Transfer learning of a temporal bone performance model via anatomical feature registration

Yun Zhou*, Ioanna Ioannou†, Sudanthi Wijewickrema†, James Bailey*,
Patorn Piromchai†, Gregor Kennedy‡, Stephen O'Leary†
* Department of Computing and Information Systems, yuzhou@student.unimelb.edu.au, baileyj@unimelb.edu.au
† Department of Otolaryngology, {ioannoui@, sudanthi.wijewickrema@, sjoleary@, patornp@student.}unimelb.edu.au
‡ Centre for the Study of Higher Education, gek@unimelb.edu.au
University of Melbourne

***Abstract*—Evaluation of the outcome (end-product) of surgical procedures carried out in virtual reality environments is an essential part of simulation-based surgical training. Automated end-product assessment can be carried out by performance classifiers built from a set of expert performances. When applied to temporal bone surgery simulation, these classifiers can evaluate performance on the bone specimen they were trained on, but they cannot be extended to new specimens. Thus, new expert performances need to be recorded for each new specimen, requiring considerable time commitment from time-poor expert surgeons. To eliminate this need, we propose a transfer learning framework to adapt a classifier built on a single temporal bone specimen to multiple specimens. Once a classifier is trained, we translate each new specimens' features to the original feature space, which allows us to carry out performance evaluation on different specimens using the same classifier.**

In our experiment, we built a surgical end-product performance classifier from 16 expert trials on a simulated temporal bone specimen. We applied the transfer learning approach to 8 new specimens to obtain machine generated end-products. We also collected end-products for these 8 specimens drilled by a single expert. We then compared the machine generated end-products to those drilled by the expert. The drilled regions generated by transfer learning were similar to those drilled by the expert.

***Keywords*—transfer learning, anatomy registration, automatic evaluation**

I. Introduction

Providing feedback based on the outcome (end-product) of a surgical procedure is an essential part of surgical training, which allows trainees to develop an understanding of what constitutes good performance [1]. In the discipline of Otolaryngology, surgical trainees develop this understanding by practising on cadavers or patients under the supervision of expert surgeons who evaluate their performance. However, there are limitations to the current training practices, which include a shortage of cadaveric temporal bones, limited availability of expert supervision, and the subjective manner of surgical skill assessment. These challenges have prompted increasing interest in the use of computer-based virtual reality (VR) simulators for surgical education [2]–[4].

VR simulators for surgery can offer repeated practice on multiple surgical cases of varying difficulty at the convenience of trainees. In such simulation systems, machine learning can be used to build a reference model of expert performance from sets of recorded expert surgeries. This reference model can be used to provide unbiased, objective and automated surgical performance evaluation. Sewell et al. and Kerwin et al. achieved reliable results when using such models to evaluate surgical performance in VR temporal bone simulators [5], [6]. However, a limitation of these models is that they can only be used to evaluate performance on the specimen (i.e. surgical case) on which the training data was collected. In order to train a reliable evaluation model for new specimens, the classical machine learning approaches require a new set of expert examples collected from each specimen. Given the fact that expert surgeons have very full schedules, such data can be difficult to obtain for a large number of specimens. To our best knowledge, little work has been done to address the problem of providing automated objective performance evaluation on multiple specimens within surgical simulators.

When a human surgeon attains some knowledge by operating on a patient, they have the ability to adapt that knowledge to other patients. Similarly, the knowledge represented in a performance evaluation model learnt from one bone specimen should be transferable to other temporal bones. The "transfer learning" [7] approach from the field of machine learning is a good way to deal with such problems. A popular transfer learning method is to model the differences between domains in order to transfer a model learnt from the source domain to the target domain. In this work we apply this principle to surgical simulation, by capturing the differences between specimens and using them to extend our evaluation model from the original specimen to new specimens.

