haar features of these patches are trained on multiple decision trees to predict the displacement of the patch center to the landmark. In the testing part, the global search is used for initialization of all the landmarks



(a)

(b)

(c)

Fig. 2. RF landmark Detection (a) Manually Annotated Contours (b) Reference Frame Detection (c) Segmentation Result [10].

and then random patches are sampled around each initialized landmark in the test image. These patch features are tested on the decision trees to vote for the displacement and the votes generate a point cloud for each landmark independent of other landmarks. A statistical shape model is fitted into the point cloud of all the landmarks to find the best combination of points and a result is shown in Fig. 2c. The system reported an accuracy of an overall mean point-to-curve error of less than 0.9 mm for 99% of the images from an 839 images of mixed quality making it the most accurate fully automatic system for segmenting the proximal femur in AP pelvic radiographs.

In Chen *et al.* paper, both femur and pelvis segmentation are studied based on similar concept as the Lindner *et al.* method [1]. Both Lindner *et al.* and Chen *et al.* work is based on the assumption that the displacement of a test patch should be similar to those of training patches with similar visual features. Chen *et al.* also uses random subset of features to predict the position of a landmark relative to the sampled patch but introduces a multi-scale approach. The patch features used in this method are Histogram of Orientated Gradients(HoG). The method does not incorporate the decision trees and Lindner *et al.* method predicts the position independent of the other landmarks and other patches, whereas Chen estimate the position jointly from all the patches together in a data-driven way. The method also incorporates geometric constraints into its equations to provide better accuracy. The geometric constraint is based on the assumption that if two patches are predicting the same landmark position, then there should be a relation between the patch centers and the displacement. This is explained in detail in the Background section. After landmark detection, the predicted landmark positions are regularized by a statistical shape model to get the final segmented shape contour. The sparse shape composition is used for the shape regularization which keeps the local shape information and is shown to have better performance than the general shape models [1]. Chen *et al.* method was evaluated with 100 training images, 188 test images for proximal femur, 100 training images and 163 test images for pelvis and provided a success rate of 98.4% with 1.3mm mean error and 98.8% with 2.0mm mean error for proximal femur and pelvis respectively.

There are many other landmark detection methods as well. One of the other related works in landmark detection that is interesting for our work is based on gradient profiling and shape model [14]. This is not a fully automated segmentation method. It needs 8 initialization points in the pelvis which is then registered to the mean model of the SSM. The main step we are interested in is the statistical appearance model(SAM) generated for the refinement step. The SAM is similar to gradient profiling. The principle behind SAM is that for each landmark, a gradient profile along the normal of that landmark is computed for each training image. From the gradient profiles of all the training images, a statistical model can be constructed for the gradient profile. Thus, we can generalize the appearance variations of that landmark and use this to refine the position of the detected landmark position iteratively. In this paper, instead of taking a 1D gradient profile along the normal, a rectangular patch along the normal is considered for generating the gradient profile as shown in Fig. 3. For each landmark, the mean of the rectangular patch is used for constructing the SAM. This provides more information about the gradient at that landmark than considering just one direction.

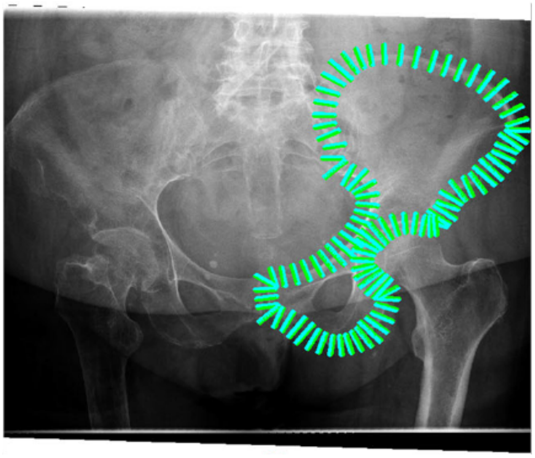


Fig. 3. Sampling Regions for Gradient Profiling [14]

Most of the landmark detection methods incorporate the statistical shape model which is typically used to regularize the detected landmark positions using global topological information. There has been a lot of new shape models proposed, most of them inspired by Cootes *et al.* Active Shape Model [5]. In this paper, we base our work on Chen *et al.* as it works on multiple bones - femur and pelvis - with competent accuracy. For the refinement step, the general Active Shape Model followed by 2D gradient profiling is used and provided accurate results. The shape information already encoded in Chen *et al.* method through their use of multiple landmarks(subshapes) already provide a good local topological information. The refinement step which includes ASM followed by gradient profiling in each scale provides improved accuracy in each scale.

3. BACKGROUND

This section will provide the necessary background knowledge to understand the technical context of this work. It will briefly explain the ASM from Cootes *et al* [5]. The section also explains the feature vectors which are extensively used throughout this work, feature matching based on geometric constraints from Chen *et al* [1] and feature selection for reducing dimensionality [2]. For more in-depth understanding on these topics, we refer to the literature.

3.1 Active Shape Models

Active Shape Models is a model based segmentation method introduced by Cootes and Taylor [5]. The method is based on the principle that a shape can be represented by a mean shape and its variations. The mean shape and the variances constitutes the statistical shape model(SSM) which contains all the parameters that are needed to define that shape. The SSM is used to find a shape of the Shape model in a new image. Initially, a set of landmarks in the new image are defined and then the shape defined by these landmarks is deformed according to the SSM to provide the best fit possible within SSM. The deformation is based on finding correspondences between the new shape and the shape defined by different SSM components and iteratively minimization a cost function for the fit. Thus the deformation is constrained by the variations defined in the SSM. The overview of the method is briefly explained below. A more detailed explanation can be found in the Cootes *et al.* paper [5][4].

There are two main phases in ASM - constructing the SSM and finding a shape of the SSM in a new image. For constructing the SSM, all the shapes in the training data have to be aligned to the same shape as a reference which is the first shape in the training data initially. The simplest way to do alignment is using Procrustes Analysis [7]. The reference shape is then updated to the mean shape of the new aligned shapes. All the shapes are then aligned to the new reference and this step is repeated until convergence. After the shapes have been aligned, Principal Component Analysis(PCA) is applied to find the principal components or modes that explains as much of the shapes as possible, the variances in each of these components and the mean shape. The SSM components are the eigen vectors of the centered shapes in the training data and the variances are the eigen values of these shapes. If we apply PCA to the data, we can approximate any of the training set, $\mathbf{x}$ which is an n-dimensional shape with L points using equation(1).

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}. \tag{1}$$

where $\bar{\mathbf{x}}$ is the SSM mean shape having point correspondences with $\mathbf{x}$ and same dimension, $\mathbf{P}$ is the SSM principal components and the $\mathbf{b}$ is the shape parameters corresponding to the SSM components that deforms the shape, $\mathbf{x}$ [4]. This is the basis for fitting a new set of landmarks to the SSM. The steps for the fitting are explained in the Method section.

3.2 Histogram of Orientated Gradients

Feature vectors are a common terminology that is used throughout this paper. A feature vector of an image is an n-dimensional vector of numerical features that represent the image and distinguishes images that are different for a given purpose. Histogram of Orientated Gradients(HoG) is a feature descriptor and is based on the principle that local patch appearance or the shape can be described by the gradient intensities and directions [6].

HoG is described for a local patch of the image and is computed by dividing this patch into cells. Each of the cells of the patch contains pixels and for each pixel, the dominant gradient in n-directions are computed and then the histogram of these gradients are added together for all pixels in the cells as shown in Fig. 4. The number of bins in the histogram is the number of gradient orientations which is known. These histogram bins per each cell is concatenated together to generate the feature descriptor. Fig. 4 shows an 8×8 patch with cell size (2,2) and assuming the number of orientations is 9. Then each cell is 4×4 which gives 16 pixels per cell. This leads to 9 orientated gradients for each pixel and these are added across one cell, so there are 18 histogram bins for each cell. Since there are 4 cells and the histograms from each cell is concatenated together to form a HoG vector of a patch, the feature vector dimension for the patch is 36. The dimension is obtained by multiplying the number of orientations and total number of cells in the patch.

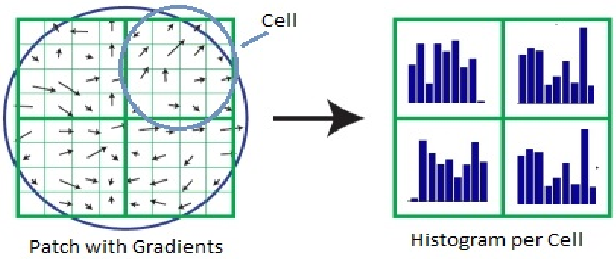


Fig. 4. Sample HoG Descriptors - 4 HoG descriptors, 1 per each cell(right) is computed from accumulating the gradient magnitude and direction of 16 pixels/cell(left)

3.3 Feature Matching with Geometric Constraints

The basis of machine learning methods is to match feature vectors from a new test image to the training feature vectors. Once they are matched, use the known parameters of that matched training patch to predict the unknown test patch parameters. One of the major contributions of Chen *et al.*'s paper [1] was predicting the displacement of the patch from the landmark, by using geometric constraints which provide information about inter-patch dependence. The feature matching based on this geometric constraint is explained with a simple example.

The basic principle of feature detection is that the displacement of a test patch should be similar to those of training patches with similar features, i.e. HoG. From the Fig. 5, this can be seen as the test patch feature $\mathbf{f_1}$ is similar to $\tilde{\mathbf{f_1}}$ and $\tilde{\mathbf{f_6}}$ features from the training data, so the assumption is that displacement vector $\tilde{\mathbf{d_1}}$ and $\tilde{\mathbf{d_6}}$ are

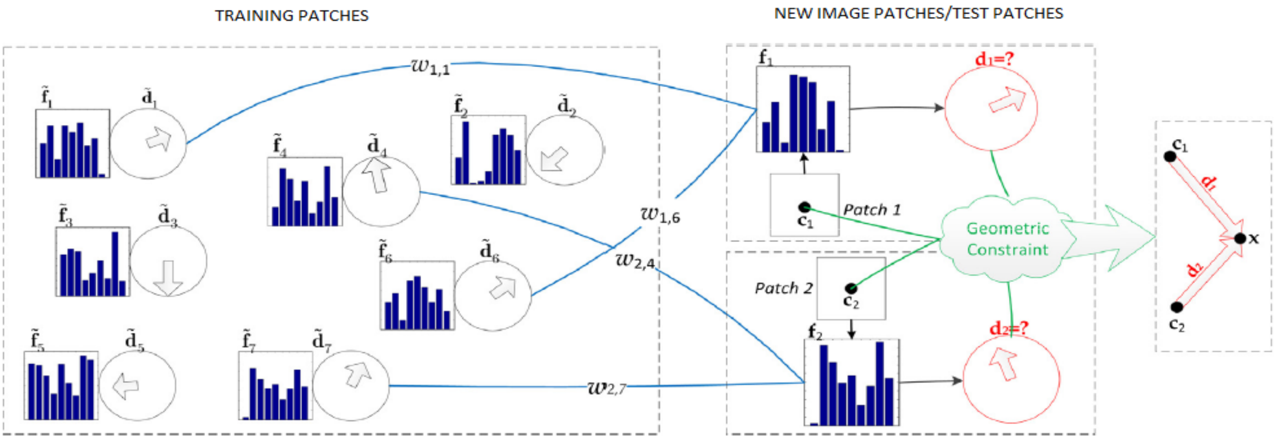


Fig. 5. A Simple Scenario of Feature Matching with Geometric Constraints for Landmark Position Prediction [1]

similar or their difference must be minimized, the same for feature patch 2. $\mathbf{f_2}$ is similar to $\tilde{\mathbf{f_4}}$ and $\tilde{\mathbf{f_7}}$. In Chen *et al.*'s paper, the displacement is also constrained based on the assumption that if two patches vote for the same landmark position, ideally the voted position should be similar as they are voting for the same landmark. The landmark position, x is found from the patch centre of the test patch and the displacement predicted from the training data for that patch. The landmark position predicted by patch 1 is $\mathbf{c_1} + \mathbf{d_1}$ and patch 2 is $\mathbf{c_2} + \mathbf{d_2}$. So, here the patches with features $\mathbf{f_1}$ and $\mathbf{f_2}$ are predicting the same landmark $\mathbf{x}$, therefore the assumption is that $\mathbf{c_1} + \mathbf{d_1} = \mathbf{c_2} + \mathbf{d_2}$. However this is not true, therefore for an optimal prediction the $(\mathbf{c_2} - \mathbf{c_1})$ and $(\mathbf{d_1} - \mathbf{d_2})$ must be minimum. This is the geometric constraint imposed on the prediction. Thus, both feature matching and geometric constraint are considered for predicting the displacements.

