BULGED EARDRUM DETECTION FROM 3D DATA

Manuel Martinello[1], *Leonidas Spinoulas*[1], *Sofia Karygianni*[2], *Pascal Frossard*[2],
Mary Ann Haralam[3], *Timothy R. Shope*[3], *Nader Shaikh*[3], *Alejandro Hoberman*[3], *Ivana Tošić*[1]

[1]Ricoh Innovations Corp. (USA), [2]EPFL (Switzerland), [3]Children's Hospital of Pittsburgh (USA)

ABSTRACT

Bulging is a medical characteristic of the eardrum that is crucial for the diagnosis of acute otitis media. This work proposes a novel classification method for distinguishing bulged eardrums from non-bulged ones. The method uses novel key features extracted from 3D data of the tympanic membrane, captured using a new type of otoscope, the light-field otoscope, capable of non-invasive 3D imaging of the middle ear. We first introduce a variety of geometrical and statistical descriptors (based on *isocontours*), and then select the most discriminative ones. Results on clinical data show that, when using the proposed feature descriptors, eardrum bulging can be automatically detected with an average accuracy of approximately 82%.

***Index Terms*—** medical imaging, classification, 3D data, feature extraction.

1. INTRODUCTION

The 3D shape and the position of the eardrum is the most important factor for recognizing middle ear infection, called acute otitis media (AOM) which is a bacterial infection of the middle ear that requires antimicrobial treatment. Distinguishing AOM from otitis media with effusion (OME) is very important, as wrong diagnosis - especially in children - often leads to overprescription of antibiotics, which can result in the emergence of bacterial resistance. Diagnostic accuracy of clinicians viewing perfect images captured from digital endoscopes is at most 75% [1, 2]. Past studies have shown that the most important sign of AOM, which is absent in OME, is *bulging* [3]. Therefore, an accurate and automated diagnostic algorithm that can detect bulging of the eardrum is essential.

Most of the work in this field is based on analyzing two-dimensional (2D) data captured by traditional otoscopes and otoendoscopes. Instead, in this paper we present for the first time an algorithm for automated classification of eardrum shape that uses three-dimensional (3D) input data. We captured our data in a clinic by using a light-field otoscope, such as the one described in [4, 5]; however, the same type of data could be produced using different 3D sensing devices for eardrum imaging. Our classification task was focused on detecting an eardrum affected by AOM, and therefore distinguishing bulged versus non-bulged eardrum (see Fig. 1). We show that by using only shape information we can achieve an average accuracy of 82%.

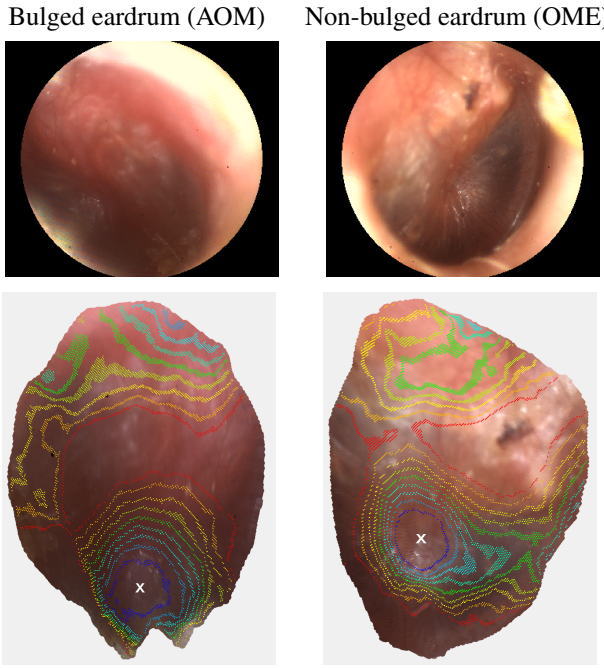


Fig. 1: **Top:** images of eardrums captured with a light-field otoscope; **Bottom:** Isocontours superimposed on registered and normalized eardrums (the white X indicates the centre of the eardrum).

