Myat Su Yin, Peter Haddawy, Siriwan Suebnukarn, Phattanapon Rhienmora (2015). Proceedings of the 2nd International Conference on Innovation in Education. Thailand: Institute for Innovative Learning, Mahidol University

Automated Outcome Scoring in a Dental Surgical Training Simulator

Myat SU YIN[a], Peter HADDAWY[b], Siriwan SUEBNUKARN[c], Phattanapon RHIENMORA[d]

[a,b] *Faculty of ICT, Mahidol University, Thailand*

[c]*Faculty of Dentistry, Thammasat University, Thailand*

[d] *School Science and Technology, Bangkok University, Thailand*

myat.su@student.mahidol.ac.th

Abstract: The traditional apprenticeship approach to dental surgical skill training has known limitations including subjectivity of evaluation, scarcity of available experts, and lack of standardization. As an attempt to address these limitations, dentistry schools have begun to incorporate virtual reality (VR) simulators into surgical curricula. However, automated outcome scoring is not fully supported in existing dental VR simulators. Without automatic outcome analysis, students must still depend on human experts for evaluation of the outcome. With the limited availability of expert supervision, students often end up in unsupervised training with delayed feedback. In this study, we present an approach to automate the process of outcome scoring in dental simulators. Automated outcome scoring is an initial step toward our larger endeavor of automated objective assessment and real-time feedback generation for surgical skill training.

Keywords: Dental surgery simulator, technical skill training, objective assessment, feedback generation, outcome scoring

Introduction

In dentistry, precise psychomotor hand skill is indispensable as it is required in day-to-day routine treatments, and therefore acquisition of the skill is crucial for dental students during their studies. The mainstream approach to dental skill training has combined didactic lectures with a surgical master apprenticeship model. A human expert observes a student's actions throughout a procedure, carries out assessment of the procedure outcome, and provides feedback. Dentistry schools have increasingly been seeking ways to address a number of limitations of this approach to training, including subjectivity of evaluation, scarcity of availability of experts, and lack of standardization. Over the past decade, a variety of computer-based simulation systems have been developed as a way to address these limitations and dentistry schools have begun to incorporate simulators into their curricula.

Among the various types of simulators, Virtual Reality (VR) simulators are becoming popular as they have the ability to record kinematic data on how a user performs each step of a task. A handful of dental VR systems has been developed for both academic and commercial use [1-8].

Typically the assessment of skill consists of the procedural analysis and the assessment of end-product or outcome of the procedure. In dental VR simulators, to the best of our knowledge, only DentSim [5] and Forsslund [2] simulators have the feature of scoring and grading the end-product. The grading module of the Forsslund VR simulator [1] measures the number of voxels cut in each area of the tooth in comparison with the expert to determine the end-product grade. In practice, the amount of tooth mass removed does not explicitly describe the quality of outcome. Using the tracking technology, DentSim captures the kinematic data of the instruments and outcome is evaluated on-line with various criteria. Students are provided with objective assessment and feedback throughout the training session. With mannequin and plastic teeth, Dentsim is suitable for skill practice but less suitable for warm-up presurgery skill practice sessions. Another limitation is the high cost of the product which many educational institutions cannot afford.

Existing work using dental VR simulators in objective assessment and feedback emphasizes analysis of operational performance without considering the quality of the outcome [7]. Without automatic assessment and real-time feedback on outcome, students must still depend on human experts. Students must call the expert dentist to check the correctness of their work upon completion of a training session. With the limited availability of expert supervision, students often end up in unsupervised training with delayed feedback. Evaluation demands dedicated time of experts, and the workload of experts is not lessened by the introduction of VR simulators into training curriculum. These issues can be addressed by incorporating mechanisms to evaluate the outcome into the training simulator.

