Auto-alignment of Knee MR Scout Scans through Redundant, Adaptive and Hierarchical Anatomy Detection

Yiqiang Zhan, Maneesh Dewan, and Xiang Sean Zhou

Siemens Medical Solutions, USA

Abstract. 3D knee magnetic resonance (MR) scout scan is an emerging imaging sequence that facilitates technicians in aligning the imaging planes of diagnostic high resolution MR scans. In this paper, we propose a method to automate this process with the goal of improving the accuracy, robustness and speed of the workflow. To tackle the various challenges coming from MR knee scout scans, our auto-alignment method is built upon a redundant, adaptive and hierarchical anatomy detection system. More specifically, we learn 1) a *hierarchical redudant* set of anatomy detectors, and 2) ensemble of group-wise spatial configurations across different anatomies, from training data. These learned statistics are integrated into a comprehensive objective function optimized using an expectation-maximization (EM) framework. The optimization provides a new framework for *hierarchical* detection and *adaptive* selection of anatomy primitives to derive optimal alignment. Being extensively validated on 744 clinical datasets, our method achieves high accuracy (sub-voxel alignment error), robustness (to severe diseases or imaging artifacts) and fast speed (∼5 secs for 10 alignments).

Keywords: MR, Scout scans, Knee, Alignment, Detection, Learning.

1 Introduction

Magnetic resonance imaging (MRI) has been successfully used to diagnose various knee diseases ranging from acute knee injuries to chronic knee dysfunction, e.g., ligament tears, patella dislocation, meniscal pathology and arthropathies, etc [8]. Owing to the inherent imaging physics and speed limitations of MR, the diagnostic MR images have high resolution with-in slice and low resolution across slices (c.f. Fig. 1 (a)). Hence, the diagnostic ability and quality of MR images highly depends on the accuracy of positioning of the MR slice group (imaging planes), such that the anatomy of interest lies in the imaging plane of high resolution. Recently, a new MR sequence, 3D knee scout scan, has been introduced to improve the quality of MR knee workflow. A 3D knee scout scan has low but **isotropic** image resolutions. Although it is not of diagnostic quality, it provides 3D context for accurate positioning of the high-resolution (high-res) diagnostic slice groups. In the current workflow, the technicians will position

G. Székely and H.K. Hahn (Eds.): IPMI 2011, LNCS 6801, pp. 111–122, 2011.

and adjust (rigidly align) the high-res slice groups in accordance with the relevant anatomies using the MR scout scans. The high-res MR knee scans are then done using these manually aligned slice groups. For example, as shown in Fig. 1 (b), for a high-res menisci scan, the transversal imaging slice group should be parallel to the meniscus plane and the transversal in-plane rotation should be aligned to the line connecting the two lower edges of condyle (the yellow dashed line). Since the manual alignment is time consuming and often not reproducible for follow-up studies, an automatic alignment method is desired to increase the speed, accuracy and reproducibility of the MR knee workflow.

However, automatic alignment of MR knee scout scans is not trivial due to the following challenges coming from MR knee scout scans: (1) **Complex transformation:** Compared to other alignment problems, knee transformation is much more complex, including 3D rotation, scaling, and articulation (Fig.3(c)(e)). This not only increases the dimensions of the parameter space, but also induces additional appearance variations of the same anatomy. (2) **Large variations of intensity and contrast:** Different knee coils and magnetic fields are used in various clinical sites, which causes the intensities of the same anatomy and contrast between different anatomies to vary a lot across MR scout scans . (3) **Abnormality/Variations in knee scout scans:** Besides variations introduced from patient ages and sizes ((Fig.3(c)(d)), knee diseases, implants and imaging artifacts dramatically change the appearance of anatomies (Fig.3(f)(g)). Moreover, owing to patient constraints or technician errors, sometimes the key anatomy used for alignment might be out of the field of view.(Fig.3(a)(b)). (4) **Runtime efficiency requirements:** Auto-align algorithm must have high runtime efficiency as its key target is to speed up the MR knee workflow.

To handle all the aforementioned challenges simultaneously, we propose an iterative alignment method that uses the learned local appearance characteristics and spatial configurations of anatomy primitives from a large set of training dataset. In principle, our method belongs to feature-based alignment/registration category. However, our method has three distinctive properties, which are particularly designed to tackle the challenges of knee MR scout scan alignment.