3.4 Feature Selection

The feature vectors are an important part of the detection process. Usually these feature vectors are high dimensional(for example, 72 dimensions if there are 18 orientations as explained for HoG before). This leads to high processing time during testing as feature matching is performed during testing. This makes it the highest computational bottle neck. A simple way to optimize this is by feature selection. The concept behind it is that a subset of the features is sufficient to represent the entire feature space. One such dimensionality reduction method is based on similarity and dissimilarity measure between multiple feature vector pairs from the training data introduced by Chen *et al* [2].

Their method selects features such that the data similar in the displacement corresponds to similarity in the feature space [2]. The steps involved in reducing the dimension of the feature vectors from m to n are as follows:

(1) Create N pairs of feature vectors and the corresponding displacement pairs
(2) A similarity and dissimilarity weight is computed using the euclidean distance between the displacement pair, the variance and maximum of these distances.
(3) A similarity and dissimilarity measure in feature space is computed using the difference between the feature pair, the similarity and dissimilarity weight.
(4) Randomly select n indices from m
(5) A transform matrix is constructed that selects the features at the chosen indices.
(6) A score is calculated using the similarity, dissimilarity metric and the transform matrix for each m
(7) Sort the values of m according to descending order of the score.
(8) The first n indices of the sorted m are the new updated indices.
(9) Repeat from (5) until convergence.

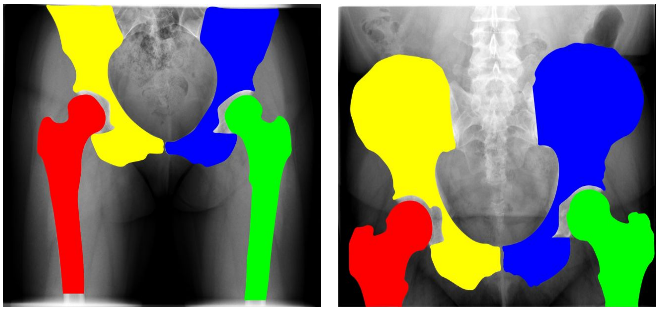
For a more detailed explanation and equations, Chen *et al.* paper can be referred [2]. The output of the method is a transform matrix that selects the n features from the m-dimensional feature vector.

4. DATA

As training data is a major part of learning algorithms, this section describes the preprocessing step required to prepare the data for creating the training data. As the segmentation is based on landmark detection, the preprocessed data corresponds to the landmarks along the edge of femur and pelvis bone. This section describes the process of obtaining these landmarks from the input which is DICOM X-ray datasets.

The data for this project has been obtained from the Reiner de Graaf Hospital at Delft in Netherlands. We obtained around 400 DICOM files containing the X-ray image and additional information like pixel spacing, image size, etc. The X-ray image have high resolution with most of their size in the range of 2048 × 2500 or more with pixel spacing around 0.168mm in x and y directions or 0.143mm. Since the images have different spacing, they have been normalized to have the same spacing which is 0.1mm/pixel in our case. This standardizes the process for computing the joint space width. Out of the 400 X-ray images obtained, 114 femur and 72 pelvis including both left and right bone structures were manually segmented for the training data.

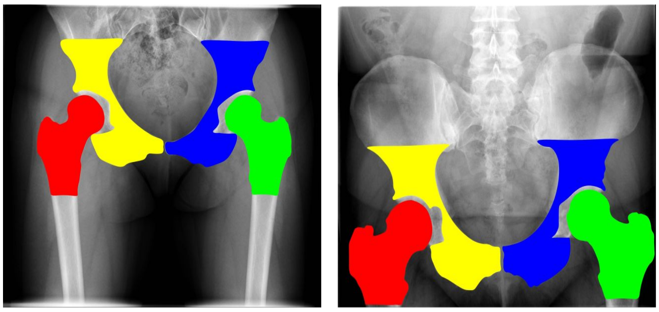
The accuracy of the result depends on the reliability and robustness of the training data. The more variation in the training data, better the detection, as the detector can detect as many variations as available in the training data. So, we tried to include as much training data as possible and as many variation as possible. The basis of the training data is the manually annotated contours of femur and pelvis. One of the main drawback of Chen *et al.* [1] is that the X-ray images for the training data were homogeneous with all of them containing full femur and pelvis. But this is an ideal case. In practice, the femoral shaft might be too small and top part of pelvis might be cut off due to the pose variance of the patient. This will lead to manual segmentation as shown in Fig. 6.



(a) (b)

Fig. 6. Manual Segmentation (a) Missing Top Part of Pelvis (b) Small Femoral Shaft

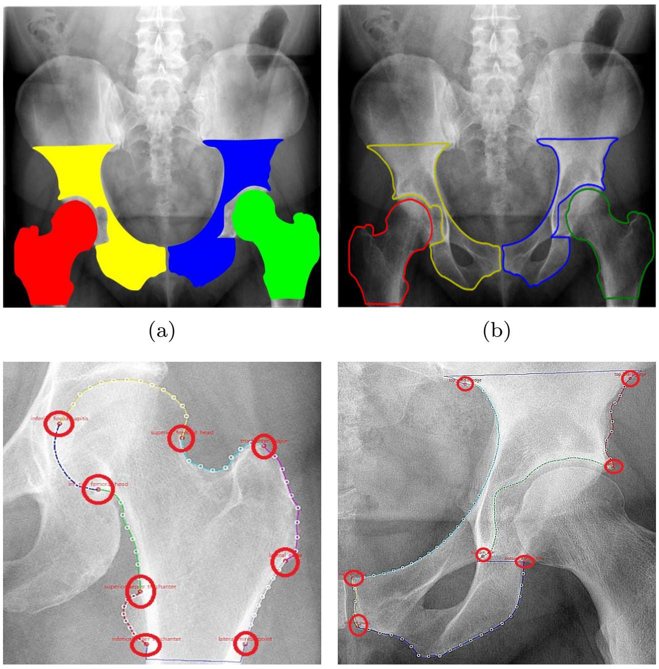
This will lead to error in the detection of landmarks at these parts of the femur and pelvis which will compound to more error. Therefore, the manual segmentation was redefined so that it only segments the part that is common in all X-ray images as shown in Fig. 7.



(a) (b)

Fig. 7. Redefined Manual Segmentation

The segmentations are based on guidelines provided by experts in hip anatomy. The masks are created using Adobe Illustrator with the help of gradient images to aid in better edge visualization and each bone structure is labeled by distinct primary colors. The contours are extracted from the mask as shown in Fig. 8b. The contours contain around 300 to 400 points but these are not landmarks. Landmarks need to have distinct features like corners or unique edges. Also, the landmarks of all the training data have to be aligned with each other. This implies the n^{th} landmark in one image must correspond to the same landmark at similar position in all the other images. Point-set registration is a common solution for finding such correspondences. The iterative closest point(ICP) based on least square differences[3] and the Gaussian mixture model based registration[8] was implemented, but did not yield satisfactory results. The main reason was that the shapes varied from X-ray to X-ray ranging from bigger to rounder femoral head, missing lesser trochanter and for the pelvis, it ranged from being wide to narrow. These registration methods could not account for these variations.



(a) (b)

(c) (d)

Fig. 8. Complete Pre-processing Pipeline (a) Create Mask using Illustrator (b) Extract Contours (c) Femur Landmark Distribution (d) Pelvis Landmark Distribution

Therefore, a tool has been used with which the X-ray image and the contour extracted from the mask can be imported and experts can place dominant landmarks in femur and pelvis manually. Once these landmarks are placed, a defined number of points are distributed evenly between these landmarks to generate the complete landmark set. There were 8 landmarks identified as significant for femur and 7 for pelvis. The defined landmarks and their position is explained in the Appendix A. Considering the evenly spaced points between these dominant landmarks, the landmark count for femur is 69 and for pelvis is 95. The full landmark set for femur and pelvis is shown in Fig. 8. This landmark set is used to create the training data as explained in the next section.

5. METHOD

This section explains the stages of the pipeline needed to get from the DICOM dataset to the final joint space width(JSW) metric. The pipeline starts with the pre-processing of the X-ray datasets into landmarks which was explained in the section 4. These landmarks are then used to create the training data for the three main stages of the method - landmark detection, Active Shape Model(ASM) and gradient profiling. Next, in the testing stage, the training data is used for detecting the landmarks in femur and pelvis in the initialization scale and in the other scales, the landmark detection is followed by refinement using ASM and gradient profiling to get the final segmentation. The segmentation of the femur and pelvis is used to define the JSW. These different stages of the pipeline are shown in Fig 9 with descriptive images and detailed explanation of each stage is covered in this section.

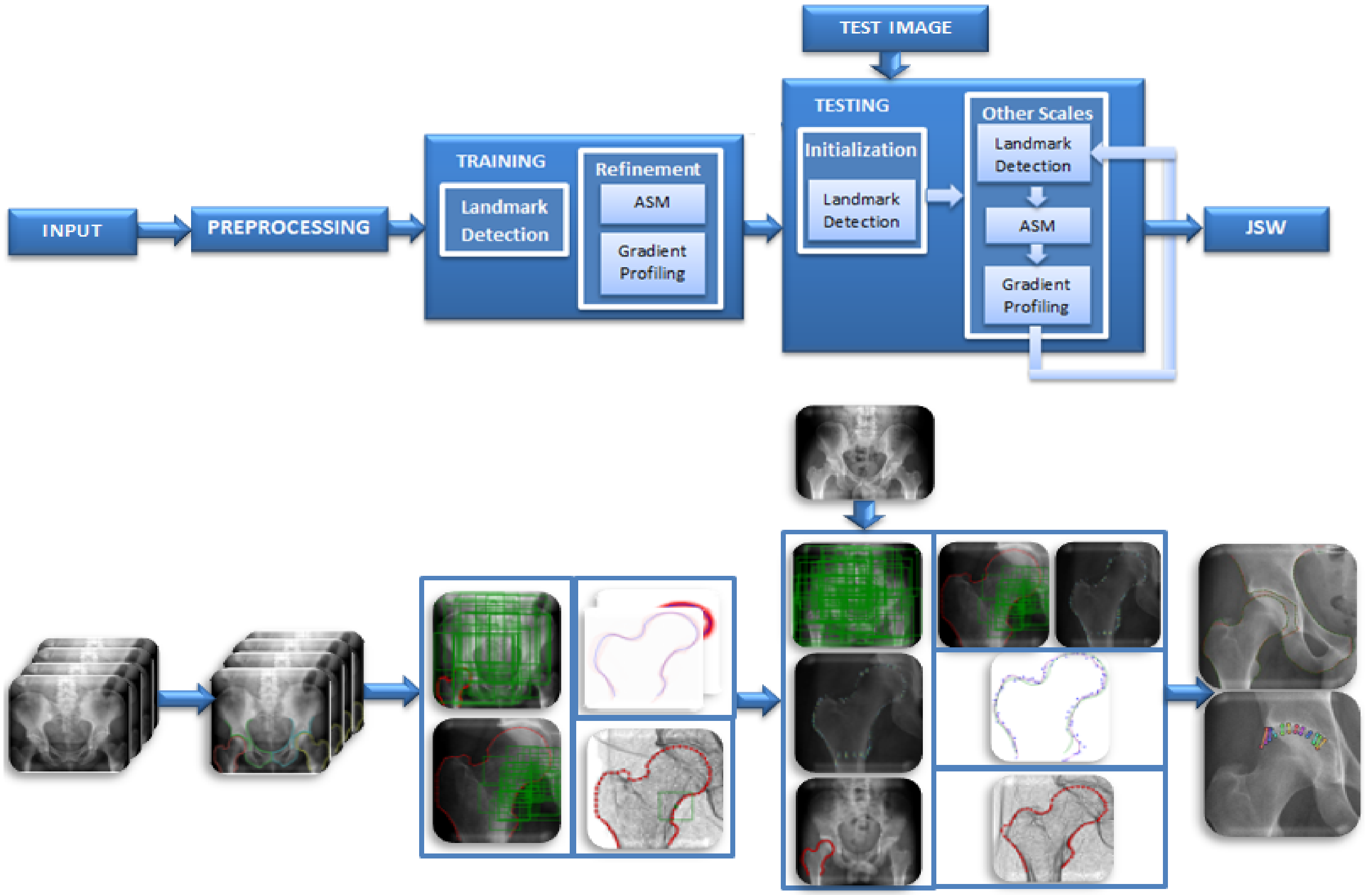


Fig. 9. The global pipeline of this work to get from a DICOM X-ray dataset to a JSW metric given a new X-ray image. The corresponding visual pipeline of each stage is also shown.

5.1 *Training*

Since all the three main stages of the method - landmark detection, ASM and gradient profiling - are learning based, they need training and testing. For all the methods, the registered landmarks obtained as explained in section 4 and the X-ray image is the basis for creating the training data. There are 114 femurs and 72 pelvis manually annotated contours for creating the training data. The training data needed for these methods are extracted as explained below.

