

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

ALIGN TECHNOLOGY, INC.,  
Petitioner,

v.

DENTAL MONITORING SAS,  
Patent Owner.

---

Case No. IPR2023-01369  
Patent No. 10,755,409

---

**DECLARATION OF DR. HASSAN FOROOSH**

## **TABLE OF CONTENTS**

|       |                                              |    |
|-------|----------------------------------------------|----|
| I.    | QUALIFICATIONS .....                         | 1  |
| II.   | SCOPE OF WORK.....                           | 6  |
| III.  | LEGAL STANDARDS .....                        | 7  |
| IV.   | BACKGROUND IN THE ART .....                  | 11 |
|       | A. Image Analysis and Computer Vision .....  | 11 |
|       | B. Edge Detection and Contour Detection..... | 14 |
|       | C. Alignment and Registration .....          | 15 |
|       | D. Deep Learning .....                       | 16 |
| V.    | THE '409 PATENT.....                         | 18 |
| VI.   | PRIOR ART.....                               | 20 |
|       | A. Salah .....                               | 20 |
|       | B. Carrier.....                              | 22 |
|       | 1. Preamble.....                             | 25 |
|       | 2. Element 1.1 .....                         | 25 |
|       | 3. Elements 1.2 and 1.3.....                 | 28 |
|       | 4. Element 1.4 .....                         | 29 |
|       | C. Maninis .....                             | 31 |
| VII.  | OVERVIEW OF THE LEVEL OF SKILL.....          | 35 |
| VIII. | CLAIM CONSTRUCTION .....                     | 36 |
|       | A. image attribute.....                      | 37 |
|       | B. tooth attribute .....                     | 38 |

|     |                                                                                 |    |
|-----|---------------------------------------------------------------------------------|----|
| IX. | CLAIMS 1-15 WERE OBVIOUS AS ILLUSTRATED BY SALAH,<br>CARRIER, AND MANINIS ..... | 40 |
| A.  | The Combination of Salah, Carrier, and Maninis .....                            | 40 |
| B.  | Claim 1 .....                                                                   | 48 |
| 1.  | Claim 1 preamble .....                                                          | 48 |
| 2.  | Step 1(a') .....                                                                | 49 |
| 3.  | Step 1(b') .....                                                                | 53 |
| 4.  | Step 1(c') .....                                                                | 59 |
| 5.  | Step 1(d') .....                                                                | 67 |
| 6.  | Step 1(e') .....                                                                | 68 |
| C.  | Claim 2 .....                                                                   | 74 |
| D.  | Claim 3 .....                                                                   | 78 |
| 1.  | Step 3(1) .....                                                                 | 78 |
| 2.  | Step 3(2) .....                                                                 | 81 |
| 3.  | Step 3(3) .....                                                                 | 81 |
| 4.  | Step 3(4) .....                                                                 | 84 |
| E.  | Claim 4 .....                                                                   | 85 |
| F.  | Claim 5 .....                                                                   | 86 |
| G.  | Claim 6 .....                                                                   | 87 |
| H.  | Claim 7 .....                                                                   | 89 |
| 1.  | Step 7(1') .....                                                                | 90 |
| 2.  | Step 7(2') .....                                                                | 91 |
| 3.  | Step 7(3') .....                                                                | 92 |
| I.  | Claim 8 .....                                                                   | 93 |
| 1.  | Step 8(A) .....                                                                 | 94 |

|      |                                   |     |
|------|-----------------------------------|-----|
| 2.   | Step 8(B) .....                   | 98  |
| 3.   | Step 8(C) .....                   | 100 |
| 4.   | Steps 8(D)-(F) .....              | 102 |
| J.   | Claim 9 .....                     | 105 |
| 1.   | Step 9(A') .....                  | 105 |
| 2.   | Step 9(B') .....                  | 107 |
| 3.   | Step 9(C') .....                  | 107 |
| 4.   | Steps 9(D)-(F) .....              | 110 |
| K.   | Claim 10 .....                    | 111 |
| L.   | Claim 11 .....                    | 112 |
| M.   | Claim 12 .....                    | 113 |
| N.   | Claim 13 .....                    | 114 |
| O.   | Claim 14 .....                    | 116 |
| P.   | Claim 15 .....                    | 118 |
| X.   | CONCLUDING STATEMENTS.....        | 119 |
| XI.  | APPENDIX – LISTING OF CLAIMS..... | 120 |
| XII. | APPENDIX – EXHIBITS .....         | 125 |

I, Hassan Foroosh, declare as follows:

**I. QUALIFICATIONS**

1. I have been retained by counsel on behalf of Align Technology, Inc. to provide opinions in this matter related to my experience and expertise. My opinions are based on my skills, knowledge, training, education, and experience in matters of this nature, and my examination of the materials utilized in preparing this report.

2. I am a Professor of Computer Science at the University of Central Florida. My primary research interests include image/video processing, computer vision, and image-based modeling, including 3D modeling.