Much of the work on automated outcome scoring has been done in the context of otology surgical skill training simulators. In evaluating mastoidectomy simulation performance, Sewell et al. [11] build a Naïve Bayes classifier based on estimates for the probability that each voxel is removed by an expert and a novice and use it in skill level classification of users. The classifier was later improved by calculating mutual information for each voxel and they build the classifier using only the 1,000 most discriminative voxels. The regions of bone which are most likely to be removed by experienced surgeons but left by novices and vice versa are obtained by visualizing the most informative voxels and are subsequently used in feedback generation. They later [12] extend the work by integrating various process-based features and incorporating metrics for exposure to critical anatomic structure into scoring. Similarly, Kerwin et al. [13] present an approach to automated scoring of virtual mastoidectomy performance on a voxel level. Firstly, they create a fully partitioned segmented dataset by defining surgically important regions on an iconic temporal bone. Then using earth's mover distance (EMD), parts of an expert-drilled bone are compared with a student-drilled bone. A decision tree is created using the features derived from these comparisons to determine scores of resident surgical performance. By comparing with multiple expert examples, the scores is averaged to provide a reliability metric.

In this study, we present an approach to automate the process of outcome scoring in a dental VR simulator. Our score cube based automated scoring begins with the registration of the template tooth to the training tooth to duplicate the expert outcome to the training tooth. Using registered expert templates, voxels values in the score cube volume are assigned according to their proximity with the templates. With the score cube, automated outcome scoring evaluates student's outcome at fine-grained voxel levels. Automated outcome scoring can be considered as an initial step toward our larger endeavor of automated objective assessment and real-time feedback generation framework for surgical skill training.

1. Dental VR simulator data collection

Figure 1 A dentist performing access opening to the root canals with a haptic virtual reality simulator

The VR simulator used in this study is the Haptic Virtual Reality Simulator of Rhienmora et al. [7]. Two PHANToM Omni (SensAble Inc., Woburn, MA, USA) haptic devices are configured in a dual mode, representing a hand-piece and a dental mouth mirror as show in 

Figure 1.

The tooth model was acquired using three-dimensional micro-CT (RmCT, Rigaku Co., Tokyo, Japan). Tomographic images were obtained using comprehensive dental imaging software (i-VIEW, Morita Co., Tokyo, Japan). Three-dimensional reconstruction was performed using 600 of these two-dimensional images processed by volume rendering. Volumetric tooth data is available in (.Raw) file format where each byte contains its unsigned byte value [0,255], representing the density value of a particular voxel.

With the collaboration of an endodontist in our study, access opening preparation for root canal treatment is chosen as the training procedure. In this preparation phase, the endodontist drills a small access hole through the surface of a tooth crown to gain access to the pulp chamber and root canals for treatment. The expected result of access preparation is to create an unobstructed passageway to the pupal space and the apical portion of the root canal. Successful outcome of the preparation largely depends on the fine motor skill of the dentist. The types of operative errors that could occur in access opening preparation are *undercut*, *overcut* and *perforation*. The *undercut* is the case when the student's drilling does not remove the roof of the pulpal chamber completely. As a result, access to the orifices of the root canals is obscured by the remaining in the roof. The *overcut* issue is the case when the student drills through a large mass of tooth unnecessarily and leaving the weak tooth structure behind. In the worst case, *perforation* can occur during access cavity preparation.

Two essential pieces of information for grading the outcome and providing the feedback are the types of errors and the region of the errors in outcome.