- Redundancy: The learned local appearance cues of distinctive anatomy primitives can not always be reliably detected, especially when they have dramatic appearance changes or are completely out of FOV (*Challenge 2,3*). Therefore, we employ "more than enough" anatomy primitives to derive alignment. For

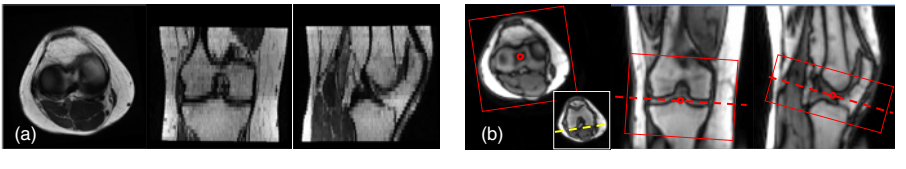

Fig. 1. (a) A typical diagnostic MR knee transversal scan. (b) A MR knee scout scan. Rectangle boxes: coverage of MR slice group. Circles: imaging centers. Yellow dashed line (in the small thumbnail) defines the transversal in-plane rotation.

example, in theory, three non-corlinear points can exclusively determine a rigid alignment. In our method, a far greater number of anatomies are used to highly increases the robustness of our approach with respect to the various abnormalities in MR scout scans.

- Adaptivity: "Redundant" anatomy primitive detectors are "adaptively" used in different stages of alignment. The anatomy primitives with high transformation-invariant appearance are used in the initial stage, while the ones with larger transformation-sensitive appearances are employed in the later stage. "Adaptivity" makes our method efficient without losing robustness (*Challenge 3,4*).

- Hierarchy: We build a hierarchy of detectors with different transformation sensitivities i.e. for each anatomy primitive, we learn detectors with different levels of alignment. At run-time, detectors less sensitive to transformation are used first to estimate an approximate but robust transformation. The detectors more sensitive to transformation (more discriminative) are run at a later stage, and hence increase the accuracy of alignment. By using this "hierarchy" strategy, appearance variations resulting from complex transformation (*Challenge 1*) are gradually removed and accurate detection and alignment can be finally achieved.

The above strategies are realized through a comprehensive objective function. The three terms in this objective function correspond to anatomy primitive-based alignment, anatomy detection using appearance characteristics and spatial constraints of the relative positions of anatomy primitives. An expectation-maximization (EM) framework is employed to optimize the objective function.

2 Related Work

As MR knee scout alignment is a new application in MR workflow, to our best knowledge, there are very few studies on this problem so far [1][5]. Although these pioneer studies open the window for automation of the MR knee workflow, they still have errors around 3-5mm, due to the aforementioned challenges in MR knee scout. At the same time, research on related problems, e.g., brain alignment and face alignment, has been conducted for more than a decade. In this section, we will review these methods and discuss if there are any off-the-shelf methods to solve our problem.

Due to various applications in neuroscience, brain alignment/registration has been extensively studied. In general, brain registration aims to maximize the similarity between two brain images within a specific transformation space. The main core of different methods lies in the definition of the similarity functions, which can be classified into voxel-based [12][3] and feature-based similarities [6][9]. In voxel-based registration methods, information from the entire image is employed to establish a voxel-wise corresponding map across different images. However, in knee scout scan alignment, the alignment of slice group usually needs a reference to only a part of the image and/or a few anatomy primitives. Global image information thus becomes unnecessary and sometimes will decrease alignment accuracy. For example, for knee menisci alignment (see Fig. 1(b)), by using a voxel-based similarity function, it is almost impossible to be immune to

the deformation/artifacts in any regions of scout scans(see Fig. 3 (f)) as these areas contribute to similarities as well. Although a pre-segmentation can help to mask out the relevant anatomies for registration, segmentation itself might not be robust to severe abnormality, either. Feature-based similarity function is more suitable to our problem as it provides the capability to focus on local information. However, in most studies [9], features are defined by low level salient structures, e.g., boundary points, crest lines. As these features do not bear exclusive anatomy definitions, the matching of features can not guarantee the matching of corresponding anatomies. (Note that corresponding anatomies might not always appear as salient structures due to disease and imaging artifacts.) [6] started to use real anatomy landmarks for feature-based registration. However, this method is not fully automatic as its landmark detection requires manual operations. In addition, it is worth noting that a large number of deformable registration methods can not be used in our application, as MR scout scan have to be rigidly aligned and the conversion from deformation fields to rigid transformation is still an open problem.