Specifically we propose a transfer learning approach to adapt a cochlear implantation surgery performance evaluation model trained on a single temporal bone specimen to multiple new specimens. First, we build a classifier using a set of expert trials on the original specimen, which we refer to as "specimen model". Then for each new specimen, we register each anatomical structure to the corresponding structure in the specimen model. Third, we transform each bone voxel[1] of the new specimen to the specimen model position according to the registration matrix of the nearest anatomical structure. This transformation allows the classifier to be used in predicting

[1]Specimens are represented in volumetric 3D grids, where a voxel represents a grid element in 3D space. This is analogous to a pixel in a 2D bitmap image.

the region that should be drilled in the new specimen. Since anatomical registration is only approximate, it will introduce some errors. To correct part of this error, the predicted drilled region is adjusted by applying surgical domain knowledge. In summary, this paper makes the following contributions:

- To the best of our knowledge, this is the first formal study in the use of anatomical registration to adapt a classifier to different specimens in a surgical simulation environment.
- The proposed algorithms are designed to consider both specimen alignment and surgical domain knowledge.
- Experimental evaluation shows that the proposed adaptation generate a high quality end-product on a set of different specimens.

Section II discusses related work. Section III explains the concept of transfer learning in general and provides a high level overview of how this can be applied to adapt a classifier for use on multiple specimens. Section III-A gives an overview of the proposed transfer learning framework, while sections III-B and III-C explain the two major components of our solution: using anatomical registration to align different specimens, and adding surgical domain knowledge constraints to adjust the decision model. Section IV describes our experiment and results, and section V concludes the paper.

II. Related Work

Automated performance evaluation within VR surgical simulators has drawn increasing attention as a crucial component of simulation-based training. Much of the existing literature has focused on evaluating surgical motions, such as hand movements, tool usage and applied force/torque [8]–[10]. In our previous work, we used surgical technique evaluation to generate meaningful automated real-time feedback in a temporal bone surgery simulator [11], [12]. The models are used primarily in evaluating *how* a trainee performed the surgery, but not whether they drilled the correct regions. While technique evaluation is important, it is equally important to evaluate the outcome of a surgical task. Therefore, other types of automated evaluation such as surgical end-product assessment need to be integrated into evaluation systems along with motion-based evaluation.

Sewell et al. and Kerwin et al. have carried out some work towards automated evaluation of simulated temporal bone surgery end-products [5], [6]. Sewell et al. [5] used the 1000 most informative voxels from a virtual temporal bone in building a Naive Bayes model to evaluate expertise. On the other hand, instead of using voxels as features, Kerwin et al. [6] derived a set of distances from each anatomical structure and used a decision tree to evaluate surgical performance. These approaches achieved high accuracy, but one major drawback is that such classifiers can only evaluate performance on a specific specimen, since a major assumption in the classifier algorithms is that the training and testing data should be in the same feature space and have the same distribution.

One of the benefits of simulation-based training is that it can - and should - expose novice surgeons to a variety of different surgical cases in order to build a complete skill set. Therefore, a need exists for surgical end-product assessment algorithms that can be applied to multiple specimens. In view of the limitations of the existing methods discussed above, we propose the use of transfer learning [7] to adapt a classifier trained from one specimen for use on multiple specimens.

Transfer learning [13], [14] has been studied widely in recent years, since the assumption that training and testing data have the same feature space and distribution does not hold in many real-world applications. For example, a classifier that was trained to distinguish between foxes and wolves is unlikely to perform well in differentiating lions and tigers, since the two domains are characterised by different features. To address this problem, transfer learning transforms different domains into a common feature space such that a classifier trained in one domain can be used in another domain. This approach has been used widely in solving text and image classification problems. A detailed survey of transfer learning can be found in [7].

III. End-product Adaptation

The principles of transfer learning can be applied to VR surgical simulation. Let us treat each specimen as a domain. If an end-product classifier built from one specimen could be successfully transferred to other specimens, it would save a lot of time and effort in collecting expert examples on other specimens. Our approach was inspired by Blitzer et al. [13], who found the correspondence between the vocabulary of two different text corpora via pivot words that occur frequently in both text corpora. Although this method works well for transferring a text classification model, it cannot be applied directly to surgical simulation due to several differences between the two fields. One specimen usually contains more than a million unique voxel positions while a single document usually contains less words. A particular word is likely to have the same or similar meaning when it appears in different texts, while a particular voxel position may represent different anatomical structures in different specimens.

