A 3D Femoral Head Coverage Metric for Enhanced Reliability in Diagnosing Hip Dysplasia

Niamul Quader$^{(\boxtimes)}$, Antony J. Hodgson, Kishore Mulpuri, Anthony Cooper, and Rafeef Abugharbieh

BiSICL, University of British Columbia, Vancouver, Canada
niamul@ece.ubc.ca

Abstract. Developmental dysplasia of the hip (DDH) in infancy refers to hip joint abnormalities ranging from mild acetabular dysplasia to irreducible femoral head dislocations. While 2D B-mode ultrasound (US) is currently used clinically to estimate the severity of femoral head subluxation in infant hips, such estimates suffer from high inter-exam variability. We propose using a novel 3D US-derived dysplasia metric, the 3D femoral head coverage (FHC_{3D}), which characterizes the 3D morphology of the femoral head relative to the vertical cortex of the ilium in an infants hip joint. We compute our 3D dysplasia metric by segmenting the femoral head using a voxel-wise probability map based on a tomographic reconstruction of 2D cross-sections each labeled with a probability score of that slice containing the femoral head. Using a dataset of 20 patient hip examinations, we demonstrate that our reconstructed femoral heads agree reasonably well with manually segmented femoral heads (mean dice coefficient of 0.71), with a significant reduction in variability of the associated metric relative to the existing manual 2D-based FHC ratio ($\sim$20% reduction, $p < 0.05$). Our findings suggest that the proposed 3D dysplasia metric may be more reliable than the conventional 2D metric, which may lead to a more reproducible test for diagnosing DDH.

1 Introduction

Developmental dysplasia of the hip (DDH) describes mechanical instability of the hip joint due to abnormal developments of the acetabulum and femoral head, and affects $0.16\% - 2.85\%$ of all newborns [1]. If the femoral head is left in an abnormal position at infancy, the surrounding anatomy may develop abnormally [2], which could lead to subsequent required expensive corrective surgical procedures and hip osteoarthritis in later years [3]. In the United States alone, the direct financial burden of failing to detect early DDH is in the order of \$1B/year [3], without considering costs of revision surgeries and socioeconomic costs.

To diagnose DDH early in infancy, the American College of Radiology [4] suggests using 2D ultrasound (US) to quantitatively estimate abnormalities in the positions of the acetabulum and femoral head relative to the vertical cortex of the ilium - these two abnormalities are complementary to one another, so both need

to be diagnosed for characterizing DDH. The acetabular abnormality is commonly quantified using the alpha angle (α_{2D}: defined as the angle between the acetabular roof and the vertical cortex of the ilium), and the abnormality in femoral head position is quantified using the femoral head coverage (FHC_{2D}: defined as the femoral head portion sitting in the acetabular cup of the hip joint) [4]. More specifically, an acetabulum is considered normal when $\alpha_{2D} > 60°$, and a femoral head position is considered normal when $FHC_{2D} > 50\%$.

Both of these commonly used dysplasia metrics are problematic - they suffer from poor within-hip repeatability, i.e., the variability between repeated measurements on the same hip tends to be high. For example, the standard deviation within repeated FHC_{2D} measurements collected by different raters is around 9% [5]. Recently, Hareendranathan et al. [6] and Quader et al. [7] suggested using 3D US to reduce variability in diagnosing acetabular abnormality in infants. Similarly, in this paper, we propose using 3D US to reduce variability in diagnosing femoral head abnormality in infants by estimating an intrinsically 3D morphology metric - the 3D femoral head coverage, FHC_{3D}, which we define as the ratio of femoral head portion medial to the plane of the ilium (Fig. 2(b), (d)).

To the best of our knowledge, no other work has proposed using FHC_{3D} or any other 3D extension of a FHC measurement. A few methods have though been proposed for automatically segmenting a femoral head in 3D US [7,8]. For example, Quader et al. [7] segmented the femoral head in 3D US using an M-estimator SAmple Consensus (MSAC) sphere-fit algorithm on hyperechoic bone and cartilage boundaries. However, this algorithm assumed that the hypoechoic femoral head touches all neighboring bone and cartilage boundaries, even in a subluxed or dislocated hip. An earlier study [8] proposed using an intensity and local texture feature (structure tensor) based level-set framework to segment the spherical femoral head, though their method requires the user to input the center coordinates of the femoral head. Furthermore, neither of these previous studies provided any quantitative validation of their segmentation accuracy.