Face alignment is another related research topic originating from the areas of face detection and recognition, which aims to remove appearance variations coming from different poses to increase the accuracy of face recognition. In [10], face detectors for different poses are exclusively performed and the pose corresponding to the strongest detector response is regarded as the aligned pose. Different machine learning technologies were also applied to solve this problem, e.g., vector boosting [4], etc. In [7], face alignment and detection are tackled altogether through learning a convolution network. Although these methods achieve tremendous success in face alignment, they can not be directly borrowed to our application due to the inherent differences between face and knee alignment. First, knee transformation is much more complex than face. While face transformation is usually modeled as a 2D rotation, knee transformation involves 3D rotation, scaling and articulate transformation. With the higher transformation space/dimension, some existing face alignment methods become computationally infeasible. Second, knee and face alignment have different accuracy requirements. Errors of face alignment are usually allowed to be up to 10 degrees. In contrast, as the accuracy of knee alignment is very critical to the subsequent high-res diagnostic scans, the accuracy of an acceptable knee alignment is much higher (<3mm translation error, <2 degree rotation error).

3 Methods

Formulation: In our method, a MR scout scan is aligned to a canonical space based on a set of automatically detected anatomical primitives/landmarks. A classic landmark-based alignment problem is usually formulated as a least square problem:

$$\min_{T} \parallel T\mathbf{m} - \mathbf{p} \parallel^2 \tag{1}$$

where $\mathbf{p}$ denotes the detected landmarks in a MR scout scan and $\mathbf{m}$ denotes corresponding landmarks in the model space (Both $\mathbf{p}$ and $\mathbf{m}$ are vectors con-

catenated by 3D coordinates of landmarks.) T is the optimal transformation that brings the scout scan to the canonical model space. Since the alignment and detection are highly dependent on each other, in order to estimate them jointly, Eq. 1 should be extended to incorporate landmark detection as:

$$\max_{T,\mathbf{p}} e^{-\|T\mathbf{m}-\mathbf{p}\|^2} + D(\mathbf{p} \mid I;T) + S(\mathbf{p}) \tag{2}$$

The first additional term, $D(\mathbf{p} \mid I;T)$, denotes the likelihood of the landmark locations, given image (scout scan) I and transformation T. This term can be derived from anatomy detectors which capture the appearance characteristics of landmarks. It is worth noting that D depends not only on image I but also on transformation T, which is in accordance to the fact that appearance characteristics are influenced by transformations as well. The second additional term $S(\mathbf{p})$ indicates the likelihood of the landmark locations in terms of their spatial relations. In other words, this term incorporates high-level spatial statistics across different anatomies and is exploited to make the alignment robust to erroneous detections. (In Eq. 2, alignment residual has an exponential format to have the same value range as the second and third term.) The current objective function lacks "adaptivity", which is critical to achieve robust alignment, especially when abnormality exists (diseases, artifacts or bad positioning). Hence, Eq. 2 is extended by inducing a selection matrix $\mathbb{C}$ as:

$$\max_{T,\mathbf{p},\mathbb{C}} e^{-\|T\mathbb{C}\mathbf{m}-\mathbb{C}\mathbf{p}\|^2} + D(\mathbb{C}\mathbf{p} \mid I;T) + S(\mathbb{C}\mathbf{p}) \tag{3}$$

Here, $\mathbb{C}$ is a diagonal matrix whose diagonal elements can be either 1 or 0. It indicates whether an anatomy landmark is used for alignment. By inducing $\mathbb{C}$, three additional properties are introduced into the alignment framework. First, $\mathbf{p}$ can be defined to contain more anatomy landmarks beyond what is "necessary". For example, although three non-corlinear landmarks is enough to derive a rigid transformation, our system exploits much more landmarks (redundancy) to achieve robustness. Second, given different values of $\mathbb{C}$, the redundant landmarks can be adaptively selected based on appearance (detection confidence) or spatial locations (spatial statistics). Third, during the optimization process, $\mathbb{C}$ is updated dynamically such that detectors with different transformation sensitivities are used hierarchically. Overall, the additional $\mathbb{C}$ in Eq.3 facilitates the alignment system to be more robust and accurate. Next, we will introduce details of Eq. 3 as well as its optimization method.

Learning-based Anatomy Detection: $D(\mathbb{C}\mathbf{p} \mid I;T)$ in Eq. 3 indicates the likelihood of the existence of landmarks, which is in fact an anatomy detection problem. Inspired by Viola's work on face detection [11], we use a learning-based method that starts from a huge set of *elementary* features. A classifier with feature selection capability is then employed to select and combine these elementary features to characterize complex anatomy appearance.

In training, we first annotate landmarks in a set of MR scout scans as positive samples. The voxels "off" of these landmarks are considered as negative samples.