To apply the approach described in [13] to temporal bone specimens, we needed to choose an appropriate set of features to use as the "pivot component" in finding the correspondence between two specimens. The anatomical landmarks of each specimen were chosen as a suitable pivot component, since they have the same meaning across specimens. This idea has been widely used in the registration of MRI images [15]. First, we aligned the anatomical structures of the specimen model with those of each new specimen using the iterative closest points (ICP) algorithm [16]. Based on this alignment of anatomical structures, we aligned each voxel of the new specimen to that of the specimen model using the nearest anatomical structures. This transfer enabled the classifier learnt on the specimen model to predict the end-product on the new specimens. Since this transfer is not globally optimal, we added a post-processing step using surgical domain knowledge constraints to refine the end-product.

A. Overview

Figure 1 provides an overview of the proposed transfer learning framework. This framework is composed of two steps.

1) Training of a drilled region classifier: A drilled region classifier is built from a set of expert simulator runs on the

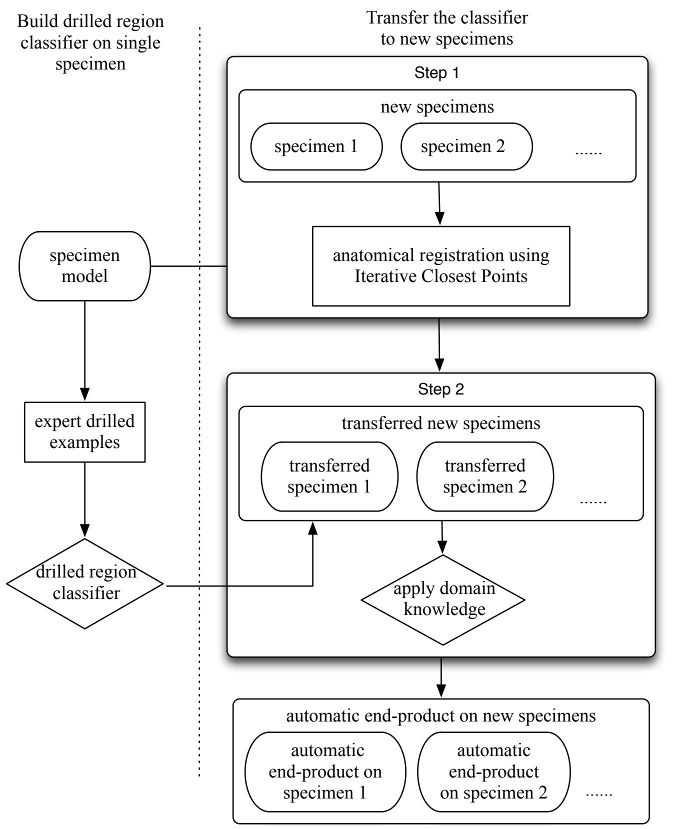

Fig. 1. Overview of the framework used to adapt a drilled region classifier to different specimens using transfer learning

reference specimen, which is referred to as the specimen model. The feature we use to train the classifier is the position of each voxel in specimen coordinates.

2) Transfer of classifier to different specimens: Each new specimen is registered with the specimen model using the ICP algorithm and the classifier is used to predict the drilled region on the new specimen. Then domain knowledge is applied to adjust the predicted drill region.

Figure 2 illustrates the 9 anatomical landmarks that human experts use to complete temporal bone surgery safely and effectively. In our study, the chosen surgical procedure consisted of a cortical mastoidectomy (removal of bone surrounding anatomical structures A to D shown in Figure 2), followed by posterior tympanotomy (removal of most of the bone between the facial nerve and the chorda tympani nerve denoted as D), and cochleostomy (drilling a small hole next to the round window (H) leading into the basal turn of the cochlea (G)). This is a considered a complex procedure that is carried out as part of cochlear implant surgery.