In this paper, we present an automatic approach for segmenting the femoral head and propose a novel 3D dysplasia metric (FHC_{3D}) from 3D US of the neonatal hip. Our specific contributions are: (1) proposing a method for automatically segmenting the femoral head in 3D US using a novel voxel-wise map that estimates the probability of a voxel belonging to the femoral head, (2) extending the femoral head coverage concept to 3D, (3) presenting an automatic method for estimating FHC_{3D} from a planar approximation of the ilium and the voxel-wise probability map, (4) quantitatively comparing the agreement between automatically segmented femoral heads against expert-labelled spherical femoral heads, and (5) investigating both inter- and intra-rater variability of FHC_{3D} in comparison to FHC_{2D}.

2   Methods

In a 3D B-mode US image of an infants hip, $U(x, y, z)$, we define FHC_{3D} as the ratio of volume of the femoral head portion that is medial to the plane of

the ilium to the total volume of the femoral head (Fig. 2(b), (d)). To estimate FHC_{3D}, we first extract a voxel-wise probability map, $P(x, y, z)$, characterizing the likelihood of a voxel belonging to the femoral head (Sect. 2.1). Next, we identify a planar approximation to the vertical cortex of the ilium, I, and use both P and I to calculate FHC_{3D} (Sect. 2.2).

2.1   Femoral Head Segmention

The femoral head in infant hips is unossified and appears hypoechoic in US (Fig. 2(a), (c)). It is surrounded by anatomical structures with distinctive sonographic properties, e.g. the ilium which has a high sonographic response at its boundary and a shadow region beneath it, the labrum, the triradiate cartilage,
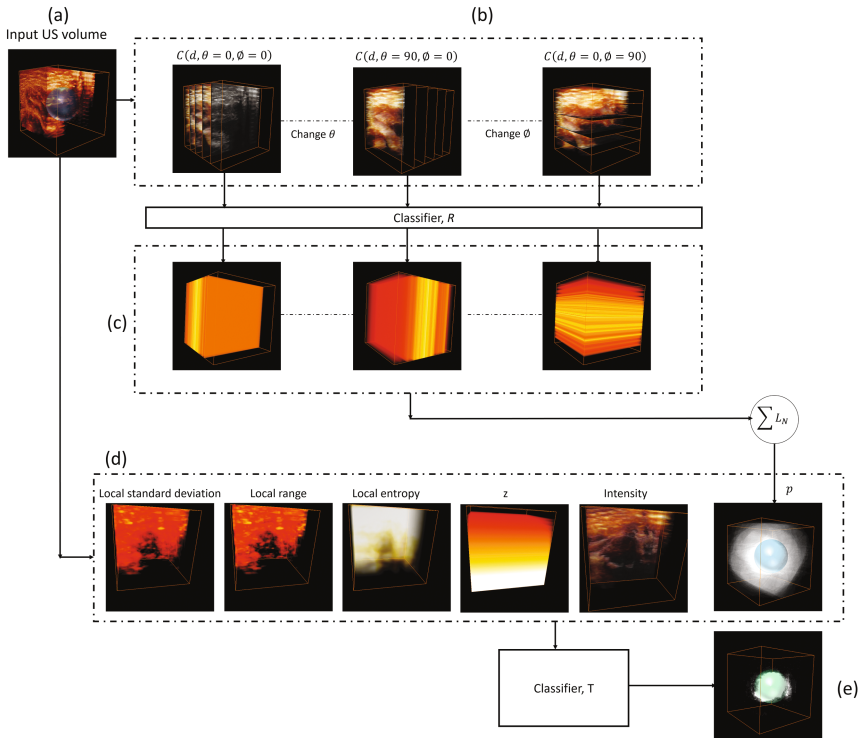
Fig. 1. Overview of our proposed voxel-wise probability map extraction. (a) Overlay of example US volume and a manually segmented femoral head. (b) N number of C are evaluated using classifier R to determine their likelihood to intersect the femoral head. (c) Back-projected likelihood scores, L, for each of the cross-sections C. (d) Back-projected responses are summed and normalized to construct voxel-wise probability map, with an overlay of the manually segmented femoral head (green). This map is used with other features, and fed in the second classifier, T, to estimate probability of each voxel belonging to the femoral head (e).