2. Automated outcome scoring

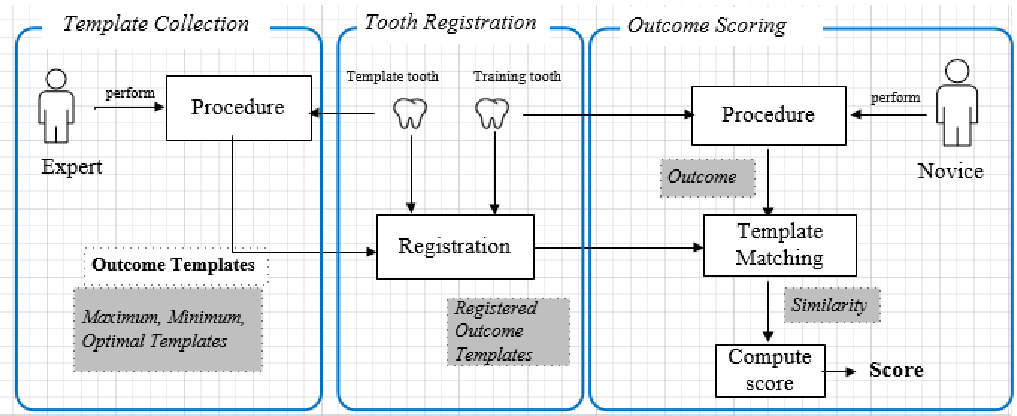


Figure 2 Automated outcome scoring

The aim of the present study is to provide a method to grade the outcome of the student's work automatically using template matching technique. This technique determines the regions of the outcome that deviate from the optimal outcome as well as the degree of deviation.

We define three categories of templates as *Maximum* (*max*), *Minimum* (*min*), and *Optimal (optimal)* templates. The *Max* and *Min* templates represent the maximum and minimum acceptable outcome size ranges for the drilling task while *optimal* template represents the ideal outcome between maximum and minimum thresholds.

The overview of automated outcome scoring is illustrated in Figure 2. Firstly, an expert dentist performs the procedures on a selected tooth (template tooth) to produce a set of canonical outcome templates that are stored in the template repository. The tooth used in the training procedure (training tooth) is from same category as the template tooth but obtained from a different patient.

Secondly, in order to make template matching feasible in a later stage, the morphological features of the template tooth are registered with the training tooth before the training session begins. Outcome templates registration entails the transfer of expert's outcome features from templates to the training tooth. When a novice trainee performs the procedure on the training tooth, the preparation outcome is compared against the previously obtained transformed expert outcomes. The score is determined on the basis of similarity between the trainee outcome and the expert's templates. The details of each step are described the next section.

2.1 Template Collection

In our experiment, the expert performed access opening on the mandibular right second molar and the mandibular right second molar from a different patient is used as a training tooth. The expert followed three stages of tooth preparation: Stage 1, cut into pulp chamber roof; Stage 2, extend the opening to one of the canal orifices; Stage 3, extend the opening to the remaining canal orifices. The expert called out each step number, which was manually entered in the system during tooth preparation. Three outcome templates, *max*, *min* and *optimal* templates are prepared by the expert for each stage. Examples of tooth templates for stage 3 are shown in
Figure 3**.**

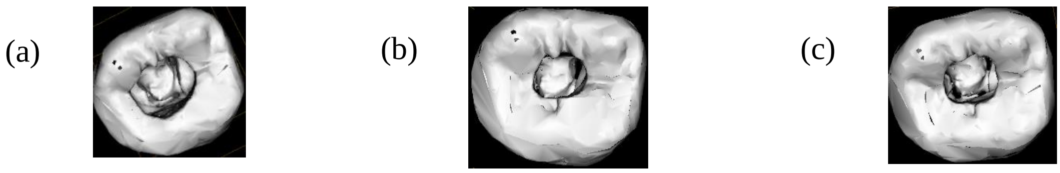


Figure 3 (a) *Max* template (b) *Min* template (c) *Optimal* template for stage 3

2.2 Template tooth registration

Since the template tooth and training tooth may be different, the max, min, and optimal outcome templates must be transformed to be appropriate to the training tooth shape and anatomy. The process of aligning and transforming the template tooth to correspond to the training tooth is called registration and is a common component in any template based technique. From the myriad of registration techniques available, we have selected to use methods based shape context descriptor because they are robust, well suited to registration of irregular objects, and efficient. In our study, we use 3D Shape Context registration implemented by Kroon et al. [10] to register two 3D tooth models.