1.1. Prior work

The inclusion of depth features into the classification pipeline is proving quite effective in several applications like object recognition [6, 7] and scene labelling [8]. Incorporationg the depth information into the eardrum classification task is not straigntforward as this task has very specific characteristics that are not taken into account in existing approaches. Traditional approaches in the field of 3D object recognition [9, 10, 11, 12] create local descriptors on points of interest that are then matched to find correspondences between different object instances. Such bag-of-features approaches however, are not suitable for our problem as the eardrums under different medical conditions have very specific differences, e.g., bulging, that can only be detected if we take into account the relative position of the different eardrum features.

Fig. 2: Steps performed to register the eardrum.

Recently a lot of effort has gone into using deep learning for training classifiers [6, 13] to recognize different objects from their RGB-D representation. While very successful at times, these networks have many parameters that require numerous training examples, which are not available in our case since the device is unique and collecting data in clinic requires substantial effort and expense.

So far, the literature on automatic classification of otitis media is limited to 2D images obtained with traditional otoendoscopes or video-otoscopes [2, 14]. These methods try to infer shape properties from 2D images (from shading or image gradients); the outcome can be erroneous and depend heavily on prior data. In contrast, the 3D information obtained with the LFO is estimated (with sub-millimeter accuracy) using the multi-view geometry from light field data [4].

2. INPUT DATA FORMATTING

First, we explain the 3D data that serves as input to our method.

2.1. 3D data representation and registration

For every eardrum we consider, we use the procedure described in [15] to obtain a 3D point cloud formed by N vertices, $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$, where each vertex is defined as a 3D vector:

$$\mathbf{v}_i = [x_i, y_i, z_i]^T, \quad \text{with } i = 1, \ldots, N \text{ and } N \in \mathbb{Z}^+. \quad (1)$$

Images of different eardrums are often captured at different viewpoints (see the two examples at the top of Fig. 1); hence, the 3D point cloud of each eardrum needs to be registered depending on its relative position to the camera, using the following three steps (also summarized in Fig. 2, and described in details in [16]):

1. Out-of-plane rotation (remove the *tilt* of the eardrum)
2. In-plane rotation (normalize the camera's *orientation*)
3. Identification of the center of the eardrum.

After registration, the points defined by the first two coordinates (x and y) lie on the tilted plane, while the third coordinate (z) represents the distance from that plane (depth). Fig. 1 shows two examples of registered eardrums whose estimated centers have been indicated with white crosses.

2.2. Isocontours

Given a 3D point cloud formed by a set of vertices V, *isocontours* are defined as line segments at a certain height (certain z value for the third dimension of each vertex). Isocontours are calculated by resampling the 3D point cloud on a regular grid and using the Marching Squares algorithm on the resampled mesh [17]. Alternatively, the Meandering Triangles algorithm could be used directly on the 3D point cloud without resampling, after applying triangulation [17].

Table 1: Enumeration of extracted descriptors from isocontours.