the greater trochanter, etc. A cross-section of U, $C(d, \theta, \phi)$ (d is the shortest distance of C from the origin of (x, y, z), and θ, ϕ are rotations about x and y axis, respectively, of a reference plane defined by $z = 0$ to a plane parallel to C, Fig. 1b) that intersects the femoral head is expected to include a hypoechoic region surrounded by cross-sections of the neighbouring anatomical structures. In order to segment a femoral head, we therefore start by training a random forest classifier, R, towards distinguishing C intersecting the femoral head from those C that do not. The classifier, R (65 decision trees, minimum number of observations per tree set equal to 1, no pruning applied), is trained using histogram of oriented gradients and local binary pattern features extracted from samples of cross-sections from US volumes. In a test US volume, the likelihood of C intersecting the femoral head (or our tomographic response for C) is evaluated using the trained classifier, R. We encode this tomographic measure throughout the coordinate space of C, and back-project or interpolate it onto the (x, y, z) coordinates, $L(x, y, z)$ (Fig. 1c). For N cross-sections sampled within the test US volume, we construct our tomographic voxel-wise probability map as: $p = norm(\sum_N L_N)$ (Fig. 1d), where $norm(.)$ is the unity-based normalization.

To further enhance the tomographic voxel-wise probability map of the femoral head, we combine this map p with seven voxel-wise features (local standard
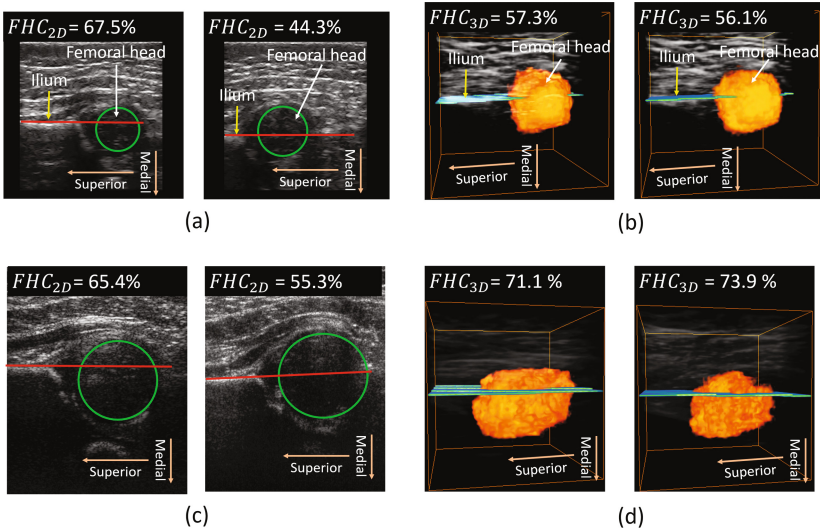
Fig. 2. Examples of manually measured FHC_{2D} and automatically extracted FHC_{3D}. (a) and (c) show example FHC_{2D} measurements made by two raters on two separate hips (hip 1 in (a) and hip 2 in (c)). (b) and (d) show automatically extracted FHC_{3D} measures from US volumes that were scanned by two separate raters on two separate hips (hip 1 in (b) and hip 2 in (d)). In these exemplary cases, US volumes and their FHC_{3D} measures appear similar whereas 2D US images and their FHC_{2D} measures vary widely.

deviation, entropy, range in a 1 mm-by-1 mm-by-1 mm neighborhood, intensity, x, y and z), and feed them into a second random forest classifier, T (10 decision trees, minimum number of observations per tree equal to 1, no pruning applied), to estimate a probabilistic score, $P(x, y, z)$, of each voxel belonging to the femoral head (Fig. 1e). In subsequent steps, we use both P and the center of the femoral head, which we calculate as: $c = [c_x, c_y, c_z] = \sum_X (PX)/\sum P$, where $X = [x, y, z]$.