3. I received my Ph.D. in Computer Science from INRIA (French National Institute for Research on Informatics and Control), University of Nice Sophia Antipolis in 1996. I also earned a Master of Science in Computer Science from the University of Nice Sophia Antipolis in 1993, a Master of Science in Electrical Engineering from Kingston University, U.K., in 1990, and a Bachelor of Science in Electrical Engineering from Kingston University, U.K., in 1989.

4. Prior to research and teaching, I worked as an R&D Engineer at Innovative Engineering Solutions Ltd. in London from 1990 to 1992. Between 1993 and 1996, I served as a Research Assistant in the Scene Analysis and Symbolic Image Processing Group of INRIA, where I performed research in areas

such as 3D super-resolution, imaging processes for registration, and shape perception. From 1997 to 2000, I served as an Assistant Research Scientist at the University of Maryland, College Park, where I investigated topics such as image registration (2D and 3D), super-resolution of video data for unmanned aerial vehicles, super-resolution of Synthetic Aperture Radar (SAR) and Foliage Penetrating SAR (FOPEN-SAR), Automatic Target Detection and Tracking (ATD/R). During this time, I authored one of the most cited works on registration and alignment based on phase correlation: (Hassan Foroosh (Shekarforoush), Josiane Zerubia and Marc Berthod, Extension of Phase Correlation to Sub-pixel Registration, IEEE Transactions on Image Processing, vol. 11, Issue 3, pp. 188-200, 2002). Between 2000 to 2002, I served as Senior Research Scientist at the University of California, Berkeley where I also served as the Technical Coordinator and Principal Research Scientist of a Multi-University Research Initiative (MURI) funded by the Office of Naval Research. As part of the project, I conducted research on multi-sensor acquisition of 3D data for autonomous 3D modeling and visualization.

5. I joined the Department of Computer Science at the University of Central Florida (UCF) as an Assistant Professor in 2002. Since then I have served as an Associate Professor, Professor, and CAE Link Professor of Computer Science in the same department. I also served in various capacities in the

department, such as the Graduate Program Coordinator, chair of various leadership committees, and most recently in the Faculty Senate Strategic Planning Council for UCF. I have taught a number of courses directed to computer vision and 3D modeling, including: 3D Computer Vision, Image Processing, Special Topics in 3D Computer Vision, and Visual Simulation.

6. I have conducted extensive research in the field of 3D computer vision and computer modeling. Some example research projects that I have conducted include *Video Analysis Techniques for Terrain Modeling and Automated Texture Mapping* (2006-2007) and *Modeling and Visualization of Musculoskeletal Structures in Dental CT-Scan* (2006-2007). For the latter project, I developed algorithms for 3D registration, modeling and visualization of maxilla and mandible of patients for dental surgery using Computer Aided Tomography (CAT) scanning.

7. I have been the principal investigator or co-principal investigator of numerous research projects funded by the National Science Foundation (NSF), NASA, DoD, and industry, with total funding exceeding \$12M since joining UCF. As one example, I recently served as the principal investigator in two projects funded by the National Science Foundation (NSF) on 3D modeling and the role of 3D information in machine understanding of digitized world. One project was directed to the registration/alignment of dental CT scans of patients over a period

of time during the treatment of a patient, with the purpose of improving the quality of 3D models, e.g., corrections and noise removal in 3D CT scans.

8. I have authored or co-authored over 180 articles in peer reviewed journals or conference proceedings. Some example publications include M. Berthod, H. Shekarforoush (Foroosh), M. Werman and J. Zerubia, *Reconstruction of High Resolution 3D Visual Information*, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages. 654-657 (1994); C. Früh, J. Flynn, H. Foroosh, and A. Zakhor, *Fast 3D model generation for urban environments*, Workshop on the Convergence of Graphics, Vision, and Video (CGVV'01), Berkeley, March 2001; Jiangjian Xiao, Xiaochun Cao, and Hassan Foroosh, *3D Object Transfer Between Non-Overlapping Videos*, Proceedings of IEEE Virtual Reality Conference, pages 127-134, (March 25-29, 2006); M. Balci, M. Alnasser, and H. Foroosh, *Subpixel Alignment of MRI Data Under Cartesian and Log-Polar Sampling*, Proceedings of IAPR International Conference on Pattern Recognition (ICPR), vol.3, pages 607-610 (2006); M. Balci, M. Alnasser, and H. Foroosh, *Alignment of Maxillofacial CT Scans to Stone-Cast Models Using 3D Symmetry for Backscattering Artifact Reduction*, Proceedings of Medical Image Understanding and Analysis Conference (MIUA) (2006); and A. Tariq and H. Foroosh, *Designing a Symmetric Classifier for Image Annotation using Multi-Layer Sparse Coding*, *Image and Vision Computing (IVC)*, Image and Vision Computing (IVC) (2018)

(initially published as *Image Annotation using Multi-Layer Sparse Coding*, at <https://arxiv.org/pdf/1705.02460> (May 6, 2017)).