B. Anatomical Registration

As discussed in section II, directly applying a classifier built from one specimen to new cases is unlikely to be successful, since features differ between specimens. An intuitive transfer learning approach to tackle this problem is to align different specimens into a common feature space in order to train a generalized classifier. However, each specimen

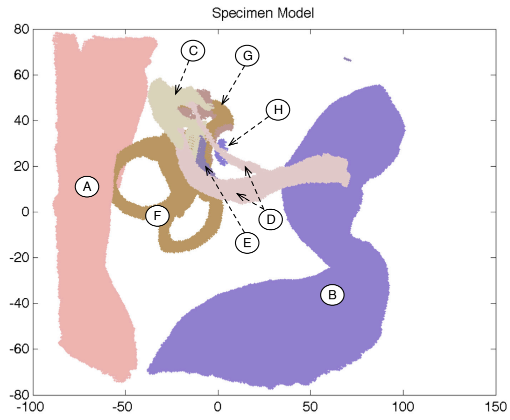

Fig. 2. Major anatomical landmarks of the temporal bone used as the specimen model. A: dura, B: sigmoid sinus, C: ossicles, D: facial nerve and chorda tympani nerve, E: stapedius tendon, F: semi-circular canals, G: basilar membrane H: round window membrane.

is quite distinct in terms of shape, scale and orientation, therefore deriving a common feature space based on raw voxel information is infeasible. On the other hand, we observe that each specimen contains a set of anatomical structures used by surgeons as landmarks to identify which regions of bone should be removed. Therefore, a common feature space derived from these landmark structures can be generalized to predict the drilled regions on new temporal bones. So we chose one specimen as our specimen model and aligned the anatomical landmarks of the specimen model to the rest of the specimens using the Iterative Closest Point(ICP) algorithm [16]. This algorithm is one of the main tools used in 3D shape registration. A detailed comparison of different improvements can be found in [17]. In our case, we use the following objective function to measure the alignment of two anatomical structures.

$$f(SM(a), S(a)) = \frac{\sum_{i=1}^{N_S(a)} ||SM(a)_i - TM(a) \times S(a)||}{N_S(a)} \tag{1}$$

$SM(a)$ is the anatomical structure voxel set of the specimen model, $S(a)$ is the anatomical structure voxel set of a new specimen, $N_S(a)$ denotes the number of voxels belonging to anatomical structure a, TM is the transformation matrix from the new specimen to the specimen model. An optimal transformation matrix TM is derived by minimizing objective function f. Figure 3 illustrates the accuracy of anatomical registration using one of the new specimens as an example.

Once we derive transformation matrix TM for each anatomical landmark, we assume that the bone voxels near each landmark follow the same pattern of transformation; therefore, we apply the TM of the nearest anatomical landmark to transfer each bone voxel from the new specimen into the feature space of the specimen model. Finally, the original drilled region classifier can be applied on the transferred bone voxels to predict which voxels need to be removed. This algorithm is described further in Figure 4.