The 2D Shape context descriptor introduced by Belongie et al. [11] represents a shape as a set of points sampled from the shape's contour. Each sample point is represented by the coarse histogram of the other point surrounding it. The similarity between two shapes is obtained by finding the correspondences between the sampled points in the two shapes. After obtaining the correspondences at sample points, the correspondence to the complete shape is achieved by estimating an aligning transformation that maps one shape onto the other. The shape context feature vector of a point in 2D is a log-polar histogram, represented as vector. The histogram collects the location of all other points relative to the selected point, described by log-distance and angle. For 3D objects, Kroon et al. [10] replace the log-polar histogram by a 3D log-spherical histogram. A point in a 3D space is described by the log of the radius, inclination (polar angle) and azimuth. For each template tooth, we randomly sampled N_s (31290) points from the tooth surface and obtained the shape context histograms.

The training tooth undergoes the identical sampling and histogram computation steps. The optimal correspondence between the template tooth points and the training tooth points is determined by computing the similarity between them. The similarity between two normalized histograms is simply the χ^2 statistic. The optimal correspondence is the permutation of the points minimizing the summed dissimilarity of the matched points. It corresponds to a weighted bipartite graph matching problem and is solved in $O({N_s}^3)$ time using the Hungarian method in the 2D case. But due to the large number of sample points required for irregular 3D objects, the computation time of the Hungarian method is not acceptable. Therefore, Kroon et al. [10] use multiple-to-one connections with Iterative Closest Point (ICP) matching. The matching pairs of shape contexts provide constraints for a B-Spline based Free Form Deformation (FFD) grid used in constructing a transformation field, which warps the points to better align the tooth point clouds. The quality of the final transform is iteratively improved by repeating the correspondence estimation and transform estimation steps, using the results of the previous step as a starting point. In our experiments, we iterate the process two times. Template tooth, training tooth and registered tooth used in our study are presented in Figure 4.

(a) (b) (c)

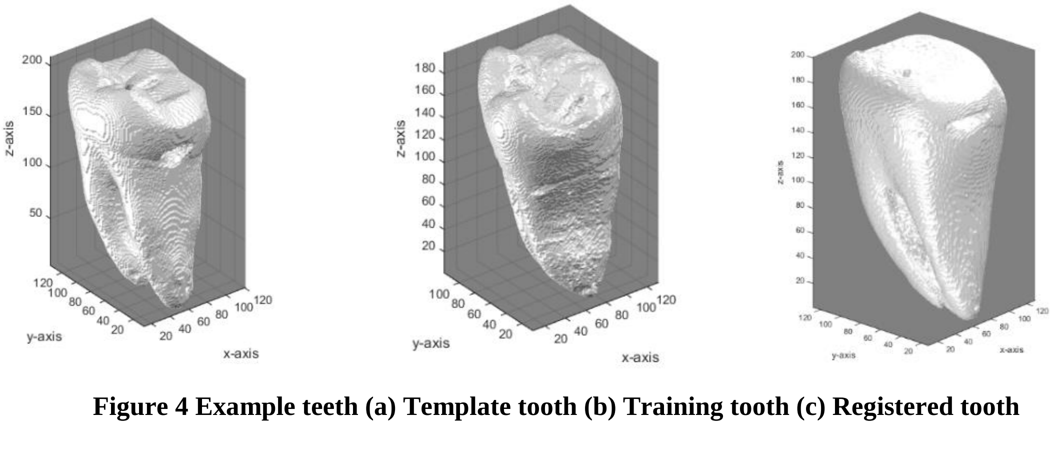


Figure 4 Example teeth (a) Template tooth (b) Training tooth (c) Registered tooth

2.3 Template Matching and Outcome Scoring

Once the expert outcome template is warped into the training tooth, the drilled area on the tooth is extracted by taking the volumetric difference between the original tooth and the drilled tooth. The extracted drilled volumes are used as the reference templates in template matching step. Similarly, the drilled area is extracted from the training tooth once the novice performed the procedure. We identify areas of drilled volume which overlap with areas in predefined templates and assign the score to the areas based on their respective location within each template.

Given a tooth, a wide range of outcome forms is possible. Consequently a robust scoring mechanism is needed to account for a variety of outcome forms. Our approach is to evaluate the voxels in the tooth volume and label them with score points with respect to reference templates. A tooth volume filled with the voxels tagged with their score points describes the geographical mapping of all reference templates on a training tooth. We call such tooth volume a "score cube" and will discuss its construction in detail.