ID	Descriptor Name	Descriptor Vectors ($\mathbf{f}_H$)
1	Total # of closed contours	$\mathbf{f}_1 \in \mathbb{Z}$, $\mathbf{f}_1 \geq 0$
2	Is the lowest contour closed?	$\mathbf{f}_2 \in \{True, False\}$
3	Level of lowest closed contour	$\mathbf{f}_3 \in \mathbb{R}$.
4	Mean depth of contours	$\mathbf{f}_4 \in \mathbb{R}$.
5	Median depth of contours	$\mathbf{f}_5 \in \mathbb{Z}$, $\mathbf{f}_5 \geq 0$
6	Num. of contours above $\mathbf{f}_5$	$\mathbf{f}_6 \in \mathbb{Z}$, $\mathbf{f}_6 \geq 0$
7	# contours above $\mathbf{f}_6$	$\mathbf{f}_7 \in \mathbb{Z}$, $\mathbf{f}_7 \geq 0$
8	# contours below $\mathbf{f}_5$	$\mathbf{f}_8 \in \mathbb{Z}$, $\mathbf{f}_8 \geq 0$
9	# contours below $\mathbf{f}_6$	$\mathbf{f}_9 \in \mathbb{Z}$, $\mathbf{f}_9 \geq 0$
10	# closed contours above $\mathbf{f}_5$	$\mathbf{f}_{10} \in \mathbb{Z}$, $\mathbf{f}_{10} \geq 0$
11	# closed contours above $\mathbf{f}_6$	$\mathbf{f}_{11} \in \mathbb{Z}$, $\mathbf{f}_{11} \geq 0$
12	# closed contours below $\mathbf{f}_5$	$\mathbf{f}_{12} \in \mathbb{Z}$, $\mathbf{f}_{12} \geq 0$
13	# closed contours below $\mathbf{f}_6$	$\mathbf{f}_{13} \in \mathbb{Z}$, $\mathbf{f}_{13} \geq 0$
14	Contour eccentricity	$\mathbf{f}_{14} \in \mathbb{R}^{N_c \times 1}$, $\mathbf{f}_{14} \geq 0$
15	Contour line waviness	$\mathbf{f}_{15} \in \mathbb{R}^{N_c \times 1}$, $0 \leq \mathbf{f}_{15} \leq 1$
16	Contour area waviness	$\mathbf{f}_{16} \in \mathbb{R}^{N_c \times 1}$, $0 \leq \mathbf{f}_{16} \leq 1$
17	$\frac{\text{contour perimeter}}{\text{mesh boundary perimeter}}$	$\mathbf{f}_{17} \in \mathbb{R}^{N_c \times 1}$, $0 \leq \mathbf{f}_{17} \leq 1$
18	$\frac{\text{contour area}}{\text{mesh area}}$	$\mathbf{f}_{18} \in \mathbb{R}^{N_c \times 1}$, $0 \leq \mathbf{f}_{18} \leq 1$
19	$\frac{\text{circumscribed circle perimeter}}{\text{inscribed circle perimeter}}$	$\mathbf{f}_{19} \in \mathbb{R}^{N_c \times 1}$, $0 \leq \mathbf{f}_{19} \leq 1$
20	$\frac{\text{circumscribed circle area}}{\text{inscribed circle area}}$	$\mathbf{f}_{20} \in \mathbb{R}^{N_c \times 1}$, $0 \leq \mathbf{f}_{20} \leq 1$
For all multidimensional features above (with $H \in \{14, \ldots, 20\}$):		
$H.a$	Mean of non-zero elements	$\mathbf{f}_{H.a} \in \mathbb{R}$.
$H.b$	Mean difference of subsequent non-zero elements	$\mathbf{f}_{H.b} \in \mathbb{R}$.
$H.c$	Median of non-zero elements	$\mathbf{f}_{H.c} \in \mathbb{R}$.
$H.d$	Slope (rate of change) of the line that best fits the non-zero elements	$\mathbf{f}_{H.d} \in \mathbb{R}$.
There is a total of 48 possible isocontour feature descriptors.		

3. ISOCONTOURS-BASED FEATURES

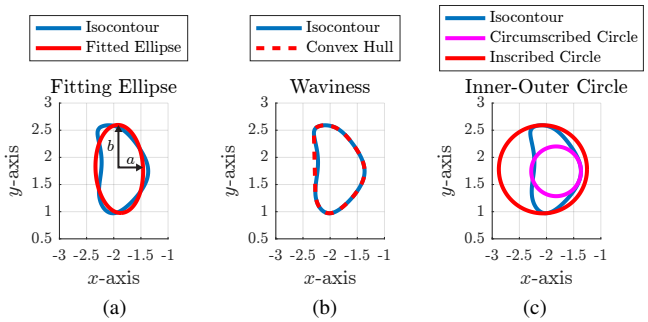
Using the registered 3D meshes, we extract the isocontours and select only the N_c isocontours that enclose the center of the eardrum (indicated with a white X in Fig. 1).

3.1. Isocontour Feature Descriptors

We now calculate a wide variety of statistical information capturing descriptors as well as shape oriented descriptors. Successively, we will test all the proposed descriptors and selected the most important ones for our classification problem. The enumeration of the extracted descriptors together with their dimensionality is presented in Table 1.

The statistical descriptors (#1-13) and their accumulated descriptors (#$H.a$ - $H.d$), with $H \in \{14, \ldots, 20\}$, are trivially described by their names provided in the table. For the remaining descriptors, we provide explanations below:

Contour eccentricity. For each closed contour we fit an ellipse to its coordinates as shown in Fig. 3(a) with principal dimensions a and b and the eccentricity of the ellipse is calculated as the ratio $\frac{a}{b}$. The closer the ellipse is to a circle the closer to 1 the value is.