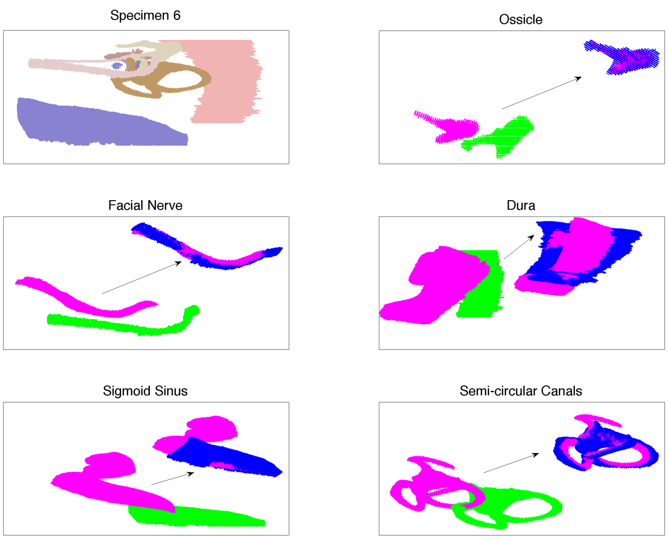

Fig. 3. Anatomical registration example using specimen 6. The first graph shows the anatomical structure of specimen 6. In each of following graph, the two sets of magenta points represent the anatomical structure voxels in the specimen model. The set of green points represent the corresponding anatomical structure in specimen 6 before registration. The blue points represent the anatomical structure from specimen 6 after registration.

Input: Trained Classifier c, Source Specimen ss, Target Specimen ts

```
tm = [] {initialize transfer matrix list}
for lstructureID a in ss do
    tm[a] = ICP(ss.structure[a], ts.structure[a])
end for
tts = [] {initialize transferred target specimen}
for bone voxel v in ts do
    a = findClosestStructure(v, ts.structure)
    tts[v] = tranferVoxel(v, tm[a])
end for
r = predictDrillRegion(c, tts)
return r
```

Fig. 4. Transfer learning of drilled bone region

C. Application of Domain Knowledge Constraints

Since we use anatomical structure registration to adapt a drilled region classifier to new specimens, the geometric relationship between landmarks is being transferred approximately. A drawback of this approximation is that the further a voxel is from its closest anatomical landmark, the less accurate its transfer will be. For example, the transfer of voxels deep inside the bone is more reliable due to proximity to many anatomical structures, compared to that of voxels on the surface of the specimen which are far from any landmark. Therefore, it may be necessary to apply post-processing based on surgical domain knowledge to the bone voxels near the surface, to improve upon the end-product approximation provided by transfer learning. One simple but useful constraint is that if a particular bone voxel is drilled, then all voxels covering it must have also been drilled, because it is not possible to remove a voxel that is not visible. Equation 2 shows how this constraint is applied to a surgical end-product $D(v)$.

$$D'(v) = \begin{cases} 1 & v.z \geq u.z \text{ and } ||v, u|| \leq 1 \text{ and } D(u) = 1 \\ D(v) & \text{otherwise} \end{cases} \tag{2}$$

v and u represent voxel absolute positions, D represents the end-product derived by transfer learning while D' is the end-product after post-processing. The *if* condition checks whether a voxel v is covering a drilled voxel u. If a voxel is assigned a value of 1, it is considered drilled.

IV. Evaluation

We evaluate the quality of the transfer models by comparing the results of the machine generated end-products to those drilled by a human expert. To perform the comparison, a quantity that measures the similarity between two drilled regions should be defined. To this end, Sewell et al. [5] applied information gain to select a set of "most significant voxels" drilled by experts, and used the presence or absence of these voxels in the drilled region to evaluate the quality of an end-product. In our analysis, instead of considering the binary presence of significant voxels, we use a weighted euclidean distance to measure the difference between two end-products A and B. We separate this weighted distance into two categories: $D1$: the sum of weighted euclidean distance between voxels that have only been drilled in A to the closest voxel drilled in B; $D2$: the sum of weighted euclidean distance from voxels drilled only in B to the closest voxel drilled in A.

Let us suppose that end-product A was drilled by a human expert while end-product B was obtained from a transferred model. We calculate the weight for each voxel drilled by the human expert in A based on the time at which it was removed. Voxels drilled at the end of the procedure are assumed to be more important than those closer to the start, as they define the path for the insertion of the cochlear implant. However, when calculating weights for $D2$, our model is unable to predict when a voxel should be drilled. Therefore, we find the nearest voxel v' drilled by the human expert in A and use its time stamp to compute the weight for $D2$.