9. Since joining UCF, I have advised and chaired the dissertation of over 30 doctoral students, 24 of whom have already successfully defended their doctoral dissertation, and secured academic positions or industry jobs in major high-tech corporations. As one example, I served as the advisor for a dissertation directed to “*Sub-pixel Registration in Computational Imaging and Applications to Enhancement of Maxillofacial CT Data*”. This research project focused on registration and alignment of 2D images and 3D digitized data obtained by dental CT scans and dental imprints of a patient (e.g., casts), with the aim of improving the quality of CT scans, e.g., eliminating backscattering issues. Methods were developed to register CT models to the physical imprints of patients (e.g., a stone cast model of the patient’s teeth).

10. I have been serving as the editor and/or committee member on numerous professional organizations. For instance, I have been involved with the following organizations: IEEE Transactions on Image Processing (Associate Editor, 2003-2008, 2011-2016); IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Member, Technical Committee, 2004-2012, 2015-2016, 2019-2020); IEEE Transactions on Pattern Analysis and Machine Intelligence (Reviewer, 1996-present).

11. For my work, I have been honored with numerous awards, including the *Piero Zamperoni Award*, International Association for Pattern Recognition (IAPR) (2004). I have also received several Best Paper awards, including “*The Best Scientific Paper Award*” in the IAPR International Conference on Pattern Recognition in 2008, “*The Best Paper Award*” in the IEEE International Conference on Image Processing (ICIP) in 2018, and the *Distinguished Researcher*, UCF College of Engineering and Computer Science (2004-2005, 2005-2006).

12. In sum, I have studied, taught, practiced, and conducted research in the area of computer vision and computer modeling of physical structures, including dental structures.

13. A copy of my *Curriculum Vitae*, attached as EX1003, contains further details on my education, experience, publications, and other qualifications to render an expert opinion in this matter.

## **II. SCOPE OF WORK**

14. I understand that a petition is being filed with the United States Patent and Trademark Office for *Inter Partes* Review of U.S. Patent No. 10,755,409 to Salah et al. (“the ’409 patent,” EX1001), entitled “Method for Analyzing an Image of a Dental Arch.”

15. I have been retained by Align Technology, Inc. (“Align”) to offer an expert opinion on the validity of claims 1-15 of the ’409 patent assigned to Dental Monitoring SAS (“Dental Monitoring”). I receive \$650 per hour for my services. No part of my compensation is dependent on my opinions or on the outcome of this proceeding.

16. I have been specifically asked to provide my opinion on the unpatentability of claims 1-15 of the ’409 patent. In connection with this analysis, I have reviewed the ’409 patent. I have also reviewed and considered various other documents in arriving at my opinions and may cite to them in this declaration.

### **III. LEGAL STANDARDS**

17. I understand that a claim is not patentable under 35 U.S.C. §102 as anticipated if the claim as a whole is described, either expressly or inherently, in a single prior art reference. I understand that this also requires that the prior art reference disclose that each claim element is arranged as recited in the claims.

18. I understand that differences between the prior art reference and a claimed invention, however slight, invoke the question of obviousness, not anticipation.

19. I have been advised that a claimed invention is not patentable under 35 U.S.C. §103 if it is obvious. A patent claim is unpatentable if the claimed invention would have been obvious to a person of ordinary skill in the field at the

time the claimed invention was made. This means that even if all of the requirements of the claim cannot be found in a single prior art reference that would anticipate the claim, a person of ordinary skill in the relevant field who knew about all this prior art would have come up with the claimed invention.

20. I have further been advised that the ultimate conclusion of whether a claim is obvious should be based upon several factual determinations. That is, a determination of obviousness requires inquiries into: (1) the level of ordinary skill in the field; (2) the scope and content of the prior art; (3) what difference, if any, existed between the claimed invention and the prior art; and (4) any secondary evidence bearing on obviousness.

21. I have been advised that, in determining the level of ordinary skill in the field that someone would have had at the time the claimed invention was made, I should consider: (1) the levels of education and experience of persons working in the field; (2) the types of problems encountered in the field; and (3) the sophistication of the technology.

22. I have also been advised that, in determining the scope and content of the prior art, in order to be considered as prior art, a reference must be reasonably related to the claimed invention of the patent. A reference is reasonably related if it is in the same field as the claimed invention or is from another field to which a person of ordinary skill in the field would look to solve a known problem.

23. I have been advised that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was independently known in the prior art. In evaluating whether such a claim would have been obvious, I may consider whether there is a reason that would have prompted a person of ordinary skill in the field to combine the elements or concepts from the prior art in the same way as in the claimed invention.

24. I have been advised that any obviousness rationale for modifying or combining prior art must include a showing that a person of ordinary skill would have had a reasonable expectation of success.