LANDMARK DETECTION For a given shape, there are $\tilde{N}$ number of landmarks. If the shape is complex, it is divided into subshapes where each subshape contains L landmarks as shown in Fig 10a where the femur shape with $\tilde{N}$ landmarks is divided into several subshapes(each subshape is denoted by a color) containing L landmarks. The number of subshapes for a given shape is given by $\tilde{N}/L$. The landmark detection is done independently for each subshape. The principle behind landmark detection is to approximate the position of a landmark in a given subshape based on the image features of the area surrounding that subshape. The image features are usually defined for a patch extracted from the image and the area is defined by two parameters - patch size and sampling radius.

The patch size defines the dimension of the patch which can be square or rectangular and sampling radius defines how far the patches are sampled from the subshape as shown in Fig 10d where a patch is sampled for a subshape at (c_x, c_y) where d<sampling radius. The image features used here are multilevel Histogram of Orientated Gradients(HoG) as per Chen *et al.* paper. A multilevel HoG of a patch is obtained by computing the HoG of the patch at different cell sizes and concatenating the obtained feature vectors. Since the concatenated feature vector is high dimensional, the feature selection method explained in section 3.4 is used to compute a transform matrix to reduce the dimension. The generation of training data for each subshape and also the multiscale system are explained below.

Per Subshape

The HoG feature vectors for patches are sampled for a subshape, the patch centres of these patches and the displacements of the patch centre to the landmarks in the subshape constitute the training data for landmark detection. This is done for all subshapes for a given shape in an X-ray image and for all X-ray images in the dataset. The overview of the steps involved in creating training data for a subshape with L landmarks of a shape with N landmarks in an X-ray image is as follows:

(1) Randomly sample k patches around the subshape with the given patch size and sampling radius
(2) Find displacement in x and y direction from each patch center to the L landmarks in the subshape

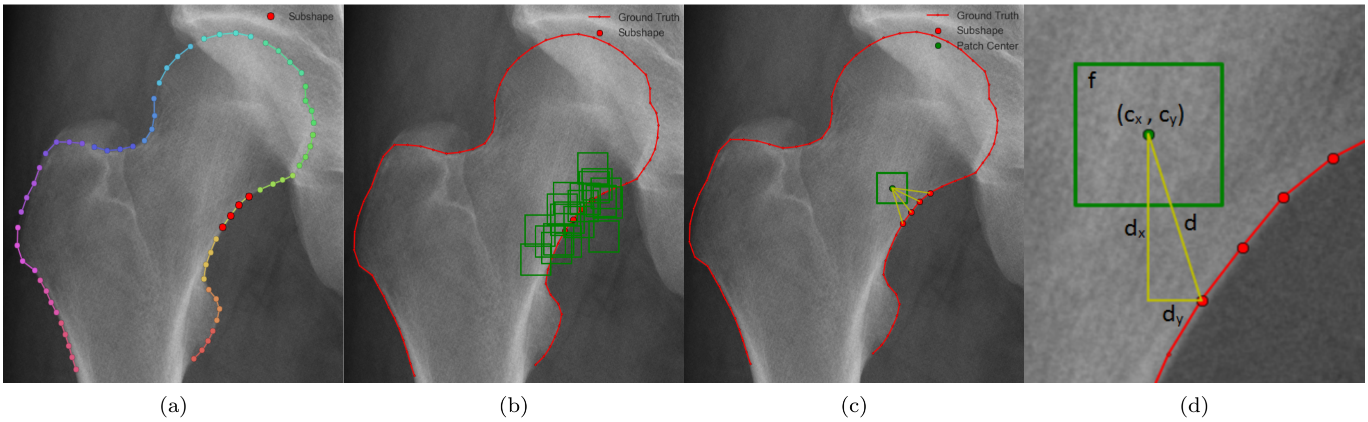


Fig. 10. Training Pipeline for a Subshape (a) A Subshape with L Landmarks (b) Sampling Patches around the Subshape (c) Computing Displacement from Patch Center to Each Landmark in the Subshape (d) Training Data for Each Patch and for a Given Landmark in the Subshape - Feature Vector **f**, Patch Centre (c_x, c_y) and Displacement Vector (d_x, d_y)

(3) Extract multilevel HoG feature vectors for all the sampled patches
(4) The patch centers, displacements and feature vectors for all the sampled patches for a given subshape is stored.

The detailed visual explanation of generating the training data for an X-ray image for a given subshape is shown in Fig 10. This is done for all the X-ray images with manually annotated landmarks for a given subshape. The training data of the k patches for an X-ray image for a given subshape in matrix form is as follows:

$$\tilde{\mathbf{D}} = \begin{bmatrix} \tilde{dx}_{11} & \tilde{dx}_{12} & \dots & \tilde{dx}_{1k} \\ \tilde{dy}_{11} & \tilde{dy}_{12} & \dots & \tilde{dy}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{dx}_{L1} & \tilde{dx}_{L2} & \dots & \tilde{dx}_{Lk} \\ \tilde{dy}_{L1} & \tilde{dy}_{L2} & \dots & \tilde{dy}_{Lk} \end{bmatrix} \quad \tilde{\mathbf{F}} = \begin{bmatrix} \tilde{f}_{11} & \tilde{f}_{12} & \dots & \tilde{f}_{1k} \\ \tilde{f}_{21} & \tilde{f}_{22} & \dots & \tilde{f}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{f}_{m1} & \tilde{f}_{m2} & \dots & \tilde{f}_{mk} \end{bmatrix}$$

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{cx}_1 & \tilde{cx}_2 & \dots & \tilde{cx}_k \\ \tilde{cy}_1 & \tilde{cy}_2 & \dots & \tilde{cy}_k \end{bmatrix}$$

$\tilde{\mathbf{D}}$ is the displacement matrix where $\tilde{dx}$ and $\tilde{dy}$ are the displacement in the x and y direction for one landmark and $\tilde{\mathbf{D}} \in R^{2L \times k)}$. $\tilde{\mathbf{F}}$ is the feature vector matrix where m is the dimension of a HoG feature vectors and $\tilde{\mathbf{F}} \in R^{m \times k}$. $\tilde{\mathbf{C}}$ is the patch centres matrix where $\tilde{cx}$ and $\tilde{cy}$ are the patch centre position in the x and y position for one patch and $\tilde{\mathbf{C}} \in R^{2 \times k}$. As the feature vectors are high dimensional, dimensionality reduction is done to reduce the feature vector dimension from m to n are computed as follows:

(1) Accumulate HoG feature vectors for all X-ray images for a given subshape
(2) Compute tranform matrix using feature selection method explained in section 3.4 for dimensionality reduction[2]
(3) Compute the dimension reduced feature vector by multiplying the feature vector with the computed transform matrix. This is done for all X-ray images.

The transform matrix and the dimension reduced feature vectors are also included in the training data. The new feature vectors are stored as matrix $\tilde{\mathbf{F}} \in R^{n \times k}$. All the steps mentioned above are repeated for all subshapes of a given shape to obtain the complete training dataset.

Multiscale System

The method proposed is multiscale with the whole pipeline being executed N times sequentially where the output of the first scale is the input of the second. The first scale is the initialization scale which aims at finding the global position of the bone in the X-ray image. For the other *N-1* scales, the training data for the landmark detection is obtained as explained above.

For all the N scales, the patch size and the sampling radius in pixels remains the same and only the resolution differs. Thus, the information contained in each patch becomes more detailed as the resolution increases. Fig 11 shows an X-ray images with increasing resolution and a patch size of 40×40 is shown for all the scales. The lower scale covers more area of the image but contains less detail. Thus, the training data for landmark detection is created for N scales with the patch size and sampling radius remains the same.

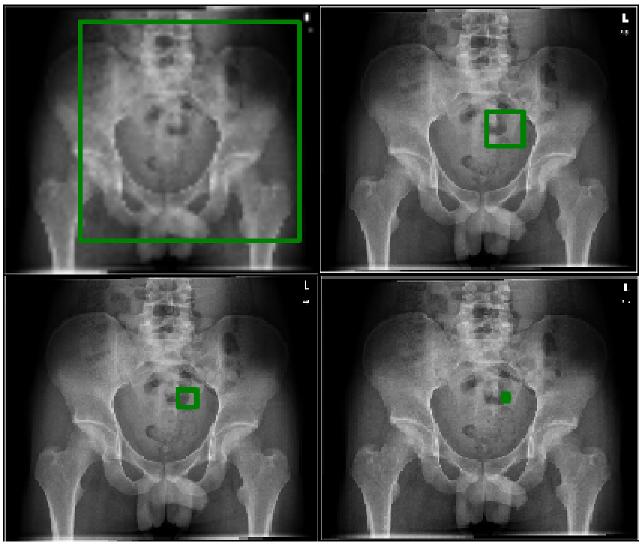


Fig. 11. Same Patch Size in 2%, 12%, 25% and 100% Resolution of the X-ray Image with the Same Patch Centre

In the initialization scale or the first scale, the concept of subshape is not used. A patch stores the displacement to all the landmarks in the shape thus $L = \tilde{N}$ where L is the number of landmarks in the subshape/subshape length and $\tilde{N}$ is the number of landmarks in the shape. Also, in the initial scale, the patches are sampled all around the X-ray image and not around the landmarks of the shape. Apart from the patch sampling and the subshape, the training data is created the same way as for other scales.

REFINEMENT The refinement step consists of 2 stages - ASM and gradient profiling. This step is mainly for regularizing the shape approximated by the detected landmarks. The refinement step is also multiscale and uses the same N number of scales as for the landmark detection. However, there is no refinement step in the initial scale as the goal of this scale is approximation rather than precision and the edges are not distinctive at that resolution. So refinement step is skipped for this scale. In the next section, creating the training data for ASM and gradient profiling is explained.

Active Shape Models

The principle behind ASM has been explained in section 3.1. The data needed for creating the training data is the set of registered landmarks of femur and pelvis. The steps for creating the training data for the ASM are as follows:

- Align the shapes to the first shape in the dataset
- Generate the mean shape from the aligned dataset
- Align the shapes to the mean shape and generate the new mean shape from the newly aligned shapes
- Generate the eigen vectors and eigen values of the aligned shapes which is the SSM components and SSM standard deviations respectively.

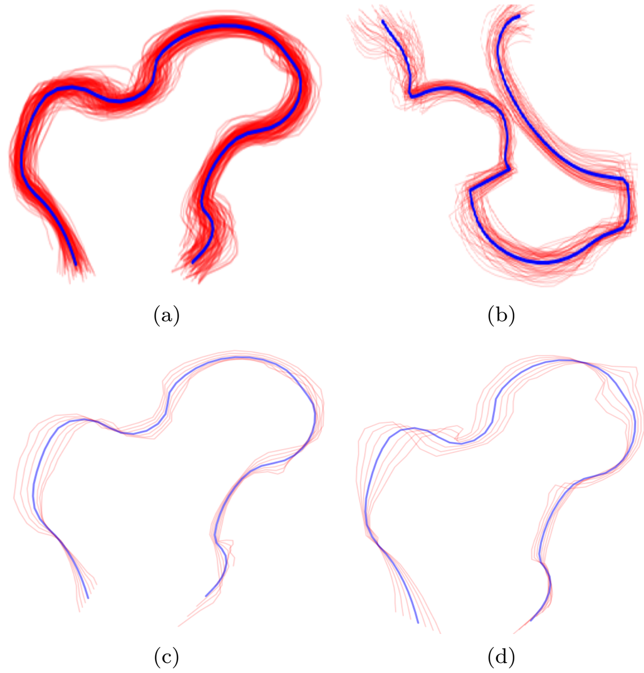


Fig. 12. ASM Training Data - Blue Line Indicates the Mean Shape (a) Femur SSM Mean Shape with Aligned Shapes (b) Pelvis SSM Mean Shape with Aligned Shapes (c) Femur SSM Component 1 - Explains 31.27% Shape Variance (d) Femur SSM Component 2 - Explains 21.78% Shape Variance

The training data for ASM is the same for all the scales. However, an SSM model for the femur and pelvis has to be created separately. Fig 12 shows the femur and pelvis SSM mean shapes and also show the first 2 SSM components of femur. This is later used for ASM fitting in the testing stage.

Gradient Profiling

The 2D gradient profiling is based on the assumption that each landmark is on a distinct edge of the whole contour. The edge is usually defined using gradients and there are different methods for computing 2D gradient of an image. Some of methods that have been tried are sobel gradient, morphological gradient with structuring element and gradient magnitude using Gaussian derivative. The Gaussian derivative gradient was found to be less sensitive to noise and edges were visually better as shown in Fig 13. Hence this was used for computing the gradients of the X-ray image.