$D1$ and $D2$ are defined in Equations 3 and 4. $||*||$ denotes the euclidean distance between two voxels, v is a voxel of a drilled specimen, the nearest voxel v' drilled by the human expert in A, and $time(*)$ is the time at which a voxel was drilled by the human expert. The similarity measure between end products A and B is defined as the sum of $D1$ and $D2$. A good prediction model should result in a lower weighted distance, meaning the end-product derived by the model is similar to that drilled by a human expert.

$$D1(A, B) = \sum_v \frac{time(A[v]) \times \min(||A[v], B||)}{\max(time(A))} \tag{3}$$

$$D2(A, B) = \sum_v \frac{time(A[v']) \times \min(||B[v], A||)}{\max(time(A))} \tag{4}$$

A. Data Collection

We collected 16 simulation trials conducted by 7 expert otologists performing the surgical procedure on a left ear specimen (referred to here as the specimen model). From this dataset, we trained different classifiers (Logistic Regression, Naive Bayes, Decision Trees) to predict the region drilled by experts on the specimen model. Decision trees achieved the best cross validation accuracy with 93.95%. This shows that drilled voxels are an accurate predictor of expertise in this surgical procedure.

To evaluate the performance of our transfer learning method, we used 8 new specimens (3 left ears and 5 right ears). These differed to the specimen model in many ways, such as orientation, size, the extent to which each anatomical landmark was segmented, and the distance between landmarks. We obtained a surgical end-product conducted by an expert on each of these new specimens, to serve as our standard reference.

B. Baseline

As discussed above, a human expert provided an end-product on each of the new specimens to be used as a standard for comparison. We then evaluated the accuracy of the transfer models with respect to this standard. We used three models in this comparison: 1) A classifier model without transfer learning where the original decision tree model was used to generate an end-product for a new specimen (referred to here as ‘D: Direct decision tree model’), 2) a classifier built on the absolute positions of the voxels (x, y, z coordinates) of each anatomical structure (referred to here as ‘T: Transfer learning model using absolute positions’), and 3) a classifier built on the distances to anatomical structures as features (referred to here as ‘R: Transfer learning model using relative distances’), where the distances of each bone voxel to anatomical landmarks rather than its absolute position are used as features for the decision tree model. To be more specific, for each bone voxel of a specimen, we derived the closest distance to each of the 11 anatomical landmarks. Since these distances are relative features, they should be intrinsically transferable between different specimens.

In addition to baselines D and R, we derived a lower bound corresponding to the weighted distances between the 16 expert performances on the specimen model to illustrate the amount of variation in the end-product produced by different experts due to different surgical styles, even when operating on the same specimen. From each expert performance, we computed the weighted distances according to equations 3 and 4 to the other 15 performances. We derived the mean value and standard deviation of these weighted distances. The lower bound of the weighted distance mean plus two standard deviations is illustrated as a dashed line in Figure 5.

C. Experiment Results

As shown in Figure 5, methods D and R performed worse than the proposed method T for the majority of specimens. The proposed transfer learning method T achieved the best end-product and was significantly better than baseline D [$t(7)=3.314$, $p=0.013$]. It was also significantly better than the relative distance method R [$t(7)=2.646$, $p=0.033$].

It is unsurprising that D performed worse, as there were significant differences between the test specimens and the specimen model on which the decision tree was built. The relative distance method R did not achieve significant improvement compared to D, especially for specimens 1, 6, and 8, where it performed considerably worse than method D. There are two possible reasons for this: 1) The relative distance features did not model the expert end-product precisely, as two voxels at different locations can have the same closest distance to an anatomical landmark. In such a case, a systemic mistake is introduced. 2) Due to segmentation limitations, different specimens may have different portions of the anatomical structures segmented. For example, specimen 6 only has the bottom part of the sigmoid sinus while the specimen model has more of this structure. This introduces an error when calculating the relative distance of a voxel to anatomical structures.