25. I have been further advised that there is no single way to define the line between true inventiveness on the one hand (which is patentable) and the application of common sense and ordinary skill to solve a problem on the other hand (which is not patentable). For example, market forces or other design incentives may be what produced a change, rather than true inventiveness. I may consider whether the change was merely the predictable result of using prior art elements according to their known functions, or whether it was the result of true inventiveness. I may also consider whether there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the patent. I may consider whether the innovation applies a known technique that had been used to improve a similar device or method in a similar way. I may also

consider whether the claimed invention would have been obvious to try, meaning that the claimed innovation was one of a relatively small number of possible approaches to the problem with a reasonable expectation of success by those skilled in the art.

26. I understand that hindsight must not be used when opining on the patentability of claims of a patent. Therefore, my opinion will reflect the understanding of a person of skill in the art as of the relevant patent application filing dates, which I have been instructed is July 21, 2017, for this matter, the earliest date to which applicants have claimed priority for the claims of the '409 patent.

27. I have been instructed for claims to be entitled to an earlier application's filing date, the claims must find written description support in the earlier application to which benefit is claimed by an applicant.

28. I also understand that, in deciding whether the patent satisfies this written description requirement, one must consider the description from the viewpoint of a person skilled in the field of technology of the patent when the application was filed. The written description requirement is satisfied if a person having skill in the art reading the original patent application would have recognized that it describes the full scope of the claimed invention as it is finally claimed in the publication that claims priority to the earlier application and that the

inventor actually possessed that full scope by the filing date of the original application.

#### **IV. BACKGROUND IN THE ART**

29. The '409 patent describes methods involving computer vision and deep learning, with specific focus on analysis of dental images. In the background below, I provide a brief overview of some important elements of computer vision and deep learning that are relevant to understanding what a POSA would have known in these areas at the time of filing of the '409 patent.

##### **A. Image Analysis and Computer Vision**

30. Images are a rich source of information that humans typically manage to analyze with relative ease. Image analysis is often described as the process of extracting information from images that may be subsequently used by either humans or computers to perform certain tasks. Developing methods and computer algorithms to provide the capability of extracting accurate and meaningful information from images has been the subject of study in computer vision in the past few decades.

31. Historically humans have been outperforming computers in tasks that require recognition and understanding of images in terms of their contents and the contextual relationships. It was not until after the early years of the last decade, with major breakthroughs in deep learning (both in terms of architecture and

hardware support), that computers managed to also outperform humans in tasks such as recognition and image content and context understanding. See EX1015, 1-2, 8. Developments of parallel processing hardware (in particular Graphic Processing Units) made it possible for deep learning networks to be trained in an end-to-end manner and on much larger training sets, typically many orders of magnitude larger than before, involving tens of thousands to millions of images.

32. Computer vision is often described in layman's terms as the science of making computers see. One way to conceptually divide vision tasks, based on ideas pioneered by David Marr, divides such tasks into three levels: low-level, mid-level, and high-level tasks (also referred to as computation, algorithm, and implementation tasks). EX1016, 1538-39, 1541. Accordingly, computer vision algorithms were traditionally developed to handle low-level, mid-level, and high-level vision problems. Low-level computer vision refers to computational methods and algorithms that process images at pixel-level, modifying pixel properties to achieve tasks such as quantization, encoding, image restoration, image enhancement, etc. Mid-level computer vision refers to computational methods to produce feature-level representations of images, such as color features, edge features, 3D geometric features, etc. A common feature-level representation widely used in computer vision tasks is boundary-based representations, such as edges and contours that define the separation between regions in images. Finally, high-level

computer vision refers to computational methods and algorithms that aim to understand image contents and contextual relationships, including tasks like object classification, object localization, object instance recognition, etc.

33. Low-level and mid-level processing steps can be intermediate steps towards performing high-level vision tasks. For example, one may perform some restoration or enhancement (low-level) before extracting edges (mid-level) to align the edges with a template contour (low-level) to determine the presence and location of an object in the image (high-level), or to recognize an object or its class (high-level).

34. There were several motivating factors that led to using deep learning methods in computer vision, one of which was to overcome certain shortcomings, such as the treatment of different visual levels independently, as opposed to the way that humans integrate all levels of vision into a unified process. This led to a new class of computer vision methods that are designed to learn features from data, but also made it possible to take advantage of high-level labels to learn better features, while still relying also on low-level image properties and values to extract features. This also made it possible for these methods to allow a network to learn all it needs in order to solve for a vision task, taking raw image data directly as input.

## **B. Edge Detection and Contour Detection**

35. Edges, contours, boundaries, and discontinuities are various terms that have been used interchangeably in the literature to refer to similar concepts involving a representation of a discontinuity in an image. Edges and contours have long been considered as important features used for performing various high-level computer vision tasks, such as classification, object recognition, instance identification, etc. They have also been proven extremely useful for other low-level vision tasks, such as registration and alignment, or template matching that lead to high-level processing and annotation.