2.2   Localizing Vertical Cortex of Ilium and Estimating FHC_{3D}

Once we have identified the femoral head, we localize the vertical cortex of the ilium. The ilium is a ossified bone boundary that presents as a hyperechoic image response and attenuates the US signal beyond it [9]. Furthermore, the ilium is also spatially continuous. To extract ilium's boundary, we first enhance the hyperechoic responses in U using the phase symmetry feature, $PS(x, y, z)$ [10] - an intensity invariant symmetry-enhancing feature that has been shown to be robust to US signal dropout [10,11]. Next, we compute the Hessian of PS and apply Descoteauxs sheet-enhancing filter [12] to selectively enhance hyperechoic responses that are part of sheet-like structures, $SPS(x, y, z)$. We then incorporate attenuation features [11] with SPS to get a probabilistic measure of bone boundaries, B, in a region of interest nominally superior to the femoral head defined by $x = 1 : c_x - r, y = c_y - r : c_y + r, z = c_z - r : c_z + r$, where r is set to 8 mm, which represents an average dimension for the radius of the femoral head in infants [9].

Having extracted B, we approximate the vertical cortex of the ilium, I, to be a plane within 15° of x, y plane since I tends to be perpendicular to the US beam in coronal scans. We approximate this plane I using a M-estimator SAmple Consensus (MSAC) algorithm [13]. Finally, we estimate FHC_{3D} as $FHC_{3D} = \sum_M P/\sum_X P$, where $X = [x, y, z]$ and M represents all X locations that are medial to I.

3   Results and Discussion

Data Acquisition and Experimental Setup: In this study, two orthopaedic surgeons and four orthopaedic surgeon-trainees participated in collecting B-mode 3D US images and B-mode 2D US videos from 20 infant hips using a SonixTouch Q+ machine and a 4DL14-5/38 Linear 4D transducer at its default penetration settings (obtained as part of routine clinical care under appropriate institutional review board approval). To investigate inter-rater and intra-rater repeatability (i.e., within-hip variability), each hip examination involved two raters - each rater acquired two 3D US images (i.e., four 3D US images in total) and two 2D US videos (i.e., four 2D US videos in total). The surgeon who acquired each of the 2D US videos further chose a 2D US image from the 2D US video and measured of FHC_{2D}. FHC_{3D} angles were calculated for each of the 3D US images using our proposed method of segmenting the femoral head and also using the method described in [7]. All 2D and 3D scans were collected in the coronal plane.

While performing a 3D US scan, the operator ensured that the entire femoral head was available within the stored US volume.

Validation Scheme: To quantify segmentation accuracy, we investigated agreement between manually and automatically estimated femoral heads in 3D US images. We also compared variability (σ) in both FHC_{2D} and FHC_{3D} for each infant hips. We estimated σ by calculating the standard deviation within repeated measurements made on the same hip. We used a leave-one-out-cross validation scheme.

Discrepancy in Femoral Head Segmentation: Femoral heads resulting from our tomographic method seemed to agree reasonably well with manually segmented femoral heads (mean dice coefficient, $D = 0.71$ ($SD = 0.05$) when $P > 0.8$ was classified as femoral head). This agreement was significantly higher than the MSAC sphere fit-based segmentation in [7] ($D = 0.53$ ($SD = 0.12$), $p < 0.01$), suggesting our proposed femoral head segmentation agrees more closely with an expert's judgement. We also found that the tomographic voxel-wise probability map had the highest out-of-bag feature importance (value of 2.9) among all the features used in the second random forest classifier (Fig. 1d), suggesting that this feature contributes most in segmenting the femoral head.