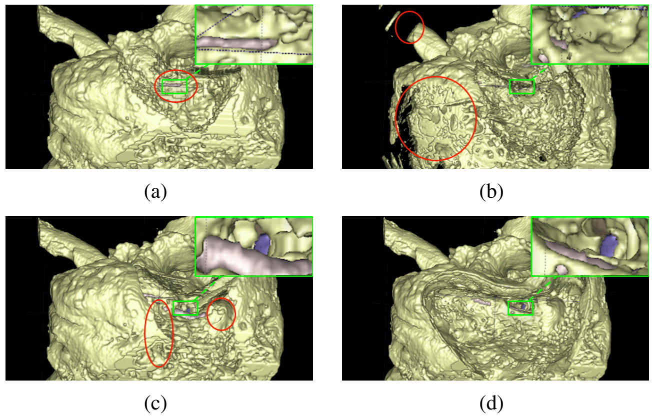

Fig. 6. End-products for specimen 6 produced by different machine learning approaches and a human expert. The green square indicates the facial recess triangle which must be drilled during the procedure. This region is magnified at the top right corner at each sub-figure. The purple/blue round window is not exposed in (a) and (b) but is exposed in (c) and (d). The red circle indicates a major error in each machine-generated end product, either drilling too much or not enough. (a): Directly applied decision tree trained from original specimen. (b): Transfer learning model using relative distances. (c): Transfer learning model using anatomical registration. (d): Human expert.

To illustrate the benefits of the anatomical registration transfer learning approach, the end-products for specimen 6 using the methods discussed above approaches are shown in Figure 6. The end-product generated by the directly applied decision tree(D) did not remove the bone (rendered in yellow) between the facial nerve and chorda tympani, which is necessary to expose the round window and complete the cochleostomy. Therefore from a surgical point of view, this end-product failed to complete the procedure. The end product generated by the transfer learning model using relative distances(R) revealed a small part of the round window (rendered in purple/blue), which is not large enough to perform the next step of cochlear implantation. In addition, this model removed many areas of bone which did not need to be drilled, such as that on top of the sigmoid sinus and the areas on the left boundary of the sigmoid. The end-product produced by the transfer learning model using anatomical registration(T) successfully exposed the round window, and was most similar to the end-product conducted by a human expert in the latter parts of the procedure. The only drawback is that bone regions drilled in the early stages of procedure were not opened as widely as the human end-product.

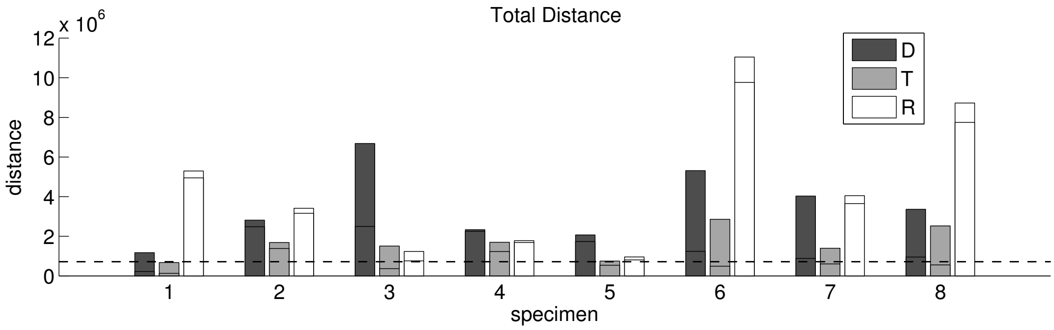

Fig. 5. Comparison of weighted distance for different methods. D: Direct decision tree model, T: Transfer learning model using anatomical registration, R: Transfer learning model using relative distances. Each bar is composed of $D1$ and $D2$. The dashed line denotes a lower bound corresponding to the weighted distance derived from expert end-products on the specimen model (representing variability between experts).