The score cube is 3D volume replica of the training tooth. To fill in the cells of the score cube, we first formulate a voxel scoring function with a score scale ranging between 1 and 3. The value of 1 is given by the scoring function to the voxels which are beyond or below the acceptable maximum - minimum ranges in respective *max* and *min* templates. The voxels in the area of *optimal* template take the value of 3 while the voxels at the maximum and minimum thresholds are assigned with the value of 2. The voxels falling into the area

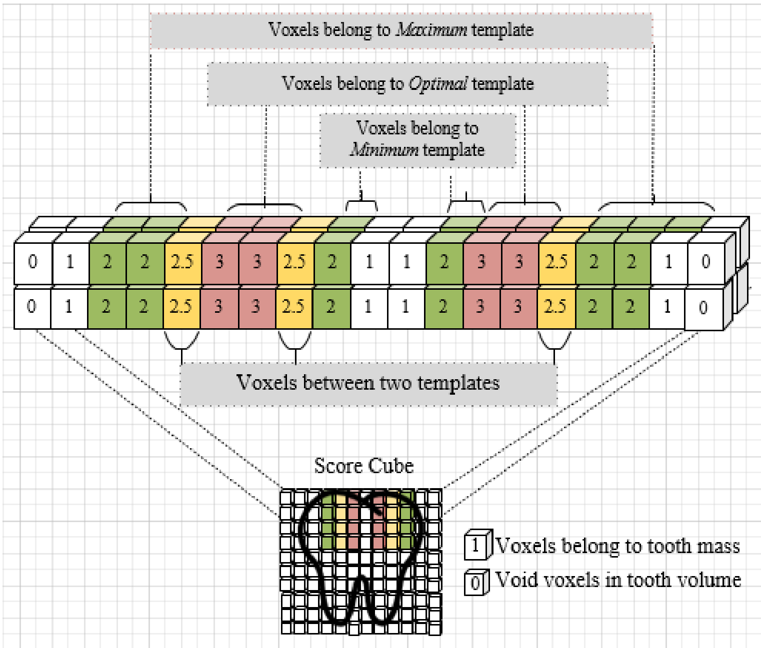


Figure 5 Example score cube

between two templates are scored according to their respective location. That is, the values of the voxels in these areas are determined by their proximities with each template and varies linearly between 2 and 3 based their locations.

Figure 5 shows an example of the score cube.

(a) (b) (c)

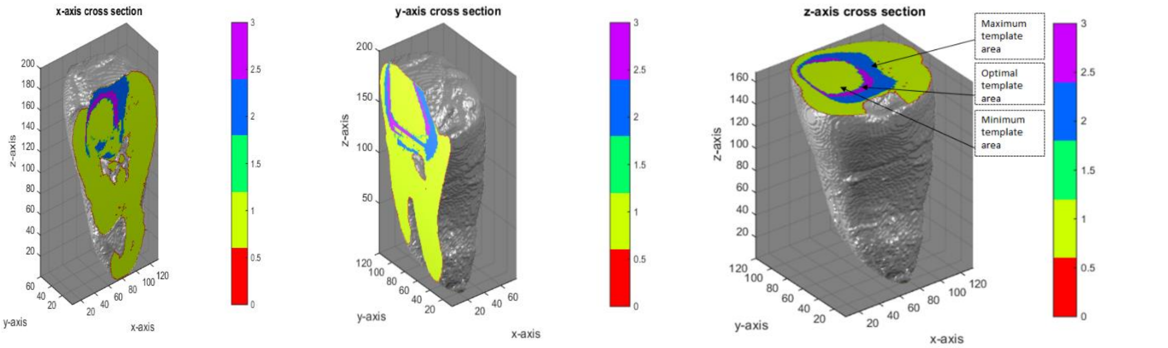
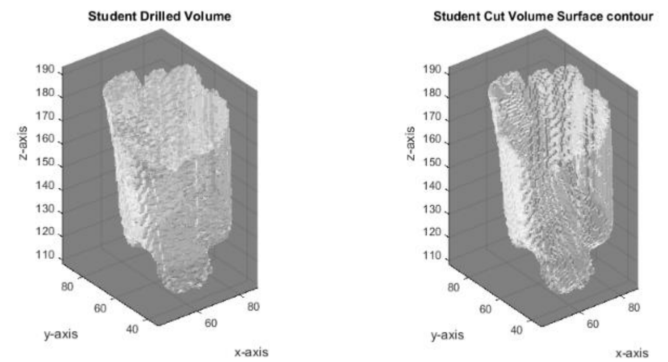