A large number of elementary features are then extracted in the neighborhood of each training sample. Our elementary features are generated by a set of mother functions, $\{H_l(\mathbf{x})\}$, which consists of one or more 3D rectangle functions with different polarities. By scaling the mother functions and convoluting them with the original image, a set of spatial-frequency spaces are constructed. For any voxel $\mathbf{x_0} \in \Re^3$, its feature vector $\mathfrak{F}(\mathbf{x_0})$ is finally obtained by sampling these spatial-frequency spaces in the neighborhood of $\mathbf{x_0}$ as Eq. 4.

$$\mathfrak{F}(\mathbf{x_0}) = \bigcup_l \{F_l(\mathbf{x_i}, s_j) | \mathbf{x_i} \in \mathbb{N}(\mathbf{x_0}), s_{min} < s_j < s_{max}\} \tag{4}$$

where $F_l(\mathbf{x}, s) = H_l(s\mathbf{x}) * I(\mathbf{x})$ denotes the spatial-frequency spaces constructed by the mother functions. These feature vectors are fed into *Adaboost* algorithm [11] to learn appearance signatures of anatomy primitives. The output of the learned classifier $\mathcal{A}(\mathfrak{F}(\mathbf{x}))$ indicates the existence likelihood of a landmark at $\mathbf{x}$.

Our method has its hallmark in the alignment of training images, which is different from a direct extension of [11]. Since the elementary features are not rotation-invariant, spatial transformations of images will change feature values. Hence, learning appearance characteristics becomes more complicated. One can envision two extreme solutions. First one is to remove this kind of feature variations through alignment of training images. The flip side is that these detectors learned from "perfectly" aligned training samples will be fragile to spatial transformations at run-time, since they never "saw" feature variations resulting from spatial transformations. Another way is to learn local appearance detectors in presence of these transformation variations. This makes the detectors robust to transformation, though the cost one has to pay is detection accuracy. From another perspective, the appearance features of different anatomies exhibit different sensitivity to spatial transformations. As shown in Fig. 2, local appearances around the center of tibia intercondylar eminence (CTIE) is more sensitive to the rotation than that of the center of lateral meniscus (CLM), i.e., with random perturbation, the overlaid images are fuzzier around CTIE (within red rectangles). Thus, all these factors should also be considered in the training framework for generating effective learned detectors.

Therefore, we propose to train hierarchy of detectors using images with different levels of alignment. We first manually align all training images to a common canonical space. Random rigid perturbations are then applied to these aligned training images for feature extraction. Assume the alignment matrix of a training image I_i is T_i, after applying random rigid transformation $r_i \in \mathcal{R}$ (where $\mathcal{R}$ defines the random perturbation range), the construction of spatial-frequency spaces is re-formulated as:

$$F_l(\mathbf{x}, s; T_i, r_i) = H_l(s\mathbf{x}) * (r_i \circ T_i \circ I_i(\mathbf{x})) \tag{5}$$

where $\circ$ is the rigid transformation operator. Elementary features are then sampled in $F_l(\mathbf{x}, s; T_i, r_i)$ using Eq. 4 and used to train a detector.

For the same anatomy, by changing the random perturbation range $\mathcal{R}$, a set of detectors with different transformation sensitivities can be trained as

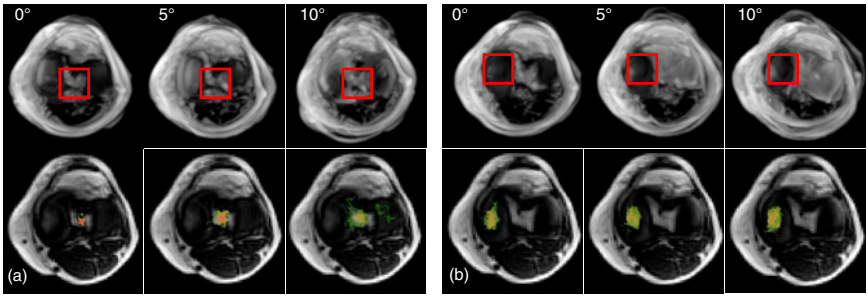

Fig. 2. Training samples with different levels of alignment and their corresponding response maps with respect to different anatomies. (a) CTIE (b)CLM. (Top row) Overlay of 20 training samples using different levels of alignment. (Bottom row) Overlay of response maps from detectors trained with images with different levels of alignment.(From left to right) 0, 5, 10 degrees random rotation after alignment, respectively.