(a) (b) (c)

Fig. 3: Calculating feature descriptors for an isocontour: (a) Eccentricity (descriptor #14 in Table 1); (b) Waviness (#15-16); (c) Ratio of circumscribed to inscribed circle (#19-20).

Contour line waviness. For each closed contour we calculate its convex hull as shown in Fig. 3(b); the ratio of the perimeter of the convex hull to the perimeter of the contour provides an indication of how wavy each contour is: wavier contours have values smaller than 1.

Contour area waviness. Similar to the line waviness above, the area waviness for a closed contour is defined as the ratio of the area of its convex hull to the area of the contour.

Ratio of contour perimeter to mesh boundary perimeter. For each closed contour we calculate the ratio of its perimeter to the perimeter of the boundary of the mesh giving rise to values between 0 and 1; This gives an indication of the size of the contour compared to the size of the whole mesh.

Ratio of contour area to mesh area. Similar to the above ratio of perimeters, we calculate the ratio between the area of each closed contour to the area of the whole mesh.

Ratio of circumscribed circle perimeter to inscribed circle perimeter. For each closed contour we find the maximum circumscribed circle (i.e., circle that fits inside the contour coordinates) and the minimum inscribed circle (i.e., circle that encloses the whole contour), as shown in Fig. 3(c); their ratio provides a shape factor for the contour (e.g., how closely it resembles a circle) with values closer to 1 determining a more circle-like contour.

Ratio of circumscribed circle area to inscribed circle area. Similar to the aforementioned ratio of perimeters, we calculate the ratio between the area of the maximum circumscribed circle to the area of the minimum inscribed circle.

Each multidimensional descriptor (#14-20) is calculated for all the N_c selected closed contours found on the current mesh. If no closed contours are found, the corresponding values are set to 0. If there are non-zero values, then descriptors $H.a$ - $H.d$, for $H \in \{14, \ldots, 20\}$ are calculated for the sequence of N_c closed contours found on each mesh.

3.2. Multidimensional Descriptor Ordering

Note that in Table 1 there is a slight abuse of notation for the multidimensional descriptors: all of them are marked as being of size $N_c \times 1$. In reality, N_c can be different across different meshes, since it defines the amount of closed contours found and flagged as selected on a specific mesh.

In order to use these multidimensional descriptors to train a classifier, their dimensions need to be the same across different meshes. We use a *priority ordering* for the contours: firstly, we find the maximum amount of selected closed contours in the meshes of the training database, $N_{c,max}$; secondly, all multidimensional descriptors are converted to dimensions $N_{c,max} \times 1$ using zero padding. For example, if for a mesh we have values $\{u_1, u_2, u_3\}$ for a multidimensional descriptor and $N_{c,max}$ is 5, we convert it to $\{u_1, u_2, u_3, 0, 0\}$.

4. CLASSIFICATION METHOD

Once we have calculated the isocontour feature descriptors for all the meshes in our database using the steps described in section 3.1, we can classify these meshes, based on their descriptors, to a set of C classes of a specific classification problem. However, the number of descriptors is very large and some of them might not be very discriminative.

In this section, we use linear Support Vector Machines (SVMs) for performing classification, but another type of classifier could be employed without loss of generality.

4.1. Selecting descriptors for classification

Our goal now is to identify the best set of descriptors that are discriminative for each type of classification problem. For that purpose, we propose an algorithm that greedily tests all isocontour descriptors and ranks them in decreasing order of importance for the classification problem at hand. We test binary classification problems (i.e., there are only two output classes to our classifier - *bulging* vs *non-bulging*), but the proposed algorithm is general and can be used for multiclass classification problems, assuming that an appropriate multiclass classifier is used. The detailed steps of the algorithm are presented in Algorithm 1, and they are here summarized:

1. Test all isocontour descriptors independently and rank them with respect to their importance.
2. Sequentially add one descriptor at a time and test the cumulative accuracy achieved by the set of isocontour descriptors selected.
3. Repeat the previous process until all isocontour descriptors have been selected.
4. The final list of descriptors ($selectedF$) ranks them in decreasing order of importance for the specific classification problem.