Agreement between FHC_{2D} and FHC_{3D}: Agreement between the $mean(FHC_{2D})$ and $mean(FHC_{3D})$ of each hip examinations was moderate (correlation coefficient, $r = 0.58$ (95% confidence interval 0.28 and 0.78)). We found similar agreement between $mean(FHC_{2D})$ of each hip within the two raters, $r = 0.47$ (95% confidence interval 0.14 and 0.71). These positive r values between FHC_{2D} and FHC_{3D} were expected since FHC_{3D} is a natural 3D extension of FHC_{2D}; however, the low to moderate r values indicated that the repeatability of either FHC_{2D} or FHC_{3D} was low.

Variability of Metrics, σ: FHC_{3D} computed using our proposed method was less variable than the manually estimated FHC_{2D} ($\sigma_{intra,FHC3D}$ of 5.4% vs. $\sigma_{intra,FHC2D}$ of 6.6%, $p < 0.05$; $\sigma_{inter,FHC3D} = 6.1$% vs. $\sigma_{intra,FHC2D} = 8.2$%, $p < 0.05$, qualitative results in Fig. 2 and box-plot in Fig. 3). While these variabilities in FHC_{3D} are still fairly large, they are 18–25% lower than the variability of the currently used FHC_{2D} metric, suggesting that our 3D-based approach can potentially improve reliability of diagnosis. We also found that the variability of the tomographic method was significantly lower than the variability in the MSAC-based approach [7] ($\sigma_{intra,FHC3D} = 10.8$% and $\sigma_{inter,FHC3D} = 9.6$%).

Computational Consideration: The complete process of extracting FHC_{3D} from a 3D US image takes around 95 s when run on a Xeon(R) 3.40 GHz CPU computer with 12 GB RAM. All processes were executed using MATLAB 2015b. In practice, a clinician would not want to wait a minute and a half to know if the scan was acceptable, but it is not necessary to deliver the head coverage metric in near-real time. In this paper, therefore, we have focused on how to

FHC_{3D} vs. FHC_{2D}

(a)

Box-plots of inter-/intra-rater variability

(b)
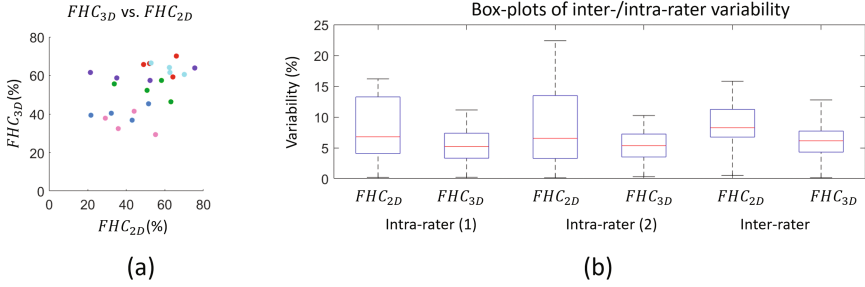
Fig. 3. Variability in FHC measurements. (a) Scatter plot showing examples of variability in FHC measurements on infant hips likely to have DDH ($FHC < 50\%$) and on healthy hips ($FHC > 50\%$), where points representing measures from each of the hips are assigned an unique color. Qualitatively, FHC_{3D} has lower spread than FHC_{2D} suggesting lower variability in FHC_{3D} measures in both healthy and DDH patients. (b) Box-plot of intra-rater variabilities for two raters separately and inter-rater variability between the two raters. Both intra- and inter-rater variabilities are significantly lower for FHC_{3D} compared to manually measured FHC_{2D}.

automate and reduce the variability in computing the FHC. Separately, we are investigating how to rapidly determine whether or not a 3D US scan is adequate for making this measurement or not (building on our previous work on classifying the adequacy of 2D images [14]), so we do not feel that the time needed to compute FHC_{3D} will block clinical use.