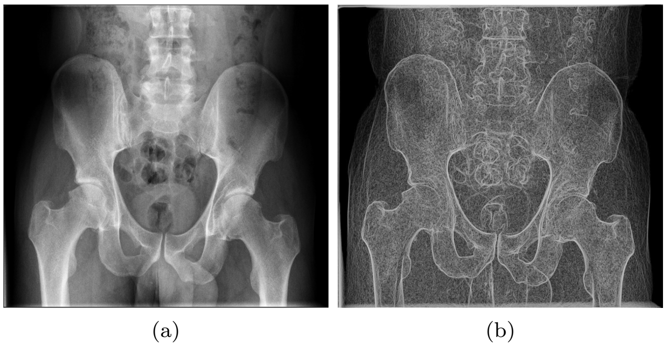
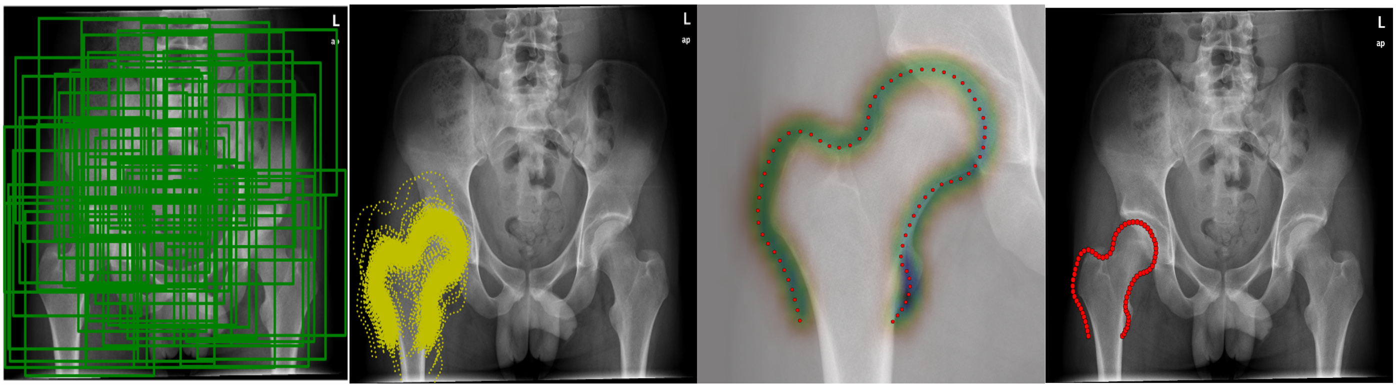


Fig. 13. Gradient Image (a) Original Image (b) Gradient Magnitude using Gaussian Derivative

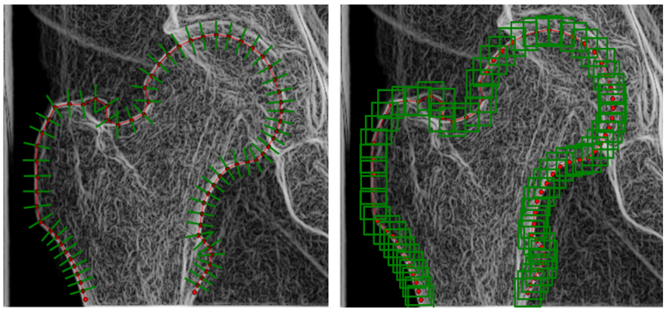
Gradient can be 1D or 2D and usually in gradient profiling, gradient is considered along a single direction, mainly the normal as shown in Fig 14 [5]. However, since the normal depends on the neighboring points and the curvature created by these points, for a single landmark, the normal direction differs a lot which leads to incorrect predictions. Xie *et al.* refined this step by considering more than one direction - a rectangular patch along the normal [14]. However, intuitively it seemed that the more direction information there is, the better the prediction. So, in our method, we considers all the directions around the landmark which is equivalent to considering a patch with the landmark as the patch centre as shown in Fig 14. This is similar to template matching - matching the edge containing the landmark in the test image to the edges in the training image for the same landmark.

For each X-ray image, the gradient using the Gaussian derivative is estimated. From the gradient image, a patch is extracted with the patch size according to the scale for each landmark as shown in Fig 14b. This patch size differs from the patch size for landmark detection. The standard deviation of the Gaussian filter, σ for finding the gradients differs for each scale. For each landmark, the gradient profiles are created by extracting the patch from the gradient image around that landmark in all the X-ray images. These gradient patches constitute the training data for the gradient profiling. The patches are extracted



(a) (b) (c) (d)

Fig. 15. Testing Pipeline for Initialization Scale (a) Sampling Patches around the Whole Image (b) Positions Voted by all the Test Patches for the Landmarks in the Shape (c) Response Image for Each Landmark with the High Density Value Highlighted (d) Landmark Detection Result for Initialization Scale



(a) (b)

Fig. 14. 1D and 2D Gradient Profiling (a) Gradient Profiles along the Normal of the Landmark (b) 2D Gradient Profiles with the Landmark as the Centre

for all landmarks in all bone structures(femur and pelvis) and all scales.

5.2 Testing

This section explains the segmentation of femur and pelvis given a new X-ray image using the training data. Since it is a multiscale system with the initial scale for approximation, this section is divided according to that. In the initial scale, there is no refinement step as explained in section 5.1. The overview of the testing step is as follows:

(1) Sample randomly all around the test image scaled and approximate the position of the bone structure using landmark detection for the initial scale. The result is the detected landmarks.
(2) Upsample the resulting landmarks from the previous scale so as to resize them to the next scale.
(3) Use the upsampled landmarks as the intialization for the next scale.
(4) Sample randomly around the subshapes of the upsampled landmarks and do the landmark detection.
(5) Scale the SSM mean shape according to the scale and fit this SSM through the landmarks detected.
(6) Update the fitted landmarks by snapping to the gradients using gradient profiling. The result is the refined landmarks.
(7) Repeat from (2) for all scales in the pipeline.

These steps are explained in detail in this section.

INITIALIZATION As explained in section 5.1, there is only landmark detection in the initial scale. The training data for landmark detection comprises of patch feature vectors, patch centres and displacement vectors. Given a new X-ray image, patches are sampled randomly all around the image. The only unknown here is the displacement vectors as the landmarks in the bone structure are unknown. Thus the aim of the landmark detection is to find the landmark position given the training data and the feature vectors and patch centres from the test X-ray image.

The test data have to be constructed into matrices the same way as explained in the section 5.1. Given a new X-ray image, the image is rescaled to the initial scale and K patches are sampled randomly all around the test image as shown in Fig 15a. Then the multilevel HoG vectors are computed for the patches. The transform matrix for the given scale and the given subshape from the training data is used to reduce the feature vector dimension from m to n. This feature vectors and patch centres are the known data for the test X-ray image and can be formulated into matrix $\mathbf{F}$ similar to $\tilde{\mathbf{F}}$ and $\mathbf{C}$ similar to $\tilde{\mathbf{C}}$ respectively where $\mathbf{F} \in R^{n\times K}$ and $\mathbf{C} \in R^{2\times K}$.

Since the training data $\tilde{\mathbf{F}}, \tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$ is saved per X-ray image, compound matrices which consists of all the X-ray images have to be created for the training data. Given that $\tilde{K}$ is the total number of training patches and the number of training images are known, equal number of random patches are selected from the training data of each image to construct the new compound matrix $\tilde{\mathbf{F}} \in R^{n\times \tilde{K}}, \tilde{\mathbf{C}} \in R^{2\times \tilde{K}}$ and $\tilde{\mathbf{D}} \in R^{2L\times \tilde{K}}$.

These matrices are used to predict the displacement matrix of the test patches, $\mathbf{D}$ using the equation (2) from Chen *et al.* work which incorporates the feature matching explained in section 3.3 and the derivation is explained in Chen *et al.* paper [1].

$$\mathbf{D} = -\mathbf{G}\mathbf{A}^{-1} \tag{2}$$

The $\mathbf{D}$ corresponds to the displacement from the test patch centres to the unknown landmarks and $\mathbf{D} \in R^{2L\times K}$ where L is the subshape length. For the initialization scale, subshape length is the number of landmarks in the femur and pelvis shape. The computation of $\mathbf{G}$ and $\mathbf{A}$ is explained in detail in the Appendix B.

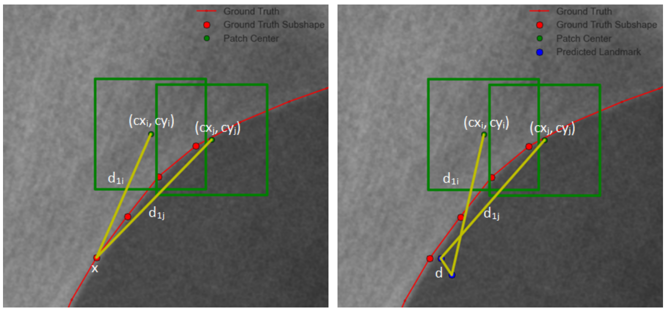
The **A** corresponds to the patch feature matching explained in section 3.3. For patch matching, a compound matrix with feature vectors of training and test patches concatenated together is obtained as follows:

$$\hat{\mathbf{F}} = [\tilde{\mathbf{F}}\ \mathbf{F}] = \begin{bmatrix} \tilde{f}_{11} & \cdots & \tilde{f}_{1\tilde{K}} & f_{11} & \cdots & f_{1K} \\ \tilde{f}_{21} & \cdots & \tilde{f}_{2\tilde{K}} & f_{21} & \cdots & f_{2K} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{f}_{n1} & \cdots & \tilde{f}_{n\tilde{K}} & f_{n1} & \cdots & f_{nK} \end{bmatrix}$$

where $\hat{\mathbf{F}} \in R^{n\times(\tilde{K}+K)}$, n is the reduced feature dimension, $\tilde{K}$ is the total number of training data patches and K is the number of test patches.

Matrix **A** have two main variables, **M** and α that relates to the matching. The $\hat{\mathbf{F}}$ matrix is used for finding a match where one column corresponds to a patch feature. A match for a single patch can be a train or test patch. The matching is done by first computing the affinity matrix **S** which is obtained by calculating s nearest neighbor for each patch. Therefore, for each patch, s nearest neighbor in feature proximity is found based on the minimum L^2 norm between the corresponding feature vectors. For example, for finding a match for the i^{th} patch, the L^2 distance between the i^{th} column and all other columns in $\hat{\mathbf{F}}$ is computed and the s least distances in these is the best matches for the i^{th} patch. $S_{i,j}$ is set to 1 for the i^{th} patch if the j^{th} is one of its s nearest neighbor for **S** $\in$R$^{(\tilde{K}+K)\times(\tilde{K}+K)}$ where $\tilde{K}$ is the total number of training patches and K is the number of test patches. The **M** matrix is computed by first estimating the Laplacian of the **S** matrix. Then the resultant matrix is normalized by dividing it by its trace to obtain **M** $\in$R$^{(\tilde{K}+K)\times(\tilde{K}+K)}$. **M** is the feature similarity measure and α corresponds to the weight of this measure and is set arbitrarily between 0 and 1.

The **G** matrix enforces the geometric constraint explained in section 3.3. According to the geometric constraint illustrated in the Fig 16, for two patches i and j, the assumption is that both the patches votes the same landmark position, **x** as shown in Fig 16a. But reality is similar to Fig 16b, when both the patch votes for the same landmark position **x**, the predictions are different. So, the constraint is to minimize the distance d so that the predictions are similar as possible. The $\mathbf{d_{1i}}$ in Fig 16 corresponds to the dx_{1i} and dy_{1i} and the geometric constraint can be restructured from $dx_{1i} - dx_{1j} = dx_{2i} - dx_{2j} = \ldots = dx_{Li} - dx_{Lj} = cx_j - cx_i$ as dx = $(cx_i + dx_{ki})$ - $(cx_j + dx_{kj})$, where $k \in$ [1,L] and L is the number of landmarks in the subshape. This is the same for the displacement and centre in the y direction as well. The aim is to minimize the dx and dy.



(a) (b)

Fig. 16. Geometric Constraint for Feature Matching (a) Assumption: Both patches predict the same position for landmark **x** (b) Reality: Both patch predicts different positions and the aim is to reduce d, the difference between the predicted positions

Matrix **G** have two main variables, $\bar{\mathbf{C}}$ and β that relates to the geometric constraint. Ideally all combination of patches that votes for the same landmark should be considered but due to efficiency reasons, only neighboring pair of patches are considered, i.e, (i,j) are the patch pairs, they have values such as (1,2), (2,3), ..., (K-1,K) where K is the number of test patches. Thus $\bar{\mathbf{C}}$ is the matrix obtained by vertically replicating L times the difference between the i^{th} and j^{th} patch centre difference. This can be written in matrix form as follows:

$$\bar{\mathbf{C}} = [\mathbf{c_2} - \mathbf{c_1}, \mathbf{c_3} - \mathbf{c_2}, \ldots, \mathbf{c_{K-1}} - \mathbf{c_K}]$$

$$\text{where } \mathbf{c_j} - \mathbf{c_i} = \begin{bmatrix} cx_j^1 - cx_i^1 \\ cy_j^1 - cy_i^1 \\ \vdots \\ cx_j^L - cx_i^L \\ cy_j^L - cy_i^L \end{bmatrix} \text{ and } \bar{\mathbf{C}} \in \mathrm{R}^{2L\times(K-1)}$$

Thus the $\bar{\mathbf{C}}$ matrix is obtained which is used to estimate the **G** matrix and the β is the weight for the geometric constraint and is set arbitrarily between 0 and 1 as well. However the matching weight,α is always set higher than the weight for the geometric constraint, $> \beta$.