36. Classical edge detection was based on low-level properties, such as detected discontinuities, which would then be assigned a binary label (e.g., a 0 or 1) to identify the location of edges. Due to the key role of edges in computer vision and the rising influence of machine learning methods, researchers developed machine-learning based methods (including what would come to be referred to as deep learning) for edge and contour detection. EX1017, 1-2; EX1007, Abstract. Instead of binary labels (such as 1 and 0) for edge locations, deep-learning methods lead to multiple labels in the feature space (i.e. the contour image). EX1017, 5-9; EX1007, 8, 9-12. What this means is that each contour is labeled differently based on various different high-level labels (e.g., class, instance, etc.) reflecting what is depicted within the region for which the contour is a boundary.

For instance, contours of an object class A might be labeled as 1, those of an object class B as 2, those of an object class C as 3, and so on and so forth. These labels can be located on the contours of the object or assigned to the pixels in regions defined by the contour. EX1007, 9-12. Thus, deep-learning methods allow categorization of different objects in an image into specific classes.

37. These type of contours or boundaries are sometimes referred to as “category-specific” or “semantic” contours of boundaries. EX1017, 9; EX1007, 11. Thus, if a neural network is trained to look for features based on a training set with high-level labels (e.g., class labels), the trained machine would be able to return also features (e.g., edges) that are specific to that class. For instance, if a machine is trained to look for the contours of different types of teeth in intraoral images, it would not only detect edges (as in classical methods) but also would be able to label edges and the regions they define based on which type of tooth they belong to instead of a simple binary labeling.

### **C. Alignment and Registration**

38. Alignment and registration refer to a processing step in computer vision allowing comparison of two or more data representations, e.g., two images, an image and a template, a 3D object and an image, a 3D point cloud and a template, two point clouds, two videos, etc. Mathematically, it is a mapping process, i.e. these methods look for a mapping function that would minimize a

distance measure between two or more data sets, where the distance measure is a measure of similarity (or equivalently dissimilarity). That score can be used to label images with annotations. Such image labels can indicate presence or absence of an object or class of objects, object location, or an alignment score with a template or with a viewpoint of a 3D object. *See, e.g.*, EX1018, Abstract, 5-7; EX1019, 391-92, 396, Fig. 3 (example images for “tooth dataset”). Prior to 2017, techniques had been developed to propagate annotations between similar 2D images (such as between video frames), as well as to project annotations from a 3D object to a corresponding region of a 2D image representing the object. *See* EX1013, 3689-90; EX1014, 3198, 3200-01.

#### **D. Deep Learning**

39. Deep learning refers to a class of machine learning methods that rely on many layers of feature learning in order to model the often-non-linear relationship between the input (i.e. the training set) and the output (e.g., probabilities of classes or instances). These models are referred to as “deep” due to the fact that they consist of many layers between the input and the output, also known as the hidden layers, with these layers potentially comprising millions of parameters to be adjusted (i.e. to be trained) through a training process using a training dataset, in order to produce accurate outputs for the unseen test data.

40. By 2017, the best results in computer vision arose from learning models based on deep neural networks (DNNs), such as deep Convolutional Neural Networks (CNNs). EX1007, 2. A key characteristic of DNNs is the fact that the output of each hidden layer may be viewed as a feature map, with features being learned through a training process that yields an optimal feature map at the output of each hidden layer for the problem at hand. In other words, each layer of a trained model effectively functions as a feature extraction algorithm with features that are learned through a training process that upon convergence yields an optimal feature map at the output of each hidden layer for the problem at hand. Furthermore, as we go deeper into the network, the extracted features may be interpreted as higher level.

41. This important property of DNNs implies that the feature maps produced by earlier layers represent lower-level features, whereas the deeper ones correspond to higher-level features. This can be seen in Maninis, for example, which shows how its DNN detects lower, level (fine-scale) features in early layers and higher-level (large-scale) features in later layers. *See* EX1007, 3 (“CNNs are by construction multi-scale feature extractors. If one examines the standard architecture of a CNN consisting of convolutional and spatial pooling layers, it becomes clear that as we move deeper, feature maps capture more global information....”). In applications involving images as input such as image

classification or object recognition, the output feature map of the first layer corresponds almost to an edge map, while deeper feature maps correspond to higher-level features, such as object parts or overall object shape and class. *See* EX1007, Fig. 2. By July 2017, DNNs had proven to be remarkably successful in computer vision and surpass human performance, due to their ability to extract multiple levels of features that are optimized through a training process with large training datasets. *See* EX1007, 7-9 (“human” and “super-human” performance in various tasks).

## **V. THE ’409 PATENT**

42. The ’409 patent is entitled “Method for Analyzing an Image of a Dental Arch.” EX1001, cover (54). The ’409 patent describes acquiring dental images using an image acquisition apparatus, which may be a cellphone or various other types of widely used devices. EX1001, Abstract, 8:56-63. The only independent claim, claim 1, recites as follows:

1. A method for acquiring an image of a dental arch of a patient, said method comprising the following steps:

a’) activation of an image acquisition apparatus so as to acquire an image, called “analysis image”, of said arch;

b’) analysis of the analysis image by means of a deep learning device trained by means of a learning base;

c') determination, for the analysis image, as a function of the results of the analysis in the preceding step, of a value for an image attribute;

d') comparison of said image attribute value with a setpoint;

e') sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof,

to check whether the analysis image respects the setpoint and, if it does not respect the setpoint, to guide the operator in order for him or her to acquire a new analysis image.