Finally, we only select the subset of feature IDs that do not cause the mean accuracy to decrease.

5. RESULTS ON CLINICAL DATA

5.1. Database of eardrums and ground-truth

As part of a clinical trial evaluating the performance of our proposed classification of otitis media in young children, we

Algorithm 1 Selecting most important descriptors

Inputs: Feature descriptor database F_D; Set of descriptor IDs $F_N = \{F_{N,1}, \ldots, F_{N,F}\}$ corresponding to the database F_D (see Table 1); Training (T_ℓ) and testing (Q_ℓ) (or validation K_ℓ) sets, with $\ell = 1, \ldots, L$, for a classification problem with C classes.
if $K_\ell \neq \{\}$ **then**
  $Q_\ell \leftarrow K_\ell$.
end if
$currentF \leftarrow \{\}$.
$selectedF \leftarrow \{\}$.
$f \leftarrow 1$.
while $F_N \neq \{\}$ **do**:
  $selectedF \leftarrow currentF \cup F_{N,f}$.
  for $\ell \leftarrow 1, \ldots, L$ **do**:
    Train classifier for C classes using training partition T_ℓ and descriptors with IDs $\in selectedF$
    Test the trained classifier using validation partition K_ℓ and descriptors with IDs $\in selectedF$
    Store classification accuracies $A_{f,\ell}$.
  end for
  $A_{f,mean} \leftarrow \sum_{\ell=1}^{L} (A_{f,\ell}) / L$.
  $\ell_{min} \leftarrow \underset{\ell}{\operatorname{argmin}} A_{f,\ell} \Rightarrow A_{f,min} \leftarrow A_{f,\ell_{min}}$.
  $\ell_{max} \leftarrow \underset{\ell}{\operatorname{argmax}} A_{f,\ell} \Rightarrow A_{f,max} \leftarrow A_{f,\ell_{max}}$.
  if $f = |F_N|$ **then**
    $\tilde{f} \leftarrow \underset{f}{\operatorname{argmax}} (A_{f,mean} + A_{f,max} + A_{f,min})$.
    $F_N \leftarrow F_N \setminus \{F_{N,\tilde{f}}\}$, (i.e., remove $F_{N,\tilde{f}}$ from set F_N).
    $currentF \leftarrow currentF \cup \{F_{N,\tilde{f}}\}$.
    $f \leftarrow 1$.
  else
    $f \leftarrow f + 1$.
  end if
end while
Outputs: Feature IDs $selectedF$ in decreasing order in terms of importance for performing classification on C classes.

collected light-field images of 482 tympanic membranes from children with AOM (bulged eardrums), OME, and no effusion. Given the challenging conditions under which the data were acquired (i.e., ear wax, lens fogging, and very narrow ear canals), some eardrums end up very dark or blurry, thus limiting the accuracy of extracting features.

Ground-truth. Before capturing images with the device, a panel of three expert otoscopists examined the patients with a conventional otoscope and provided a diagnosis independently. The panel also provided a diagnosis confidence level for each eardrum from 1 (very uncertain) to 5 (very confident). We then selected a subset for which at least two experts agreed in the diagnosis and their confidence level was 4 or 5. Eventually, our ground-truth is formed by 376 images: 68 AOM (diagnosed with bulging), and 308 non-bulged eardrums (including both middle ear effusion and no effusion).

Partitions. Since there is not much data for training and the data is very unbalanced, we divided the dataset in 5 folds and built 5 partitions: each partition ℓ uses 80% of data for training (denoted as T_ℓ) and 20% for test (Q_ℓ) or validation.

Fig. 4: Classification results for *Bulging* versus *No-Bulging* (confusion matrices for all database partitions) using most important *Priority Ordering* isocontour feature descriptors.

5.2. Classification results

After applied Algorithm 1 for *bulging* classification, we obtain that the most important 17 descriptors are the followings: $2, 3, 10, 11, 15[a, b, c, d], 16, 16[a, b, c], 19, 19[a, b], 20[b, c]$. Note that the selection of descriptors uses training and validation sets that do not overlap with the test set.