$\{\mathcal{A}(\mathfrak{F}(p); R_j)|j = 1 \cdots N\}$, where $\mathcal{R}_j$ indicates different random perturbation ranges and $\mathcal{R}_1 \subset \mathcal{R}_2 \subset \cdots \subset \mathcal{R}_N$. Finally, $D(\mathbb{C}\mathbf{p} \mid I; T)$ in Eq. 3 is defined as the integration of the responses given by these detectors across all anatomy primitives as Eq. 6.

$$D(\mathbb{C}\mathbf{p} \mid I; T) = \sum_i c_{ii} \sum_j \mathcal{A}_i(\mathfrak{F}(p_i); \mathcal{R}_j)/(\sum_i c_{ii} \sum_j \mathbf{1}) \tag{6}$$

Here, $\mathcal{A}_i(.; \mathcal{R}_j)$ denotes the detector corresponding to the ith anatomy primitive trained from samples with random perturbation range $\mathcal{R}_j$. c_{ii} is a diagonal element of selection matrix $\mathbb{C}$, corresponding to the ith anatomy primitive.

Fig 2 (bottom row) shows two interesting characteristics of detectors trained with different levels of alignment. *First*, detectors trained with looser alignment (larger perturbation) show a flatter response map, while detectors trained with a tighter alignment (less perturbation) are more discriminative and provide a peaky response map. *Second*, different anatomies show different sensitivities to spatial transformations. As shown in Fig 2 (a), responses from CLM detectors display similar patterns with variations in the random perturbation range. On the contrary, responses from CTIE detectors (Fig 2 (b)) change from a clear peaky pattern to a noisy flatter pattern with increase in perturbation range. These two characteristics indicate the necessity of a "hierarchical" detection schema to achieve both robustness and accuracy, which is realized in our iterative optimization framework.

Ensemble of Group-wise Spatial Configurations: Imaging artifacts, severe diseases and variations of field of views (FOV) can change local appearance dramatically causing erroneous detections. Hence, it is important to prune these erroneous detections by exploiting spatial correlations between different anatomy primitives (role of $S(\mathbb{C}\mathbf{p})$ in Eq. 3).

Different from the active shape model that learns global shape statistics, we propose to learn group-wise spatial configurations for each anatomy primitive p_i. Assume $U(\mathbf{p}\backslash p_i)$ is a subset of anatomy primitives which does not contain p_i, i.e., $U(\mathbf{p}\backslash p_i) \subset \{\mathbf{p}\backslash p_i\}$. The group-wise spatial configuration between p_i and $U(\mathbf{p}\backslash p_i)$ is modeled as a conditional probability following Gaussian distribution:

$$s(p_i|U(\mathbf{p}\backslash p_i)) = \frac{1}{(2\pi)^{(3/2)}|\Sigma|^{1/2}} exp(-\frac{1}{2}(p_i-\mu)^T\Sigma^{-1}(p_i-\mu)) \tag{7}$$

where μ and Σ are two statistical coefficients that are learned as follows.

We employ a linear model to capture the spatial correlation between p_i and $U(\mathbf{p}\backslash p_i)$, i.e., $p_i = \mathfrak{C}\cdot\mathbb{U}$, where $\mathbb{U}$ is a vector concatenated by $\{p_j|p_j \in U(\mathbf{p}\backslash p_i)\}$. Given a set of training samples, the linear correlation matrix $\mathfrak{C}$ can be learned by solving a least squares problem. Furthermore, μ and Σ are calculated as:

$$\begin{aligned} \mu &= E[\mathfrak{C}\cdot\mathbb{U}] \\ \Sigma &= E[(\mathfrak{C}\cdot\mathbb{U}-\mu)(\mathfrak{C}\cdot\mathbb{U}-\mu)^T] \end{aligned} \tag{8}$$

In this way, each anatomy primitive p_i receives a set of conditional probabilities from different subsets of $\{U_j(\mathbf{p}\backslash p_i)|j=1\ldots M\}$ defined as Eq. 7. It is worth noting that a correctly detected p_i might also receive a lower conditional probability from a subset $U(\mathbf{p}\backslash p_i)$ that includes erroneous detections. On the other hand, in most cases, the conditional probabilities received by an erroneously detected p_i are *all* low. (The only exception is that the erroneous detections happen to construct a correct spatial configuration, which very rarely happens.) Therefore, the maximum value of all conditional probabilities received by p_i is used to measure the "eligibility" of the spatial location of the ith anatomy locations. $S(\mathbb{C}\mathbf{p})$ in Eq. 3 is thus defined as:

$$S(\mathbb{C}\mathbf{p}) = \sum_i c_{ii} \max_{|U_j|\leq K} s(p_i|U_j(\mathbf{p}\backslash p_i))/\sum_i c_{ii} \tag{9}$$

where K is used ($K=3$ in this study) to constrain the cardinality of U_j.