The complete formulation of the equation(2) is given in Appendix B and is used to obtain the predicted displacement **D**. After the **D** matrix is computed, each test patch votes for the landmark position where each column of **D** corresponds to a test patch. The votes for the landmark are computed by using the formula **C+D** which corresponds to the patch centre of the test patches vertically replicated L times added to the predicted displacements. Each patch gives a landmark position. Thus for each landmark, there are K predicted positions as shown in Fig 15b. To estimate the true position of the landmark from this, kernel density estimation(KDE) of the positions is computed for each landmark giving a response image. The response images of all the landmarks are shown in Fig 15c. The position with the highest density in the response image is the estimated landmark. The high density point of all the response images forms the segmented shape as shown in Fig 15d. This is the result of the initialization scale.

OTHER SCALES The initialization scale only provides an approximate position of the bone structure in the X-ray image. For a more accurate landmark detection, the test patches have to be sampled nearer to the bone and hence the system is multiscale where the accuracy increases for different scales. After the initialization scale, the landmark detection is done by sampling in a smaller region, i.e. around the subshapes with L landmarks and then the detected landmarks are refined using ASM and gradient profiling. The pipeline is same for all scales other than the initial scale and is explained below. The sequential results for each scale shows increasing accuracy as shown in Fig 20.

Landmark Detection

The concept of landmark detection for the new scale works similar to that of the initialization scale. The only difference is that instead of randomly sampling all around the test X-ray image, the patches are sampled around subshapes similar to the sampling done in the training data as shown in Fig 10b. In the initialization scale, the subshape length was the number of landmarks in the bone shape whereas here it is L. The subshape is obtained from the landmarks detected from the previous scale. The overview of the steps for landmark detection for scales other than the initial scale is as follows:

(1) Upsample the resulting landmarks from the previous scale so as to resize them to the next scale by multiplying the landmarks by $(newscalefactor/previousscalefactor)$.
(2) Extract a subshape containing L landmarks from the upsampled shape.
(3) For a given subshape, randomly sample patches around the subshape and compute the multilevel HoG feature vectors for these patches.
(4) Compute the reduced feature vectors using the transform matrix for the given scale and the given subshape.
(5) Estimate the test displacement matrix, **D** and the updated landmarks in the subshape for the test image as explained in the Initialization section before using the training data.
(6) Repeat from (3) for all subshapes and estimates the landmarks in the shape.

The result are the detected landmarks for that scale.

Active Shape Models

An initial shape is obtained from the landmark detection and this shape is regularized using the ASM. Given a new shape, **Y** for testing and the SSM model of the bone which is obtained from the training part, the aim of ASM fitting is to find the model points, **x** that best fit the detected landmarks, **Y** to the SSM mean shape, $\bar{\mathbf{x}}$ based on correspondences where **x** , **Y** and $\bar{\mathbf{x}} \in R^{2\times L}$. The steps for ASM fitting are as follows:

(1) Rescale the SSM mean shape to the corresponding scale.
(2) Initialize the shape parameter, **b** to 0, implies model points = mean shape, i.e. $\mathbf{x} = \bar{\mathbf{x}}$
(3) Generate the model points positions using $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$, where **P** is the SSM principal components.
(4) Find the pose parameters transform which best align the model points **x** to the new set of landmarks **Y** using Procrustes Analysis [7].
(5) Project **Y** into the model co-ordinate frame, **Y'** by using the inverse transform from (4)
(6) Update the shape parameters, **b** to match to **Y'** by finding least squared solution of **Ax'** = **B**, where **A** is **P**, **x'** is **b** and **B** = **Y'** - $\bar{\mathbf{x}}$
(7) Repeat from (3) until convergence

This provides the new fitted shape for that scale as shown in Fig 17. The concept behind ASM is briefly explained in section 3.1. For more detailed explanation, see Cootes *et al.* paper [5].

Gradient Profiling

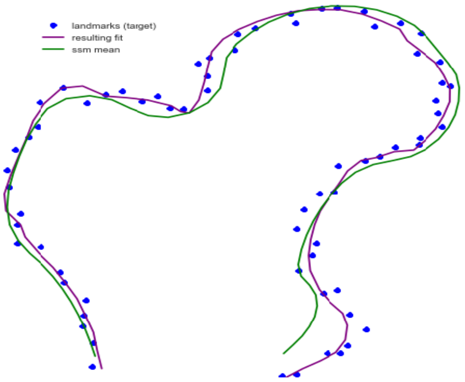


Fig. 17. Active Shape Model Fitting Result

After the shape has been regularized by the ASM, the gradient profiling is used to snap the points to the gradients. This is done by taking a patch from the gradient image with the patch centre as the landmark similar to the gradient profiling training and using a windowing approach to do template matching. The template matching is based on fast normalized cross-correlation [9]. A match for the training patch for that landmark is found by windowing through the test patch and computing a similarity measure. The highest similarity measure provides the best match for that training patch. The centre of the sub-patch in the test patch that provides the highest similarity is the new updated landmark position voted by that training patch.

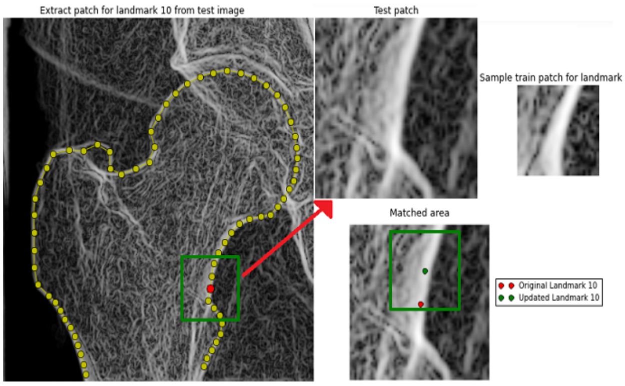

Extract patch for landmark 10 from test image
Test patch
Sample train patch for landmark 10
Matched area
Original Landmark 10
Updated Landmark 10

Fig. 18. Gradient Profiling Testing Pipeline for a Given Landmark with a Sample Train Patch

The overview of extracting the gradient patch for a given landmark and the result of template matching of the test patch with a sample train patch for the same landmark is shown in Fig 18. The matching is done for all the training patches available for that landmark and a similarity measure is computed for each training patch. Thus each training patch votes for a position and a weight is computed as well. A weighted sum of these positions is used to predict the new landmark position. The whole process is repeated for all the landmarks for that scale. The test patch size differs for different scale. The updated landmarks are the final result for the given scale and this is the initialization for the next scale if there is any. The whole pipeline of training and testing are done for femur and pelvis independently.

5.3 Joint Space Width Metric

After the femur and pelvis has been segmented and refined through all scales, the JSW metric is extracted by considering the necessary subshapes from femur which contains the joint space between femur and acetabular cup. Then, for each landmark in the subshape, the closest point in the pelvis contour that has been segmented is found and the distance to that point is considered as the joint distance at that landmark. Thus, depending on the number of landmarks in the subshapes, l distances are found as shown in Fig 25. The individual distances of the landmarks to the contour, the mean distance and the median distance has been evaluated for the JSW metric.

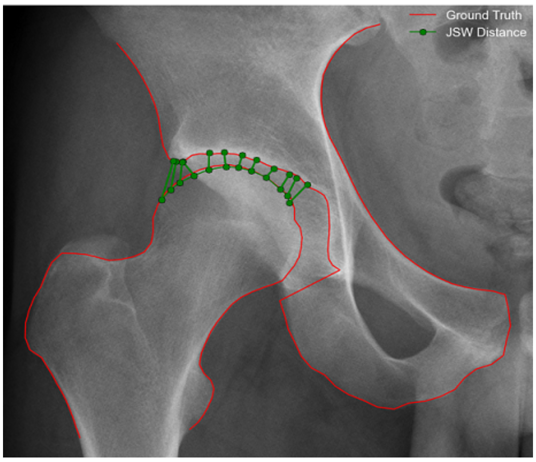


Fig. 19. Joint Space Width Distances

6. IMPLEMENTATION

The parameters needed for the implementation of the method is explained in this section. The method proposed is multiscale where the number of scales, N considered in this method is 6. The scales used in this paper are 2%, 5%, 12%, 25%, 50% and 100% of the X-ray resolution. The number of landmarks, $\tilde{N}$ in femur shape is 69 and in pelvis shape is 95.

Some of the important parameters chosen for each scale in the training and testing part of landmark detection is shown in table 1. The subshape length, L used in this paper is 4. Since L=4 and $\tilde{N}$=69 for femur, there are 17 subshapes and for pelvis $\tilde{N}$=95, hence number of subshapes is 24.

Table 1. Landmark Detection Parameters

Scale	Patch Size	Radius	Number of Subshapes	Subshapes

Initial scale	40	Whole Image	1	Femur:(0,69)
				Pelvis:(0,95)

Other scales	40	30	Femur:17 Pelvis:24	Femur:(0,4),(4,8),.. Pelvis:(0, 4),(4,8),..

A major part of landmark detection is the multilevel HoG vectors which have various parameters as explained in the section 3.2. Since it is multilevel, we use 2 levels with cell size (2,2) and (4,4) with 18 orientations which is applied on a 40×40 patch giving 72 and 288 dimensional feature vectors respectively as explained in section 5.1. Thus, the HoG dimension, m of a single patch is 360. To reduce this high dimensional vector, the feature selection method explained in the section 3.4 is used to compute a transform matrix which reduce the dimension from 360 to 100(n).

Some of the other parameters in the landmark detection is the number of train patches per image, k which is 200 in this paper, the total number of training patches during testing, $\tilde{K}$ is 2000 and the number of test patches, K is 500. For feature matching, s nearest neighbors are used. In this paper, s is 5. Other parameters for feature matching is α and β which is set to 0.05 and 0.005 respectively as default. The values are arbitrary.

The training and testing patch size for the 2D gradient profiling step for all scales except the initial scale are given in the table 2. The initial scale is excluded as there is no refinement step at this scale. The standard deviation of the Gaussian filter, σ for finding the gradients for each scale is also given in the table 2.

Table 2. Gradient Profiling Parameters

Parameter	Scale 0.05	Scale 0.12	Scale 0.25	Scale 0.5	Scale 1.0
Training Patch Size	10	20	30	40	60
Testing Patch Size	20	40	60	80	120
$\sigma_x = \sigma_y = \sigma$	0.5	0.5	1	2	4

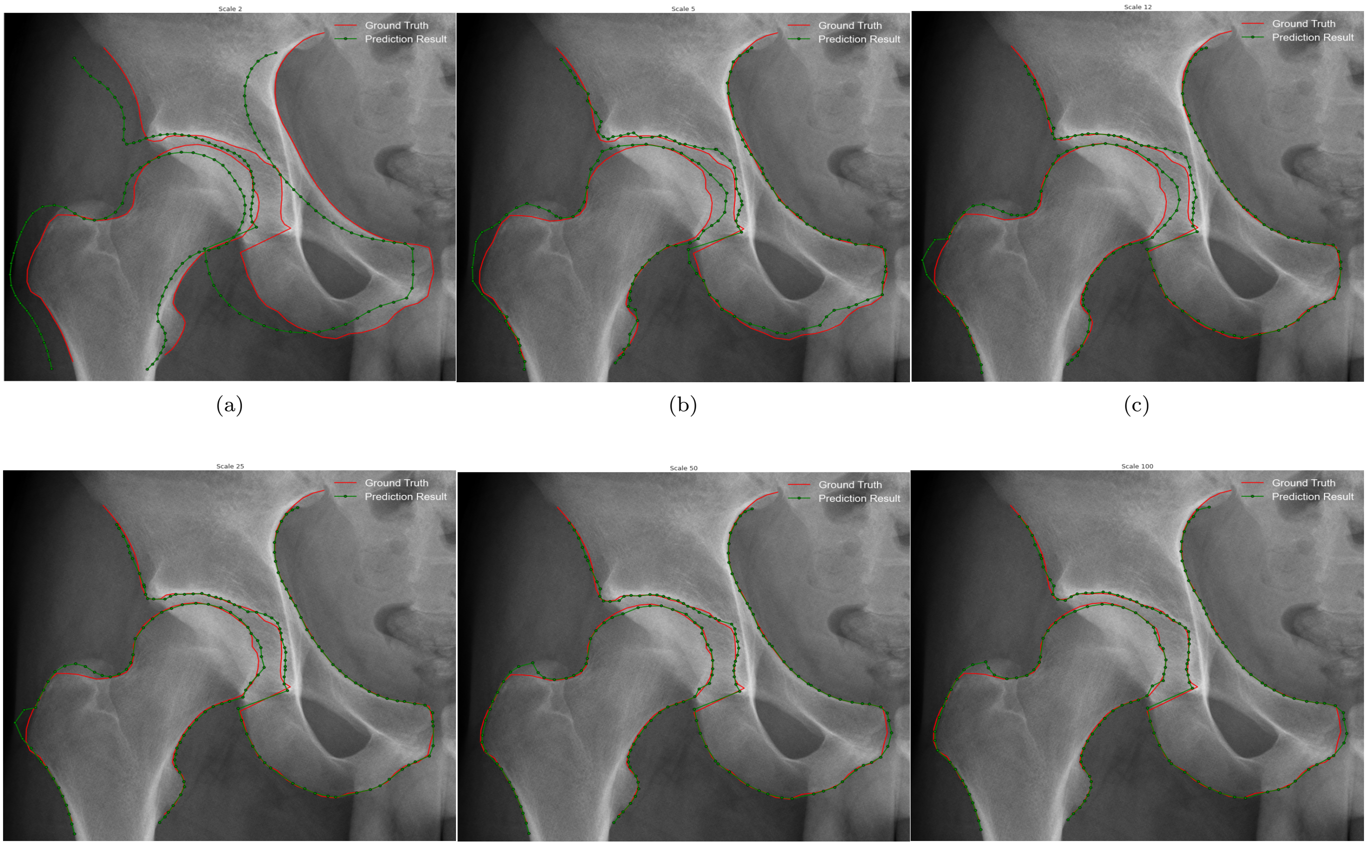
For the JSW metric, 3 subshapes from femur are considered which contains 12 landmarks in total. Thus, l is 12 as there are 12 individual distances for the JSW metric.