In order to analyse the errors of the machine-generated end-products, we consider the two types of weighted distances $D1$ and $D2$ separately. A large $D1$ implies that the model is too aggressive and has drilled too many voxels when compared to regions drilled by the human expert. In contrast, a large $D2$ occurs when the model has drilled too few voxels. The most significant error observed in the transfer learning model using anatomical registration was $D2$. Part of this error could be attributed to the fact that not all experts opened up the specimen model in the same way as the human expert who provided the reference data for the new specimens. Hence, the decision tree model, when transferred to the test specimens is concentrated around the central regions which were drilled by most experts. Despite this, the drilled regions derived from the proposed anatomical registration transfer learning method provide a good guideline for where the drilling should be done in different specimens, particularly in later parts of the procedure.

V. Conclusion

Automatic surgical end product evaluation has been extensively studied in different types of open surgery. However, the adaptation of classifiers trained on a single specimen to different specimens has not been widely researched. In this work we introduced a simple algorithm for classifier adaptation based on anatomical structure alignment and domain knowledge constraints. We combined anatomical registration with a nearest neighbour technique to transfer new specimens to the feature space of the original specimen. Then we applied domain knowledge constraints by adjusting the transferred classifier gradually moving from the deepest layer to the surface. It was observed in an experimental evaluation that the proposed method was able to identify the drilled regions in different specimens accurately and the end-products produced were similar to those drilled by a human expert. One limitation of the proposed method is that regions of bone drilled in the early stages of the procedure were not opened as widely as the human end-product. This surgical principle should be encoded as domain knowledge constraints to improve the quality of the end-product in the future.

References

[1] K. A. Ericsson, "Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains," *Acad Med*, vol. 79, no. 10, pp. S70–S81, 2004.

[2] M. Agus, A. Giachetti, E. Gobbetti, G. Zanetti, and A. Zorcolo, "Real-time haptic and visual simulation of bone dissection," *Presence-Teleop Virt*, vol. 12, no. 1, pp. 110–122, 2003.

[3] J. Bryan, D. Stredney, G. Wiet, and D. Sessanna, "Virtual temporal bone dissection: a case study," in *Proc. Conf. Visualization'01*. IEEE Computer Society, 2001, pp. 497–500.

[4] T. Kerwin, H.-W. Shen, and D. Stredney, "Enhancing realism of wet surfaces in temporal bone surgical simulation," *IEEE T Vis Comput Gr*, vol. 15, no. 5, pp. 747–758, 2009.

[5] C. Sewell, D. Morris, N. Blevins, F. Barbagli, and K. Salisbury, "Evaluating drilling and suctioning technique in a mastoidectomy simulator," *St Heal T*, vol. 125, pp. 427–432, 2007.

[6] T. Kerwin, G. Wiet, D. Stredney, and H.-W. Shen, "Automatic scoring of virtual mastoidectomies using expert examples," *IJCARS*, vol. 7, no. 1, pp. 1–11, 2012.

[7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[8] H. Lin, I. Shafran, D. Yuh, and G. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Comput Aided Surg*, vol. 11, no. 5, pp. 220–230, 2006.

[9] G. Forestier, F. Lalys, L. Riffaud, B. Trelhu, and P. Jannin, "Classification of surgical processes using dynamic time warping," *J Biomed Inform*, vol. 45, no. 2, pp. 255–264, 2012.

[10] J. Rosen, B. Hannaford, C. G. Richards, and M. N. Sinanan, "Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills," *IEEE T Bio-Med Eng*, vol. 48, pp. 579–591, 2001.

[11] Y. Zhou, J. Bailey, I. Ioannou, S. Wijewickrema, G. Kennedy, and S. O'Leary, "Constructive real time feedback for a temporal bone simulator," in *Proc. MICCAI'13*, 2013.

[12] Y. Zhou, J. Bailey, I. Ioannou, S. Wijewickrema, S. O'Leary, and G. Kennedy, "Pattern-based real-time feedback for a temporal bone simulator." in *Proc. VRST'13*, 2013.

[13] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc.Conf. Empirical Methods in Natural Language Processing'06*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 120–128.

[14] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," in *Proc. of IJCAI'09*, 2009.

[15] J. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical image analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[16] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, 1992.

[17] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *Proc. 3DIM'01*, 2001.