The advantage of our spatial correlation modeling lies in two aspects. First, instead of learning global spatial statistics, we learn spatial correlations within small groups of anatomies, which are assembled in a "democratic" way at runtime. It makes our method robust to missing or gross detection failures. Second, by constraining the cardinality of U_j and using a linear model, our spatial model will not overfit erroneous detections. In addition, the run-time efficiency is also guaranteed.

Adaptive and Hierarchical Alignment: Since the objective function after incorporating Eq. 6 and 9 becomes high-dimensional, non-linear and non-convex, it is difficult to optimize by traditional approaches. Recall Eq. 3, it has three sets of optimization variables T, $\mathbf{p}$ and $\mathbb{C}$, where T is the transformation parameters to be estimated and $\mathbf{p}$ and $\mathbb{C}$ are two sets of latent variables. This structure is similar to the maximum-likelihood estimation problem which is often tackled

using expectation-maximization (EM) algorithm. Thereby, our approach optimizes $\mathbf{p}$, $\mathbb{C}$ and T iteratively following the idea of EM algorithm. The initial T_0 is set as an identity matrix. Afterwards, "expectation stage" (E-stage) and "maximization stage" (M-stage) are performed alternatively until convergence.

E-stage: During E-stage, we aim to optimize the two latent variables $\mathbf{p}$ and $\mathbb{C}$. $\mathbf{p}$ is first optimized by maximizing $D(\mathbb{C}\mathbf{p} \mid I; T)$ in Eq. 3. It is realized by invoking learned anatomy detectors in the image with current transformation, i.e., $T(I)$, to generate response maps corresponding to different anatomies. The maximum positions of these response maps are considered as intermediate optimal $\mathbf{p}$. By setting $c_{ii} = 0$ for p_i with low response and $c_{ii} = 1$ otherwise, we obtain the initial $\mathbb{C}$, which is further optimized according to Eq. (9). This optimization is performed by iteratively "peeling away" detections that "violate" learned spatial statistics. For each iteration, we calculate the minimum value of $\max_{|U_j| \leq K} s(p_i | U_j(\mathbf{p} \backslash p_i))$ across all p_i with $c_{ii} = 1$. If this value is less than a pre-set threshold, the corresponding p_i will be considered as an erroneous detection and its corresponding c_{ii} is set as zero. This process is repeated until $\mathbb{C}$ converges, i.e., no more landmarks are removed as erroneous detection. After each E-stage, we get intermediate optimization results $\mathbf{p}$ and $\mathbb{C}$, which will be used in M-stage.

M-stage: In M-stage, the intermediately optimized $\mathbf{p}$ and $\mathbb{C}$ are fixed. The rigid transformation T is then optimized using the close-form solution presented in [2]. The derived transformation T will be considered as intermediately optimized transformation and used in the next E-stage.

The three hallmarks of our approach, redundancy, adaptivity and hierarchy are indeed realized in this optimization framework. The implementation of "*redundancy*" is straightforward by including more than enough anatomy primitives in $\mathbf{p}$. Although some of them might be erroneously detected, they can be effectively pruned through the optimization of $\mathbb{C}$ in E-stage. "*Adaptivity*" and "*hierarchical*" are realized in the iterative optimization procedure. As introduced before, anatomies have appearance characteristics with varying sensitivity to spatial transformation. In the initial iterations, the lower sensitive anatomies (e.g. CLM, c.f. Fig. 2) with high detector responses are more likely to be selected. In the later iterations, as the image alignment improves, transformation-sensitive anatomies start to show higher detector responses and become highly likely to be selected. In this way, different anatomies are "adaptively" selected in the optimization. From another perspective, each anatomy has a hierarchy of detectors with different sensitivities to spatial transformations. In the initial stages, when the image is far from the well aligned position, detectors with looser alignment (more perturbation) exhibit higher responses and dominate $D(\mathbb{C}\mathbf{p} \mid I; T)$. With the iterations, as the image becomes well aligned, detectors with more rigorous alignment start to generate higher responses as well and thereby contribute more to $D(\mathbb{C}\mathbf{p} \mid I; T)$. Hence, detectors with different levels of alignment are "hierarchically" employed to gradually remove the appearance variations resulting from transformation. These hallmarks thus make our algorithm highly accurate and robust.