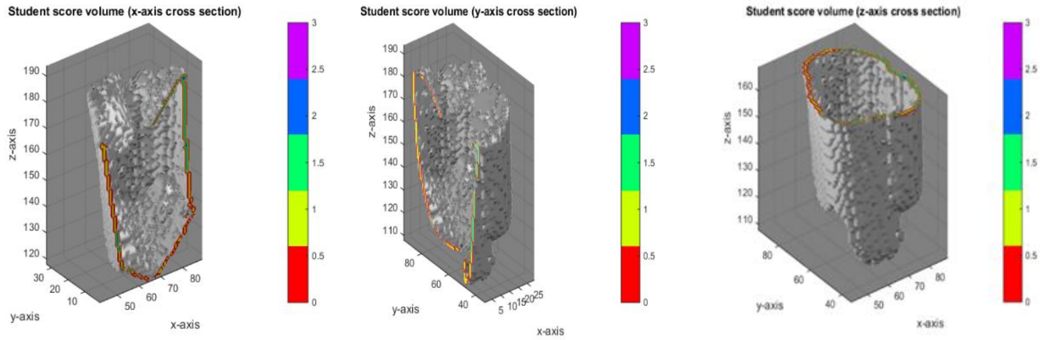
Figure 6 Cross sectional view of an example score cube for step 3 of the procedure (a) taken along x-axis (b) taken along y-axis

Typically, the score cube is a replica tooth volume of the training tooth. The score values of voxels manifested in training tooth relative to reference templates can be verified by taking cross-sectional views as in Figure 6. As shown in Figure 6 (b), score cube explicitly characterize *max*, *min* and *optimal* templates regions and their respective scores on the training tooth. Such spatial information on templates regions and spread of scores provides a visual guidance to the student during training. Students can form a mental image of the perfect drilling area and can plan the drilling route to achieve a quality outcome. Subsequently, such visual map subdues the likelihood of undercut and overcut drilling.

Once the score cube is constructed, it is applied in scoring the procedure outcome. We extract the surface contour from the drilled area of procedure outcome. The contour is then mapped onto the score cube and the voxels on the surface are assigned with the score points accordingly. Figure 7 illustrates the scoring of a student outcome using the score cube. As shown in Figure 7 (c), the color coded regions of describe a detailed information on the areas of shortcomings (low score areas), as well as areas appears satisfactory (averagely scored areas) and areas of perfection (high scored areas) inherently exhibit in a drilling outcome.



a b
c d e **Figure 7 Scoring the student outcome (a) drilled volume (b) surface contour (c) surface contour voxels with assigned scores x-axis view (d)) surface contour voxels with assigned scores y-axis view (e) surface contour voxels with assigned scores z-axis view**



3. Discussion

Automated outcome scoring based on template matching allows the use a variety of training teeth as in real clinical training sessions. As a result, the students can experience a realistic training and assessment setting with VR simulators. Score cube provides detailed insights on types and regions of errors in the student's outcome. Types of errors in outcome are determined from the fine-grained voxel level because, when a small hole is drilled through a tooth surface using the high-speed drilling instrument, a minute mistake in the drilling area have a significant impact on the resulting outcome. With voxel score values derived from the expert templates, the score cube captures any deviation as small as a few millimeter in the student outcome from the expert template. The scoring results also characterize the regions of errors. For example, low scoring areas on the outcome indicate regions of over-drilling or under-drilling, while moderately scoring areas suggest the areas that are acceptable but not optimal yet. This information is instrumental in providing corrective feedback to students. While not applicable to all dental surgical procedures, e.g. tooth extraction, the score cube outcome scoring technique can be applied to a variety of procedures beyond the access opening procedure shown in this paper. The score cube approach is applicable to dental surgery procedures involving the act of milling or drilling or where a prosthetic may need to be positioned and fitted properly. It may also be applicable for some types in orthopedic surgery.