7. RESULTS AND DISCUSSION

The datasets consists of 114 femur and 72 pelvis with manual annotations. These are used as the ground truth for the evaluation throughout this work. To improve the evaluation generalization of our approach, we use leave one out testing strategy. Using leave one out strategy, the entire dataset except the test X-ray is used for the training data. For estimating the joint space width(JSW), the ground truth JSW for the 72 femur and pelvis set are used. The image spacing for all X-rays is fixed at 0.1mm/pixel and this is used for computing the JSW metric.

The complete pipeline result for a single scan has been shown in Fig. 20. It can be seen that the results improves in each scale. There is a big offset in the first scale but since this is an intialization scale, this is to be expected. From the final result, Fig. 20f, it can be seen that there are some error according to the ground truth but the algorithm actually follows the gradients and is accurate. More importantly, the landmarks near the joint space show negligible error.

In this section, the time taken for training and testing has been evaluated. The error rate for the segmentation of femur and pelvis and the JSW metric has been assessed as well. Some of the main parameters in the landmark detection has been studied and evaluated for their performance.



(a) (b) (c)

(d) (e) (f)

Fig. 20. Complete Pipeline Results (a) Scale 2% Results (b) Scale 5% Results (c) Scale 12% Results (d) Scale 25% Results (e) Scale 50% Results (f) Scale 100% Results

7.1 Computation Time

The testing and training is performed on a system with a Intel Core i5 2.80 Ghz quad-core processor, 8GB internal memory and NVIDIA GeForce GTS 250 graphics card with 256MB GDDR3 graphics memory and 128 CUDA cores. The performance is tested for the main component of the pipeline - the training and test part.

The time taken for creating the training data depends on number of training patches per subshape per scale. In our paper, we use 200 training patches for each subshape for each X-ray image and for each scale. The training also includes the speed for creating the training data for gradient profiling training. The SSM training data time is negligible compared to the others, therefore it is not considered. The testing speed is evaluated per scale for a X-ray image. The number of test patches per image for each scale and per subshape is 500. From the second scale, the gradient profiling and ASM fitting time is included in this time as well. The computation time for both training and testing of femur is given in table 3 and for pelvis in table 4. The training time is for the whole dataset and the testing time is the average time for each X-ray image. Femur has 17 subshapes and pelvis has 24 subshapes except for the first scale 0.02.

The time taken to create the complete training data is approximately 3 hours 30 minutes. Since the training data is generic, adding a new X-ray does not require re-creating the entire training data once again. The addition of new data improves the accuracy of the system and without too much overhead as its generic. The average computation time taken for processing a new scan, i.e. femur and pelvis is approximately 21 minutes for one side(left/right).

Table 3. Computation Time in Minutes: Femur

Step	Scale 0.02	Scale 0.05	Scale 0.12	Scale 0.25	Scale 0.5	Scale 1.0
Training	0.85	15	15.2	17.6	25	57.3
Testing	0.09	1.5	1.6	1.7	2	2.8

Table 4. Computation Time in Minutes: Pelvis

Step	Scale 0.02	Scale 0.05	Scale 0.12	Scale 0.25	Scale 0.5	Scale 1.0
Training	0.45	10.8	11.2	12.6	16.5	35.2
Testing	0.08	1.9	2	2.1	2.2	2.7

7.2 Error Rate

The main error metric used in this paper is point-to-curve distance. The curve is the ground truth which is the manually segmented landmarks of the bone and the point is the detected landmark in the case of segmentation. For the JSW metric, the absolute error between the 12 distances in the ground truth and distances obtained from the corresponding 12 detected landmarks is considered as the error metric. Both the segmentation error and JSW error is evaluated on the 100% scale. The error is calculated according to the spacing 0.1mm/pixel.

FEMUR SEGMENTATION The femur segmentation error rate is shown in Fig 21 which is obtained by considering the error distribution for each subshape for 114 femurs assessed with leave-one out validation . There are 17 subshapes for femur as shown in the Fig 21. Thus, for each subshape, the closest distance between the detected landmarks to the ground truth contour is computed as the error. From the Fig 21, it can be seen that the mean error is 0.946mm. However, the area around the femoral head which is mainly interesting here (subshape 8-10) have segmentation error approximately in the range of 0.2mm which is relatively negligible.

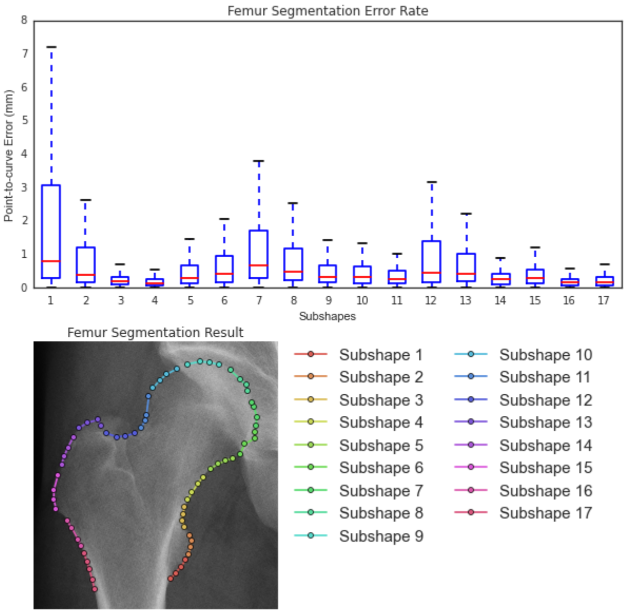
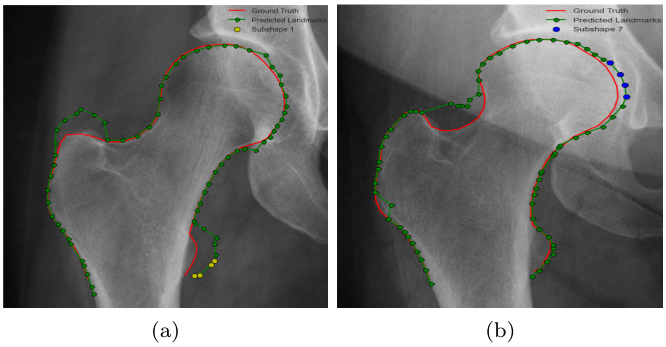


Fig. 21. Femur Segmentation Error Rate Per Subshape

From the Fig 21, it can be seen that highest error are for subshape 1 and 7 for the femur. The error at subshape 1 as shown in Fig 22a is mainly because in some of the X-ray images, the trocanter minor is invisible due to patient pose variation. There is not enough similar X-rays where trocanter minor is invisible in the training data, hence this couldn't be generalized enough to be detected. As for subshape 7 error, this is mainly due to complex shape of femur and as shown in Fig 22b, the femoral head in this X-ray image is a bit less rounder than the other femurs. However in both the cases in Fig 22, the area of the femur near the acetabular cup have negligible segmentation error.

PELVIS SEGMENTATION The pelvis segmentation error is computed similar to the femur error by considering the distance between the detected landmark points to the ground truth curve. The error is assessed per subshape as well. There are 24 subshapes for pelvis and the error rate for all the subshapes are shown in Fig 23. The mean segmentation error is approximately 1.09mm. The error around the subshapes 2 to 5 are negligibly less which is the acetabular cup area where the JSW is computed. We can intuitively assume that the accuracy can be further increased for pelvis by having more training data.



(a) (b)

Fig. 22. Incorrect Femur Segmentation Results (a) Subshape 1 Error (b) Subshape 7 Error

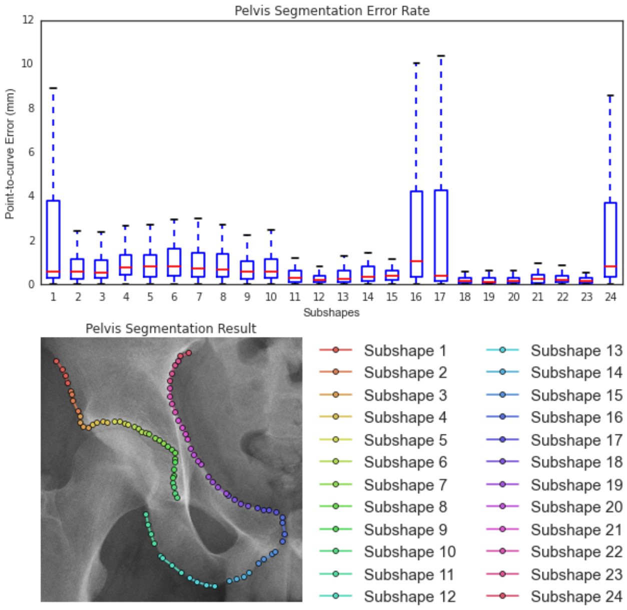
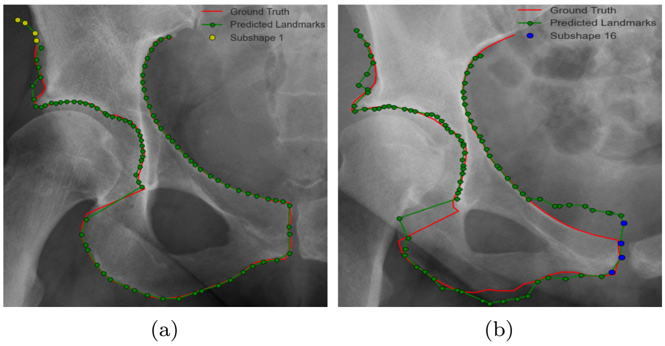


Fig. 23. Pelvis Segmentation Error Rate Per Subshape

The highest error in the pelvis segmentation is around the subshape 1, 16 and 17 according to Fig 23. Some results which shows these error are shown in Fig 24 and the error is mainly due to shape variation of pelvis. Some pelvises are wider than others. In Fig 24b, the pelvis is not as wide as the usual pelvis and hence shows an around the subshape 16. The training data for pelvis is not rich enough to detect all the minute changes. However the area of the pelvis near the acetabular cup has negligible segmentation error which is the joint space area.

JOINT SPACE WIDTH The error rate of the femur and pelvis segmentation near the acetabular cup is negligible. This is mainly because the landmarks are distributed in such a way that there are more landmarks around the area we are interested in, in this case the acetabular cup, so that accuracy near that area is more. The joint space is the area near the femur and acetabular cup and the joint space width(JSW) can be measured in many ways.

One of the metric for joint space is the landmark distances computed as explained in section 5.3. The distance is computed as the distance from the landmarks in femur to



(a)

(b)

Fig. 24. Incorrect Pelvis Segmentation Results (a) Subshape 1 Error (b) Subshape 16 Error

the pelvis contour. The femur landmarks are considered as the femur segmentation error rate for the subshapes near the acetabular cup area is less than the error rate of the pelvis subshape around that area. The femur subshapes considered for JSW are subshape 8,9 and 10 and each of them containing 4 landmarks and hence there are 12 distances. The ground truth for this JSW metric is computed from the manually annotated femur and pelvis the same way as shown in Fig 25. The result of the landmark distance as the JSW metric is shown in Fig 25.