4 Experiments

In clinical workflow, our auto-alignment is employed to generate various high-res slice groups relevant to different anatomy parts, e.g., meniscus, patella, cruciate ligaments, etc. Due to space limitation, in this section, we focus on experimental results of knee meniscus slice group, which is the most frequently used for high-res MR imaging. Our method is general and achieves similar performance on other slice boxes. As shown in Fig. 1 (b), the alignment of knee meniscus slice group is defined by the meniscus plane (norm of the slices) and the lower edge of the femur condyle in transversal view (slice in-plane rotation).

Our experimental data includes 924 knee scout scans (180 for training and 744 for testing) with isotropic resolution $2mm$. These datasets come from different clinical sites and were generated by different types of Siemens MR Scanners (Avanto, Verio, etc.) with different knee coils (KN, K15, EX, etc.), which induces additional appearance variations and makes the auto-alignment even more challenging. We validate our method in two aspects, robustness and accuracy.

Robustness: A clinically feasible auto-alignment algorithm has to be robust to the variations not only due to age, gender, demography, but also coming from disease and imaging artifacts. In this study, we test our method on 744 knee scout scans. The alignment results are visually checked and evaluated by experienced experts as "perfect", "acceptable" (clinically accepted but can be improved) and "gross failure" (obvious alignment errors). Table 1 shows the comparison of the performance of our method with other three methods. **Seg:** an in-house benchmark system that is mainly based on heuristic segmentation and detection. **Reg:** a mutual information-based registration method and **Det**: our method without using learned spatial configurations which can be considered as a direct extension of Viola's face detection method on our application. As shown in table 1, **Det** is better than **Seg** and **Reg** as local information helps in robustness. In our method, the robustness is further increased by incorporating ensemble of group-wise spatial configurations. Some representative challenging cases are shown in Fig. 3. Note that even in these challenging cases our algorithm does the knee meniscus alignment very well, i.e. the central transversal slice in the aligned volume passes through the menisci and the condyles are aligned (small thumbnail image in each subfigure).

Accuracy: We also perform quantitative evaluation of the accuracy of our method. In this study, we randomly pick 50 representative clinical cases covering

Table 1. Robustness comparisons of different auto-slice-positioning methods on knee meniscus slice group. GF: Gross Failure. AC: Acceptable. PF: Perfect.

	Seg	**Reg**	**Det**	**Proposed Method**
GF	95 (12.7%)	78 (10.5%)	20 (2.7%)	**1 (0.1%)**
AC	58 (7.8%)	80 (9.4%)	11 (1.5%)	**7 (1.0%)**
PF	591 (79.4%)	596 (80.1%)	713 (95.8%)	**736 (98.9%)**

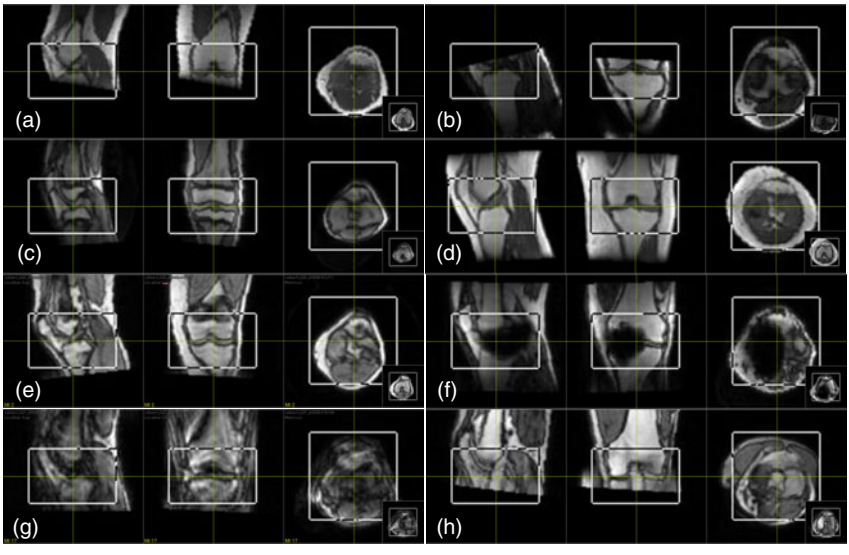

Fig. 3. Robust alignment of our algorithm on representative challenging cases. Each subfigures show the 3 orthogonal MPRs of the aligned volume at the center of the knee meniscus slice box: (a) tibia out of FOV, (b) femur out of FOV, (c) a pediatric patient, (d) an old patient, (e) a patient with bone disease, (f) a patient with metal implant, (g) motion artifacts and (h) folding artifacts.