EX1001, 32:13-33. Summarized briefly, claim 1 includes steps of acquiring an image, analyzing the image with a trained deep learning device, determining a value for an image attribute based on the analysis, comparing the value to a setpoint, and sending an information message based on the comparison. *Id.* The information message must be related to one or more of several categories: “the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof.” *Id.* If the setpoint is not respected, the apparatus guides the user to acquire a new image. *Id.*

43. The '409 patent concedes that many of the steps of its method are already found in the prior art. For example, the '409 patent repeatedly states that Salah teaches many of the steps recited in its claims. *See* EX1001, 2:42-47, 3:4-10, 7:9-33, 7:50-54, 8:55-56, 9:36-39, 14:48-50, 17:65-18:24, 19:35-37, 20:48-53, 22:7-11, 23:41-24:6, 25:21-28, 27:16-32 (repeatedly acknowledging that WO2016066651 (Salah) teaches techniques described in the specification). Moreover, many of the '409 patent's figures are copied from Salah. For example, Fig. 16 of the '409 patent depicts essentially the same process as Fig. 1 of Salah, and Figs. 10, 11A-D, and 12A-D in the '409 patent are copies of Figs. 2, 4a-d, and 5c-6b in Salah. *Compare* EX1001, Figs. 10-12D, 16, *with* EX1008, Figs. 1-2, 4a-d, 5c-6b.

44. The dependent claims of the '409 patent recite additional elements relating to deep learning, image annotation, modeling of teeth, and acquisition of images. These additions involve routine aspects known in the art and taught or suggested by the grounds references, as I explain in greater detail in section IX.

## **VI. PRIOR ART**

### **A. Salah**

45. PCT Publication No. WO2016066651 to Salah et al. ("Salah," EX1008) is entitled "Method for Monitoring Dentition." Salah was published in French, but I understand that a EX1008 is a certified translation of Salah. Salah

names 4 inventors in common with the '409 patent. Salah was published on May 6, 2016, more than 1 year before the earliest claimed priority date of the '409 patent. Accordingly, I understand that Salah may be considered in evaluating the patentability of the claims of the '409 patent.

46. Salah discloses methods for monitoring patients' teeth, which can be helpful to track the teeth during orthodontic treatment. EX1008, Abstract, 1:12-22, 8:10-18. Two-dimensional images can be obtained of a patient's teeth, and these images can be compared to a three-dimensional model of the patient's teeth, allowing the model to be updated and used to measure progress of treatment. EX1008, 1:26-2:19, 46:15-19, Fig. 1. Salah's analysis of images can include using contour detection on an acquired image to aid mapping to a reference model, and this contour detection can be performed using deep learning. EX1008, 34:11-14, 36:14. Although Salah uses the term "deeplerning," rather than deep learning, this appears to be a typographical error; a POSA would understand that Salah meant to be referring to "deep learning," as deep learning is a well-known method of computer analysis that has many uses, including specifically the detection of contours in images. EX1008, 36:14.

47. Salah teaches that its 2D images may be obtained by a patient or relative using a device such as a cell phone, which can also perform the remaining steps of the evaluation process (including comparison to a 3D model). EX1008,

2:21-27, 31:4-7. Because the users may not have training in image acquisition or dentistry, Salah also teaches that the image-acquisition apparatus can guide the user to obtain appropriate images. EX1008, 27:21-28:10. For example, the apparatus can indicate whether the image is acceptable or needs to be retaken, and it can also provide indications (such as arrows or other instructions) indicating how to adjust the camera's positioning to correctly orient it to obtain an image. *Id.*

### **B. Carrier**

48. U.S. Patent Publication No. 2021/0068923 to Carrier, et al. ("Carrier," EX1005) is entitled "Methods and Apparatuses for Dental Images." Carrier was filed on November 18, 2020, and published on March 11, 2021. Carrier claims priority to Application No. 16/827,594 (EX1010), filed March 23, 2020, and Application No. 15/803,718 (EX1011), filed November 3, 2017, as well as to Provisional Patent Application No. 62/417,985 (EX1006), which was filed November 4, 2016. EX1005, Face.

49. Carrier teaches methods to obtain images of patients' teeth for therapeutic use, including images obtained by a mobile phone. EX1006, Abstract, ¶7 ("The method can comprise viewing the patient's teeth, for example, on a screen of a mobile phone."); EX1005, Abstract, ¶8. The "user's device guides the user in taking high-quality images," including providing an overlay such as an image or outline of teeth that a user may use to align the imaging apparatus. EX1006, ¶¶54-

55 (“In particular, these apparatuses and methods may include the use of an ‘overlay’ on a real-time image of the screen, providing immediate feedback on each of the desired orientations. ... The overlay may be based on a generic image of teeth, or it may be customized to the user's teeth, or to a patient-specific category .... The overlay may be shown as partially transparent, or it may be solid, and/or shown in outline.”); EX1005, ¶¶99-100. The device can assess alignment by (1) performing edge detection on teeth, (2) measuring distance between the teeth’s edges and the overlay, and (3) comparing the distances to a threshold. EX1006, ¶63; EX1005, ¶108. The device can also assess image quality in further ways, including “using image recognition algorithms to ensure all required teeth are visible.” EX1006, ¶81; EX1005, ¶128.