4   Conclusions

We presented a new method for segmenting the femoral head in 3D B-mode US volumes, together with an automatic method for extracting a new 3D dysplasia metric, FHC_{3D}. Our femoral head segmentation method agrees reasonably well with an expert's segmentation. Furthermore the 3D US-derived dysplasia metric, FHC_{3D}, is significantly less variable than its 2D counterpart, FHC_{2D}. Though the variability in FHC_{3D} may still be clinically significant, we believe that a 3D morphology-derived dysplasia metric like ours could potentially be valuable in reducing the variability in diagnosing DDH.

References

1. Shorter, D., Hong, T., Osborn, D.A.: Cochrane review: Screening programmes for developmental dysplasia of the hip in newborn infants. Evid. Based Child Health Cochrane Rev. J. **8**(1), 11–54 (2013)
2. Gulati, V., Eseonu, K., Sayani, J., Ismail, N., Uzoigwe, C., Choudhury, M.Z., Gulati, P., Aqil, A., Tibrewal, S.: Developmental dysplasia of the hip in the newborn: a systematic review. World J. Orthop. **4**(2), 32–41 (2013)
3. Price, C.T., Ramo, B.A.: Prevention of hip dysplasia in children and adults. Orthop. Clin. North Am. **43**(3), 269–279 (2012)

4. American College of Radiology: ACR-AIUM practice guideline for the performance of the ultrasound examination for detection and assessment of developmental dysplasia of the hip (ACR guidelines) (2012)
5. Peterlein, C.D., Schüttler, K.F., Lakemeier, S., Timmesfeld, N., Görg, C., Fuchs-Winkelmann, S., Schofer, M.D.: Reproducibility of different screening classifications in ultrasonography of the newborn hip. BMC Pediatrics **10**(1), 98 (2010)
6. Hareendranathan, A.R., Mabee, M., Punithakumar, K., Noga, M., Jaremko, J.L.: Toward automated classification of acetabular shape in ultrasound for diagnosis of DDH: contour alpha angle and the rounding index. Comput. Methods Program. Biomed. **129**, 89–98 (2016)
7. Quader, N., Hodgson, A., Mulpuri, K., Cooper, A., Abugharbieh, R.: Towards reliable automatic characterization of neonatal hip dysplasia from 3D ultrasound images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9900, pp. 602–609. Springer, Cham (2016). doi:10.1007/978-3-319-46720-7_70
8. de Luis-Garcia, R., Alberola-Lopez, C.: Parametric 3D hip joint segmentation for the diagnosis of developmental dysplasia. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2006, pp. 4807–4810. IEEE (2006)
9. Graf, R.: Fundamentals of sonographic diagnosis of infant hip dysplasia. J. Pediatric Orthop. **4**(6), 735–740 (1984)
10. Kovesi, P., et al.: Symmetry and asymmetry from local phase. In: Tenth Australian Joint Conference on Artificial Intelligence, vol. 190, pp. 2–4. Citeseer (1997)
11. Quader, N., Hodgson, A., Abugharbieh, R.: Confidence weighted local phase features for robust bone surface segmentation in ultrasound. In: Linguraru, M.G., Oyarzun Laura, C., Shekhar, R., Wesarg, S., González Ballester, M.Á., Drechsler, K., Sato, Y., Erdt, M. (eds.) CLIP 2014. LNCS, vol. 8680, pp. 76–83. Springer, Cham (2014). doi:10.1007/978-3-319-13909-8_10
12. Descoteaux, M., Audette, M., Chinzei, K., Siddiqi, K.: Bone enhancement filtering: application to sinus bone segmentation and simulation of pituitary surgery. Compu. Aided Surg. **11**(5), 247–255 (2006)
13. Torr, P.H., Zisserman, A.: MLESAC: a new robust estimator with application to estimating image geometry. Comput. Vis. Image Underst. **78**(1), 138–156 (2000)
14. Quader, N., Hodgson, A.J., Mulpuri, K., Schaeffer, E., Abugharbieh, R.: Automatic evaluation of scan adequacy and dysplasia metrics in 2-D ultrasound images of the neonatal hip. Ultrasound Med. Biol. **43**, 1252–1262 (2017)