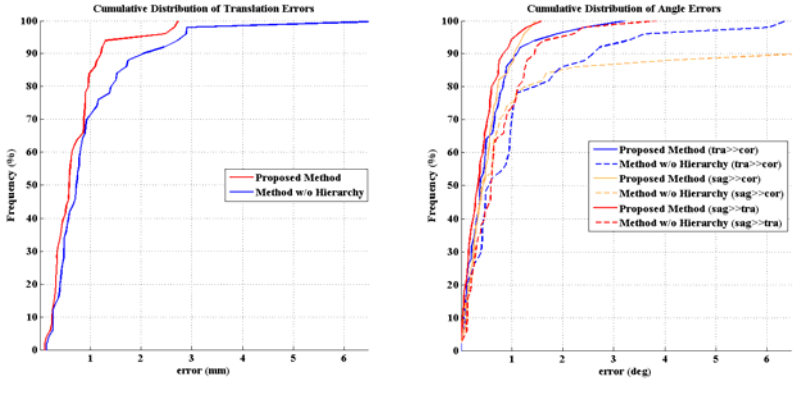

Fig. 4. Cumulative distribution of alignment errors using different alignment methods

different genders, ages and knee sizes. These 50 cases are manually aligned and used as ground truth to derive the quantitative errors. For comparison, we also implemented a method which is similar to the proposed one but without hierarchical detection and alignment, i.e., $\mathbb{C}$ in Eq. 3 is omitted. (We did not compare with **Seg**, **Reg** and **Det**, as gross failure cases make their accuracy numbers

much worse.) Our method has the root mean square error 0.93mm(translation), 0.83° (tra >> cor), 0.54° (sag >> tra), 0.63° (sag >> cor). Our method actually achieves sub-voxel accuracy. The method without hierarchical detection and alignment has larger errors 1.52mm (translation), 1.72° (tra >> cor), 1.02° (sag >> tra), 3.30° (sag >> cor), respectively. More detailed quantitative errors, cumulative error distributions of these two methods, are shown in Fig. 4.

Runtime efficiency: Running on a computer with Dual-Core AMD Opteron (tm) Processor 2212 (2GHz), our method is able to generate 10 alignment for different knee anatomies in 5 seconds. It satisfies the requirement of MR knee workflow.

5 Conclusions

In this paper, we proposed an auto-alignment method for MR knee scout scans, which exploits redundant, adaptive and hierarchical strategies to detect anatomy primitives and align scout scans accordingly. Our method is highly robust, accurate and fast with extensive demonstration on 700+ dataset. It paves the way to automate the positioning of slice boxes for high-res MR knee scanning and improve the MR knee workflow efficacy. It comes as no surprise that the proposed method has been integrated into the latest Siemens MR Scanners.

Acknowledgments. The authors would like to thank Martin Harder for a lot of valuable discussions and his data collection efforts.

References

1. Bystrov, D., Pekara, V.: Automated planning of mri scans of knee joints (2007)
2. Eggert, D.W.: Estimating 3-d rigid body transformations: a comparison of four major algorithms. Machine Vision and Applications 9, 272–290 (1997)
3. Glocker, B., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. Medical Image Analysis 12, 731–741 (2008)
4. Huang, C., Ai, H., Li, Y., Lao, S.: Vector boosting for rotation invariant multi-view face detection. In: ICCV (2005)
5. Jolly, M., Guehring, J.: Automatic Femur Segmentation and Condyle Line Detection in 3D MR Scans for alignment of High Resolution MR. In: ISBI (2010)
6. Joshi, A., Shattuck, D., Thompson, P., Leahy, R.: Surface-constrained volumetric brain registration using harmonic mappings. IEEE TMI 26, 1657–1669 (2007)
7. Osadchy, M., Cun, Y., Miller, M.: Synergistic face detection and pose estimation with energy-based models. J. of Machine Learning Research 8, 1197–1215 (2007)
8. Ostlere, S.: Imaging the knee. Imaging 15, 217–241 (2007)
9. Pennec, X., Ayache, N., Thirion, J.P.: Landmark-based registration using features identified through differential geometry. Academic Press, London (2008)
10. Schneiderman, H., Kanade, T.: A statistical method for 3d object detection applied to faces and cars. In: CVPR (2000)
11. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57, 137–154 (2004)
12. Wells, W., Viola, P., Atsumid, H., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1, 35–51 (1996)