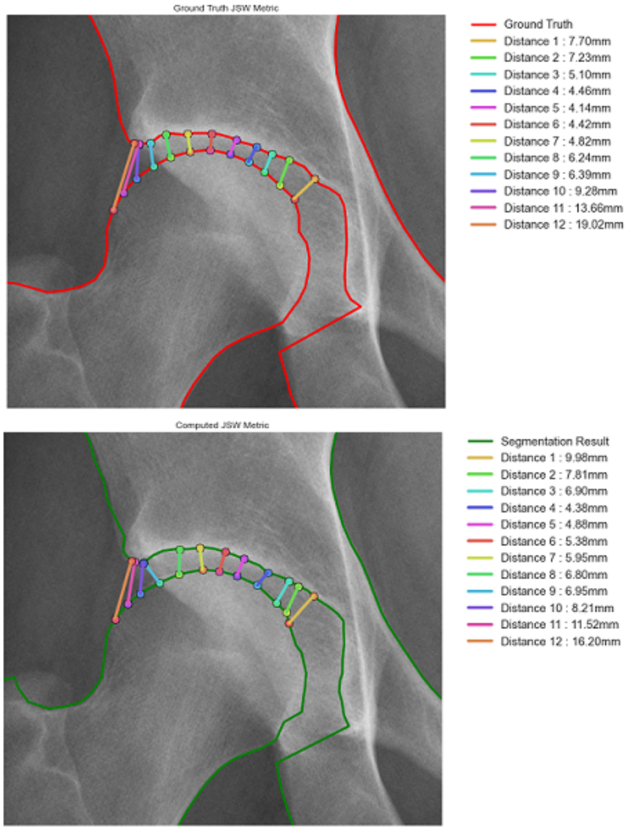


Fig. 25. Landmark Distance as JSW Metric - Ground Truth JSW Metric and JSW Metric from the Segmenation Result

The error rate for these 12 distances is calculated by taking the absolute difference of the landmark distance obtained from the segmentation result and the landmark distance from the ground truth landmarks. An example of the landmark distance is shown in Fig 25 and the absolute difference between each of the distance is the error. The error rate for the 12 distances is given in Fig 26.

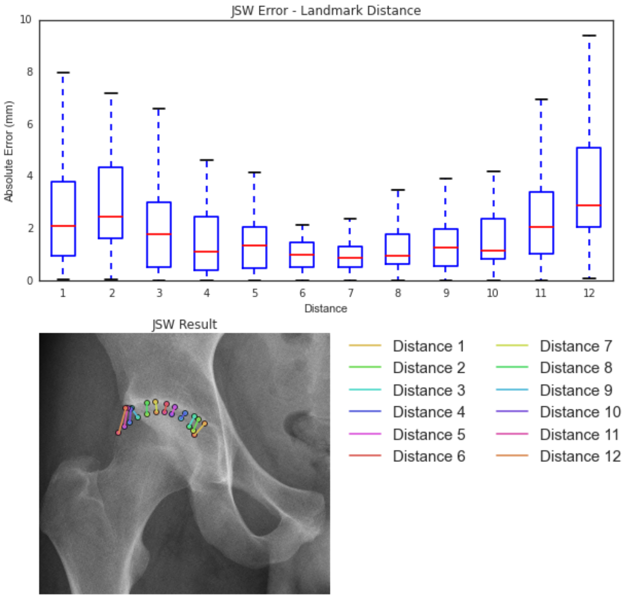


Fig. 26. JSW Error Rate - Landmark Distances as Metric

From the Fig 26, it can be seen that the distance 1 and distance 12 have the highest error. However, these landmarks do not define the joint space. They were considered as they were a part of the subshape that contains the landmarks in the joint space. The mean error of these distances is approximately 1.1mm which is reduced if the distances 1, 2, 11 and 12 are not considered as they are not part of the joint space.

Some of the other metrics that can be considered as the JSW metric are the mean of these 12 distances, their median, minimum and maximum. The maximum of these 12 landmark distances does not provide much information about the joint space as they are tainted by the 4 outlier distances. However, the mean and minimum of these distances provides an interesting insight into the joint space. As the aim of the JSW is to check if there is joint space narrowing, if the minimum distance JSW metric is found for both left and right hip, it can provide insight into which side is affected depending on how big the difference is between the metric of both sides. This is because osteoarthritis(OA) is usually only on one side of the body. So, if one side has a relatively smaller JSW metric than the other, this could indicate joint space narrowing based on the assumption that healthy hip have approximately symmetrical joint space. The error rate for these JSW metric are shown in Fig 27.

The mean error for these JSW metrics is 1.1mm for mean measure, 0.8mm for median, 1.2mm for minimum and 3.3mm for maximum. Median metric has the smallest error as this defines the actual joint space which is the middle distances of the 12 and is more accurate as the outlier distances are not that relevant for this metric.

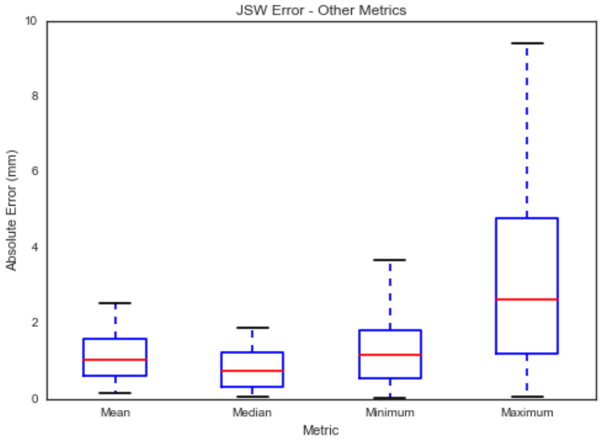


Fig. 27. JSW Error Rate - Other Metrics

7.3 Parameter Evaluation

There is lot of parameters that needs to be tuned for the complete pipeline. Most of the parameters were set based on qualitative performance. This is because there are too many combinations to be considered between the parameters and as creating the training data is time consuming, it was not plausible to estimate the best combination of parameters. Some of the major parameters for the landmark detection is evaluated and discussed to inspect the parameter sensitivity.

NUMBER OF TRAINING PATCHES The total number of training patches considered for the testing affects the system mainly for the computation time. The assumption was that more the training data, more accurate the system would be. Therefore, this parameter was evaluated to observe if the system is sensitive to the increase in the number of images. Fig 28 shows the result of the leave one out validation with varying number of patches. The minimum number of patches possible is 113 for femur as each image contributes to the training data and there are 114 manually annotated femurs.

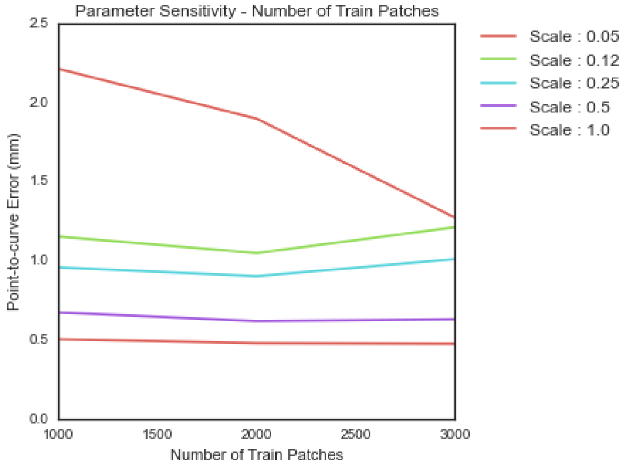


Fig. 28. Parameter Sensitivity - Number of Training Patches

It can be seen that the segmentation accuracy increases as the scale increases. However regarding the sensitivity to the number of patches, no solid conclusion could be drawn. This is because even though the point-to-curve error decreases as the number of patches increases, it increases when the number of patches is 3000 for some scales. However, for the scale 1.0, the error remains constant irrespective of the number of patches. As the JSW is extracted in this scale, this is the relevant scale. Therefore, the number of training patches is set to 2000 in this work as it shows good performance in all scales.

NUMBER OF SUBSHAPES Subshape length plays a major role in the accuracy as well as the computation speed of the algorithm. The detection of a landmark in the subshape is affected by the other landmarks in the same subshape. Therefore, landmark detection accuracy decreases if the number of landmarks in the subshape increases. This is because a single patch votes for all the landmarks in the subshape which is erroneous as the shape is too complex to be generalized by a single patch. Therefore, if the number of landmarks in the subshape decreases, the accuracy increases. But smaller subshapes means more processing time, therefore, a balance have to be made between both and a subshape with 4 landmarks was considered ideal in our case.

PATCH SIZE & SAMPLING RADIUS Another main parameter for the landmark detection is the patch size and the sampling radius. The patch size affects the computation time and the accuracy of the segmentation and the sampling radius affects the accuracy. The sensitivity of each of these parameters for the landmark detection is evaluated here. The refinement step was not included in the result while accessing this parameter.

The error rate for different patch size for all scales is shown in Fig 29. It can be seen that the error decreases when the patch size increases. However, if the patch size increases, the computation time increases as it takes longer to compute the HoG of the patch. And the declining rate of the error from using patch size 40 to 60 is lesser than from 30 to 40. Hence, the patch size used in this work is 40 × 40 pixels as the computation time is acceptable for this patch size.

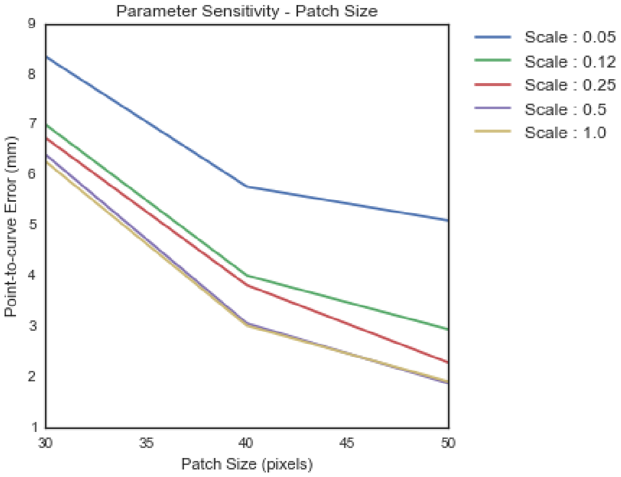


Fig. 29. Parameter Sensitivity - Patch Size

From Fig 30, it can be seen that the error is directly proportional to the sampling radius. So, the accuracy increases when the sampling radius is low. However, if the sampling radius is too low, it gives an incorrect landmark prediction especially if the initialization result has a huge error. This is because the system can only correct the landmark upto the sampling radius. So, if the prediction from the initialization scale has an error that is more than the sampling radius, then the landmark detection fails especially without the refinement step. Therefore a sampling radius of 30pixels is used in this work.

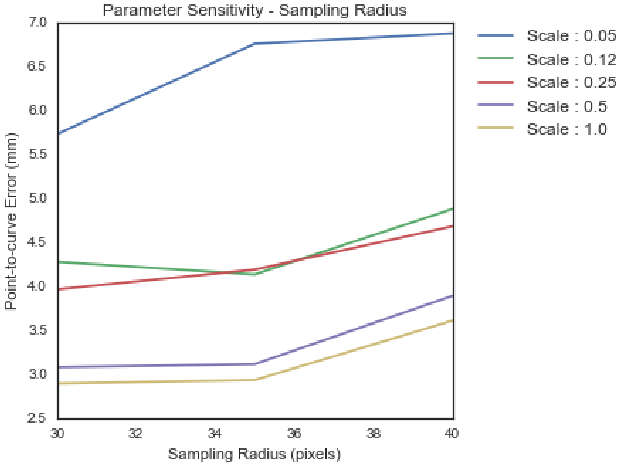


Fig. 30. Parameter Sensitivity - Random Sampling

These are the parameters that have been evaluated to investigate their sensitivity to the landmark detection.

8. CONCLUSION AND FUTURE WORK

In this work, we have presented a multiscale supervised learning method for measuring the joint space in X-ray images. This technique uses a data driven strategy proposed by Chen *et al.* to estimate the image displacements for landmark detection [1]. The multiscale approach increases the accuracy of detection. The segmentation accuracy was further enhanced using the unique refinement step introduced in this work which includes a classic ASM and 2D gradient profiling.

A quantitative performance evaluation showed that this technique provides an accurate segmentation of both femur and pelvis with a mean error of 0.946mm and 1.09mm respectively. This is superior to the performance exhibited by the state-of-the-art methods since the method works for multiple bone structures, X-rays with varying intensities and different patient posture.

The purpose of segmentation was to extract the joint space width(JSW) automatically. Only the area around the joint space is relevant where the segmentation error is less than 0.5mm. This segmentation error is viable to define a JSW metric for diagnostic purposes. The JSW metric introduced and discussed in the work shows a mean error of less than 1.2mm which is acceptable.

The highest computational bottleneck while implementing this technique was creating the training data. Even though some optimization techniques was implemented for training data in this work like feature vector dimensionality reduction, there are numerous possibilities to speed up this process for example, by using parallel processing. Further research is needed to identify the processes that could be made parallel efficiently.

Further investigation is also needed to find the best combination of parameters for the system to improve the accuracy. One solution is to use hyperparameter optimization to choose the parameters which will also ensure that the model does not overfit its data by tuning.

Likewise, other superior metrics could be explored for JSW metric to make it anatomically more precise. If there are X-ray images that have been classified into healthy hip and hip with osteoarthritis(OA), then their JSW metric can be analyzed to detect if there is any pattern emerging. It can be used to see if there is a common joint space width for healthy hip and OA hip. Another possibility that could be interesting is to classify the X-rays images based on presence or absence of joint space narrowing. This can be used to find an approximate minimum JSW to indicate joint space narrowing. These values could be insightful while building a full automated system for detecting osteoarthritis.