The final classification results using the most important features for each one of the aforementioned combinations are presented in Fig. 4. Each figure depicts the confusion matrices for the 5 different partitions of the database as well as the mean accuracy across the partitions and the amount of selected features. Note that 3 partitions (out of 5) show an accuracy for bulging detection greater than 91%.

6. DISCUSSION AND FUTURE WORK

We proposed a novel approach for detecting bulging of eardrums, the most important factor for recognizing bacterial infection and distinguishing AOM from other non-bacterial effusions or from no-effusion (healthy condition). The approach consists of classifying eardrums by using for the first time features extracted from 3D data of the eardrum; these data are now available as a result of new types of 3D imaging devices, such as the light-field otoscope used in this work, and they might be beneficial in the future for other classification tasks, other than detection of bulging. We first introduced a variety of geometrical and statistical descriptors based on isocontours, and then selected the ones that best discriminated bulging eardrums. Our approach achieves an average accuracy of approximately 82% on clinical data, reaching almost 92% in the best case. These values have been obtained by using only shape information; we expect that a higher classification performance could be achieved by also considering color information together with shape. In this work, limitations are mostly related to relatively small and unbalanced amount of training data. We are confident that with a larger and better distributed dataset, one can reduce the gap between minumum and maximum accuracies.

7. REFERENCES

[1] K. Blomgren and A. Pitkranta, "Is it possible to diagnose acute otitis media accurately in primary health care?," *Family Practice*, vol. 20, no. 5, pp. 524–527, 2003.

[2] A. Kuruvilla, N. Shaikh, A. Hoberman, and J. Kovačević, "Automated diagnosis of otitis media: vocabulary and grammar," *Journal of Biomedical Imaging*, vol. 2013, pp. 27–41, 2013.

[3] N. Shaikh, A. Hoberman, H. E. Rockette, and M. Kurs-Lasky, "Development of an algorithm for the diagnosis of otitis media," *Academic Pediatrics*, vol. 12, no. 3, pp. 214–218, 2012.

[4] N. Bedard, I. Tošić, L. Meng, A. Hoberman, J. Kovačević, and K. Berkner, "In vivo middle ear imaging with a light field otoscope," *Optics in the Life Sciences, OSA Technical Digest*, 2015.

[5] N. Bedard, T. Shope, A. Hoberman, M. A. Haralam, N. Shaikh, J. Kovačević, N. Balram, and I. Tošić, "Light field otoscope design for 3d in vivo imaging of the middle ear," *Biomedical Optics Express*, vol. 8, no. 1, pp. 260–272, 2017.

[6] R. Socher, B. Huval, B. Bath, C. D. Manning, and A. Y. Ng, "Convolutional-recursive deep learning for 3d object classification," *Advances in Neural Information Processing Systems*, pp. 665–673, 2012.

[7] S. Tang, X. Wang, X. Lv, T. X. Han, J. Keller, Z. He, M. Skubic, and S. Lao, "Histogram of oriented normal vectors for object recognition with a depth sensor," *Asian Conference on Computer Vision*, pp. 525–538, 2012.

[8] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena, "Semantic labeling of 3d point clouds for indoor scenes," *Advances in Neural Information Processing Systems*, pp. 244–252, 2011.

[9] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433–449, 1999.

[10] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 373–380, 2009.

[11] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," *European Conference on Computer Vision*, pp. 356–369, 2010.

[12] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1252–1262, 2007.

[13] A. Wang, J. Lu, J. Cai, T.-J. Cham, and G. Wang, "Large-margin multi-modal deep learning for rgb-d object recognition," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1887–1898, 2015.

[14] H. C. Myburgh, W. H. van Zijl, D. Swanepoel, S. Hellström, and C. Laurent, "Otitis media diagnosis for developing countries using tympanic membrane image-analysis," *EBioMedicine*, vol. 5, pp. 156–160, 2016.

[15] I. Tošić and K. Berkner, "Light field scale-depth space transform for dense depth estimation," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 435–442, 2014.

[16] S. Karygianni, M. Martinello, L. Spinoulas, P. Frossard, and I. Tošić, "Automated eardrum registration from light-field data," *IEEE International Conference on Image Processing*, 2018.

[17] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.