4. Conclusion and Future Work

We have presented an approach to automated outcome scoring in dental surgical simulation. The results from our prototype system are encouraging but much work remains to be done in order to realize the full functionality. The current approach performs scoring based only on the morphological aspect of the drilled areas. Our future work will incorporate the scoring criteria from the context of dental areas, such as, identification of misdirected accesses through crown area (overcut), incomplete opening of pulp chamber roof (undercut), perforation towards the crown and to the floor of the pulp chamber in the outcome and penalize them accordingly. Ultimately, we plan to incorporate the automated outcome scoring technique into an objective assessment and feedback generation framework which is based on correlating procedure and outcome.

5. Acknowledgement

The authors would like to thank anonymous reviewers for their valuable comments that helped improve the manuscript.

References

[1] Corneliu, A., Mihaela, D., Mircea-Dorin, P., Crenguta, B., & Mircea, G. (2011). Teeth reduction dental preparation using virtual and augmented reality by Constanta dental medicine students through the VirDenT system. In The International Conference Development, Energy, Environment, Economics (pp. 21–24). Puerto De La Cruz, Tenerife, Spain.

[2] Forsslund Systems. (2014). Retrieved September 01, 2014, from http://www.forsslundsystems.com/

[3] Onishi, K., Mizushino, K., Ikemoto, H., & Noborio, H. (2013). AR Dental Surgical Simulator Using Haptic Feedback. In Communications in Computer and Information Science (Vol. 374, pp. 202–205).

[4] Wang, D., Zhang, Y., Hou, J., Wang, Y., Lv, P., Chen, Y., & Zhao, H. (2012). IDental: A haptic-based dental simulator and its preliminary user evaluation. IEEE Transactions on Haptics, 5(4), 332–343.

[5] DentSim. (2014). Retrieved from http://www.dentsimlab.com/

[6] Gal, G. Ben, Weiss, E. I., Gafni, N., & Ziv, A. (2011). Preliminary assessment of faculty and student perception of a haptic virtual reality simulator for training dental manual dexterity. Journal of Dental Education, 75(4), 496–504.

[7] Rhienmora, P., Haddawy, P., Suebnukarn, S., & Dailey, M. N. (2011). Intelligent dental training simulator with objective skill assessment and feedback. Artificial Intelligence in Medicine, 52(2), 115–121.
[8] Ullah, F., & Park, K. (2013). Development of a surface-based virtual dental sculpting simulator with multimodal feedback. International Journal of Precision Engineering and Manufacturing, 14(4), 577–587.
[9] Kroon, D.-J. (2011). Segmentation of the Mandibular Canal in Cone-Beam CT Data. Enschede, The Netherlands.
[10] Belongie, S., Malik, J., & Puzicha, J. (2002). Shape Matching and Object Recognition Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 509–522. doi:10.1.1.18.8852
[11] Sewell, C. (2007). Automatic Performance Evaluation in Surgical Simulation. Standford University.
[12] Sewell, C., Morris, D., Blevins, N., Dutta, S., Agrawal, S., Barbagli, F., & Salisbury, K. (2008). Providing metrics and performance feedback in a surgical simulator. Computer Aided Surgery, 13(2), 63–81.
[13] Kerwin, T., Wiet, G., Stredney, D., & Shen, H.-W. (2011). Automatic scoring of virtual mastoidectomies using expert examples. International Journal of Computer Assisted Radiology and Surgery, 7(1), 1–11.