Considering the scope for future work, we can be sure that this work have significant role in the world of clinical research pertaining to measurement of joint space narrowing in Hip X-ray images. This is an initial step towards a developing a technique capable of automatic diagnosis of osteoarthritis.

REFERENCES

[1] C. Chen, W. Xie, J. Franke, P. Grutzner, L.-P. Nolte, and G. Zheng. 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] C. Chen, Y. Yang, F. Nie, and J.-M. Odobez. 3d human pose recovery from image by efficient visual feature selection. *Computer Vision and Image Understanding*, 115(3):290–299, 2011.

[3] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, pages 2724–2729. IEEE, 1991.

[4] T. Cootes, E. Baldock, and J. Graham. An introduction to active shape models. *Image processing and analysis*, pages 223–248, 2000.

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[7] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 285–339, 1991.

[8] B. Jian and B. C. Vemuri. A robust algorithm for point set registration using mixture of gaussians. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE*

International Conference on, volume 2, pages 1246–1251. IEEE, 2005.

[9] J. Lewis. Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123, 1995.

[10] C. Lindner, S. Thiagarajah, J. Wilkinson, T. Consortium, G. Wallis, and T. F. Cootes. Fully automatic segmentation of the proximal femur using random forest regression voting. *Medical Imaging, IEEE Transactions on*, 32(8):1462–1472, 2013.

[11] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation 1. *Annual review of biomedical engineering*, 2(1):315–337, 2000.

[12] J. T. Sharp, J. Angwin, M. Boers, J. Duryea, G. von Ingersleben, J. R. Hall, J. A. Kauffman, R. Landewé, G. Langs, C. Lukas, et al. Computer based methods for measurement of joint space width: update of an ongoing omeract project. *The Journal of rheumatology*, 34(4):874–883, 2007.

[13] J. Sobotta. *Sobotta atlas of human anatomy I.* Williams & Wilkins, 1997.

[14] W. Xie, J. Franke, C. Chen, P. A. Grützner, S. Schumann, L.-P. Nolte, and G. Zheng. A complete-pelvis segmentation framework for image-free total hip arthroplasty (tha): methodology and clinical study. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2014.

Appendix A. LANDMARK DEFINITIONS

The tool has been constructed in Javascript by Clinical Graphics B.V, Delft for providing accurate registration of the landmarks for all the training data. With the tool, an image can be uploaded along with the contours and desired number of landmarks can be placed manually on the contour. The landmark can also be placed near the desired place and it will be snapped to the nearest position in the contour. For femur, 8 landmarks were placed manually whose description is given in table A.1 and the corresponding landmarks are shown in Fig A.1.

Table A.1. Femur Landmark Definitions

Index	Landmark Definition
(1)	Inferior end point of lesser trochanter
(2)	Superior end point of lesser trochanter
(3)	Inferior edge of femoral head
(4)	Inferior edge of fovea capitis
(5)	Superior edge of femoral head
(6)	Proximal point of greater trochanter
(7)	Lowest tipping point of lateral edge
(8)	Lateral point opposite the inferior point of the lesser trochanter

Note: For a more detailed explanation of the terminology of the anatomy, the *Sobotta's* book on "Atlas of Human Anatomy" can be referred [13].

For pelvis, there are 7 landmarks manually placed which are defined in table A.2 and shown in Fig A.2.

The tool distributes n number of landmarks between these manually placed landmark. The n is the number of landmarks that have to be distributed between 2 neighboring manually placed landmarks. The n can be different for each pair of landmarks. The fully landmark dataset is shown in Fig A.1 for femur and Fig A.2 for pelvis.

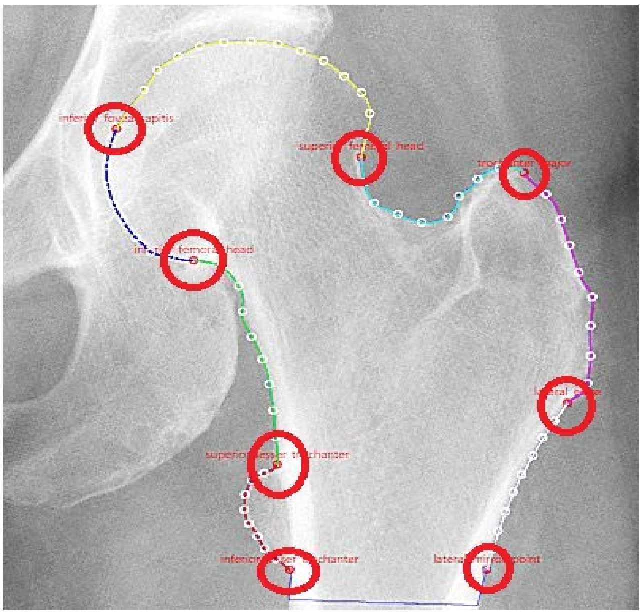


Fig. A.1. Femur Landmarks

Table A.2. Pelvis Landmark Definitions

Index	Landmark Definition
(1)	Top medial edge near the spine
(2)	Top lateral edge - mirror of medial edge
(3)	Labrum near the acetabular cup/Superior acetabulum
(4)	Bottom of the tear
(5)	Inferior acetabulum
(6)	Inferior point of pubic
(7)	Superior point of pubic

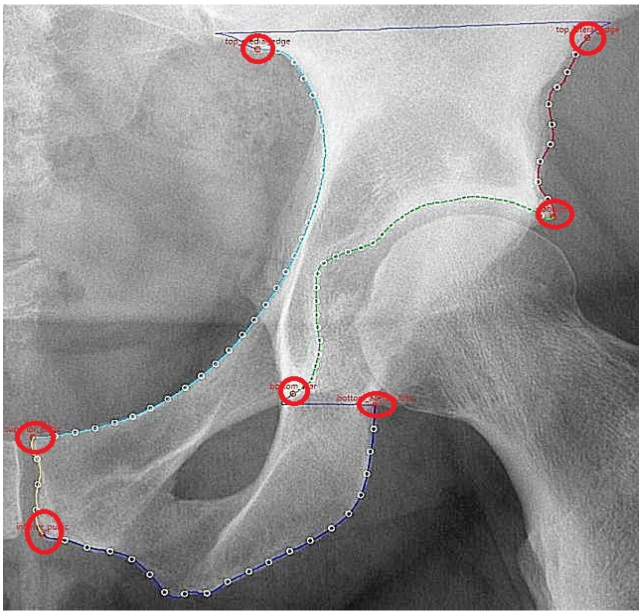


Fig. A.2. Pelvis Landmarks

The entire work was constructed to be used as a Python library and is completely done in Python. The database is stored in the HDF5 format and uses pytables to interact with it. Although python have lot of image processing libraries like Scipy and Sklearn for nearest neighbor algorithms, template matching and gradient operations, all the basis functions like the registration algorithms, preprocessing, landmark detection and refinement algorithms had to be implemented from scratch. The Ipython Notebook is used as the interactive environment and the Matplotlib library is used for visualization.

Appendix B. LANDMARK DETECTION EQUATION

The training data is constructed as explained in section 5.1 and the compound matrices $\tilde{\mathbf{D}}, \tilde{\mathbf{F}}$ and $\tilde{\mathbf{C}}$. For the test image, patch centre matrix $\mathbf{C}$ and the feature vector matrix $\mathbf{F}$ is extracted according to the scale similar to training data. Assuming K patches are sampled randomly from the test image, the dimension of $\mathbf{C}$ is (2, K) and $\mathbf{F}$ is (k, K). In our case, K is 500. Compound matrices which are needed later containing the training and test data jointly can be constructed as follows:

$$\hat{\mathbf{D}} = [\tilde{\mathbf{D}}\ \mathbf{D}] = \begin{bmatrix} \tilde{d}x_{11} & \dots & \tilde{d}x_{1\tilde{K}} & dx_{11} & \dots & dx_{1K} \\ \tilde{d}y_{11} & \dots & \tilde{d}y_{1\tilde{K}} & dy_{11} & \dots & dy_{1K} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{d}x_{L1} & \dots & \tilde{d}x_{L\tilde{K}} & dx_{L1} & \dots & dx_{LK} \\ \tilde{d}y_{L1} & \dots & \tilde{d}y_{L\tilde{K}} & dy_{L1} & \dots & dy_{LK} \end{bmatrix}$$

$$\hat{\mathbf{F}} = [\tilde{\mathbf{F}}\ \mathbf{F}] = \begin{bmatrix} \tilde{f}_{11} & \dots & \tilde{f}_{1\tilde{K}} & f_{11} & \dots & f_{1K} \\ \tilde{f}_{21} & \dots & \tilde{f}_{2\tilde{K}} & f_{21} & \dots & f_{2K} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{f}_{k1} & \dots & \tilde{f}_{k\tilde{K}} & f_{k1} & \dots & f_{kK} \end{bmatrix}$$

where $\hat{\mathbf{D}} \in R^{2L\times(\tilde{K}+K)}$ and $\hat{\mathbf{F}} \in R^{k\times(\tilde{K}+K)}$

After the matrices $\tilde{\mathbf{D}}, \hat{\mathbf{F}}, \hat{\mathbf{C}}$ are constructed, the $\mathbf{D}$ matrix is found using the equation (B.1).

$$\begin{aligned} \mathbf{D} &= -\mathbf{G}\mathbf{A}^{-1} \qquad , where \\ \mathbf{G} &= -\frac{\tilde{\mathbf{D}}\mathbf{P}^T}{L\tilde{K}} - \frac{\beta\bar{\mathbf{C}}\mathbf{U}^T\mathbf{Q}^T}{LK} \\ \mathbf{A} &= \frac{1}{L\tilde{K}}\mathbf{P}\mathbf{P}^T + \frac{2\alpha}{L}\mathbf{M} + \frac{\beta}{LK}\mathbf{Q}\mathbf{U}\mathbf{U}^T\mathbf{Q}^T \end{aligned} \tag{B.1}$$

The $\mathbf{P}$ and $\mathbf{Q}$ matrices are defined as $\tilde{\mathbf{D}} = \hat{\mathbf{D}}\mathbf{P}$ and $\mathbf{D} = \hat{\mathbf{D}}\mathbf{Q}$. $\mathbf{P}$ and $\mathbf{Q}$ are made up of identity matrices $\mathbf{I}_{M\times N}$ and zeros matrices $\mathbf{0}_{M\times N}$ to extract the corresponding training and test displacement matrices respectively.

$$\begin{aligned} \mathbf{P} &= [\mathbf{I}_{\tilde{K}\times\tilde{K}}; \mathbf{0}_{K\times\tilde{K}}] \in R^{(\tilde{K}+K)\times\tilde{K}} \\ \mathbf{Q} &= [\mathbf{0}_{\tilde{K}\times K}; \mathbf{I}_{K\times K}] \in R^{(\tilde{K}+K)\times K} \end{aligned}$$

The $\bar{\mathbf{C}}$ and $\mathbf{U}$ corresponds to the geometric constraint in the matrix form. Ideally all combination of patches should be considered but due to efficiency reasons, only consecutive pair of patches are considered, i.e, (i,j) have values such as (1,2), (2,3), ..., (K-1,K). $\mathbf{U}$ is the matrix $[\mathbf{e_1} - \mathbf{e_2}, \mathbf{e_2} - \mathbf{e_3}, ..., \mathbf{e_{K-1}} - \mathbf{e_K}]$ which is obtained from

$$\mathbf{D}(\mathbf{e_i} - \mathbf{e_j}) = \begin{bmatrix} dx_{1i} - dx_{1j} \\ dy_{1i} - dy_{1j} \\ \vdots \\ dx_{Li} - dx_{Lj} \\ dy_{Li} - dy_{Lj} \end{bmatrix}$$

where $\mathbf{e_i}$ is a K dimensional column vector with the i^{th} elements as 1 and all the other elements 0s. Thus $\mathbf{U}$ is a matrix of 1, -1 and 0s and $\mathbf{U}$ $\in \mathrm{R}^{K\times(K-1)}$. The $\bar{\mathbf{C}}$ is the matrix obtained by vertically replicating L times the difference between the i^{th} and j^{th} patch centre difference. This can be written in matrix form as $\bar{\mathbf{C}} = [\mathbf{c_2} - \mathbf{c_1}, \mathbf{c_3} - \mathbf{c_2}, \dots, \mathbf{c_{K-1}} - \mathbf{c_K}]$

where $$\mathbf{c_j} - \mathbf{c_i} = \begin{bmatrix} cx_j^1 - cx_i^1 \\ cy_j^1 - cy_i^1 \\ \vdots \\ cx_j^L - cx_i^L \\ cy_j^L - cy_i^L \end{bmatrix}$$ and $\mathbf{C}$ $\in \mathrm{R}^{L\times(K-1)}$

All the matrices used to solve equation(B.1) is explained here. For the derivation of these equation, Chen *et al.* paper can be referred [1].