50. I understand that subject matter that is disclosed in the Carrier Publication (EX1005) has an effective filing date of the Carrier Provisional (EX1006) if two conditions are satisfied: first, the same subject matter must be disclosed by both applications, and second, if the disclosure in the Carrier Provisional provides written description for at least one claim of the Carrier Publication. Based on the satisfaction of both of these conditions, I understand that Carrier may be considered in evaluating the patentability of the claims of the ’409 patent.

51. Regarding the first condition, I illustrate that the same subject matter is present in both Carrier references by providing citations to the matching disclosure in both references throughout my declaration. Regarding the second condition, I have reviewed the Carrier Provisional, and it is my opinion that it supports at least one claim in the Carrier Publication. For example, the Carrier Provisional describes each element recited in claim 1 of the Carrier Publication, as I explain below.

52. Claim 1 of the Carrier Publication recites the following:

[Preamble] A computer-implemented method comprising:

[1.1] providing a mobile application on a mobile device associated with a patient, the mobile application configuring a user interface for display on the mobile device;

[1.2] guiding the patient through the user interface to take a first plurality of photos of teeth of the patient, the first plurality of photos corresponding to a first plurality of predetermined views of the teeth, wherein guiding the patient to take the first plurality of photos comprises showing a first plurality of overlays over the teeth in accordance with the first plurality of predetermined views;

[1.3] guiding the patient through the user interface to take a second plurality of photos of the teeth, the second plurality of photos corresponding to a second plurality of predetermined views of their teeth, wherein guiding the patient to take the second plurality of photos comprises showing a second plurality of overlays over the teeth in accordance with the second plurality of predetermined views; and

[1.4] transmitting the first plurality of photos and the second plurality of photos so the first plurality of photos and the second plurality of photos can be evaluated for medical treatment.

EX1005, claim 1. Below, I illustrate that the Carrier Provisional discloses a method including each of these claim elements.

## **1. Preamble**

53. The Carrier Provisional discloses a computer-implemented method. For example, it discloses that “described herein are apparatuses adapted to perform any of the methods described herein, including in particular software, firmware, and/or hardware adapted to perform one or more of these methods.” EX1006, ¶18. Carrier further discloses “computer-readable storage media storing a set of instructions capable of being executed by a processor (e.g., of a mobile phone)” for implementing its methods. EX1006, ¶18. A method performed by a device using software, firmware, and/or hardware, such as one using processor-executable instructions stored in computer-readable storage media, is a computer-implemented method. The Carrier Provisional thus discloses computer-implemented methods.

## **2. Element 1.1**

54. The Carrier Provisional discloses “providing a mobile application on a mobile device associated with a patient, the mobile application configuring a user interface for display on the mobile device.” The Carrier Provisional teaches “a photo uploading mobile app” that may “control a smartphone camera to take dental

images of a patient,” including taking a “predetermined series of images, such as those shown in FIG. 5A, and described” earlier in Carrier. EX1006, ¶¶94-95.

Carrier teaches that its device is “configured to take a series of predetermined specific views of the teeth,” with an app that “guides a user in taking a sequence of specific images, including a user interface as shown in Fig. 5A that allows a user to select an overlay” used as part of the guidance. EX1006, ¶60. This overlay-configuring element is shown in Fig. 5A, reproduced below:

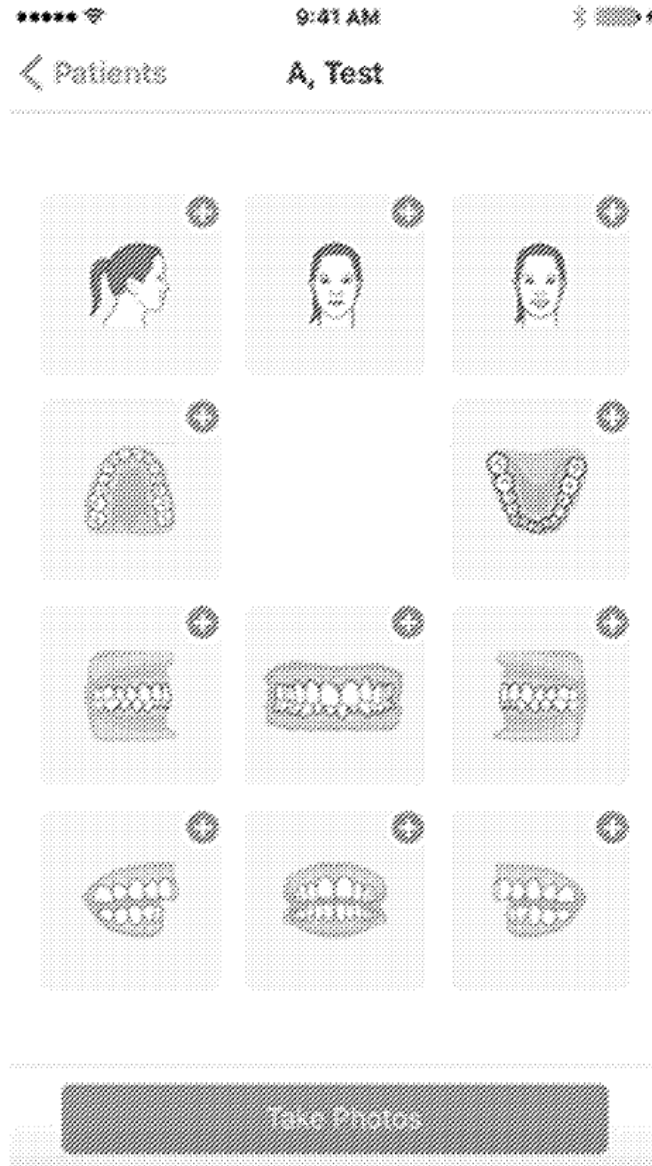


FIG. 5A

EX1006, Fig. 5A; *see also id.*, ¶¶73-75 (describing operation of mobile app to capture multiple pluralities of dental images).

55. The smartphone can be the user's own device—that is, the patient's own mobile device can be used to execute the app and guide the acquisition of

images, according to the process depicted in Fig. 8. EX1006, ¶54, Fig. 8; *see id.*, ¶55 (overlay “may be customized to the user’s teeth”).

56. Accordingly, the Carrier Provisional discloses element 1.1.

### **3. Elements 1.2 and 1.3**

57. Element 1.2 recites “guiding the patient through the user interface to take a first plurality of photos of teeth of the patient, the first plurality of photos corresponding to a first plurality of predetermined views of the teeth, wherein guiding the patient to take the first plurality of photos comprises showing a first plurality of overlays over the teeth in accordance with the first plurality of predetermined views.” Element 1.3 recites the same requirements regarding a “second” plurality of photos, predetermined views, and overlays, respectively. Thus, together, elements 1.2 and 1.3 recite guiding a patient to take two pluralities of photos of teeth, each corresponding to a respective plurality of predetermined views, and each using a respective plurality of overlays. The Carrier Provisional discloses these elements.

58. The Carrier Provisional discloses providing a “plurality of overlays for a plurality of dental images in a plurality of predetermined views.” EX1006, ¶36; *see id.*, ¶¶61, 75. The app “guides a user in taking a sequence of specific images,” which involves using 8 specific overlay images for the required types of photos.” EX1006, ¶¶73-74, Figs. 4A-5A (showing a different overlay for each of 8

different tooth views); *see also id.*, Figs. 8 (showing process for acquiring images), 10 (showing multiple examples of different tooth images from different perspectives). The eight different photo types, corresponding to eight different predetermined views with eight different overlays, include both the first and second sets of pluralities recited in claim 1 of the Carrier Publication. Indeed, Carrier expressly divides its sets of images into subsets, each including a distinct plurality. *See* EX1006, ¶73 (describing several distinct pluralities of images: two “frontal/anterior” view with mouth closed/open; two “right buccal” views with “mouth closed/open”; two “left buccal” views with “mouth closed/open”; as well as “upper jaw, lower jaw, profile, face, etc.”). Thus, the Carrier Provisional discloses guiding a patient through the user interface to take first and second pluralities of photos. respectively corresponding to first and second pluralities of overlays for respective first and second pluralities of predetermined views.

59. Accordingly, the Carrier Provisional discloses elements 1.2 and 1.3.

#### **4. Element 1.4**

60. Element 1.4 recites “transmitting the first plurality of photos and the second plurality of photos so the first plurality of photos and the second plurality of photos can be evaluated for medical treatment.” The Carrier Provisional discloses that the “captured dental images can be transferred to server part for performing more precise estimation of treatment progress.” EX1006, ¶86; *see also*

*id.*, ¶97 (“Images/photos taken may be uploaded by the apparatus into a remote server or site .... Images may be individually uploaded, or they may be uploaded as a composite of multiple separate images.”), Fig. 8. Uploading the captured images would include uploading the captured first and second pluralities recited in elements 1.2 and 1.3. Accordingly, the Carrier Provisional discloses elements 1.4.

61. As I have shown above, the Carrier Provisional provides a written description of each element of claim 1 of the Carrier Publication. The following claim chart summarizes the locations of relevant support for each element of claim 1 of the Carrier Publication:

| <b>US20210068923</b>                                                                                                                                                                                                                                                                                                                                                                                              | <b>US62/417,985</b>                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|
| 1. A computer-implemented method comprising:                                                                                                                                                                                                                                                                                                                                                                      | EX1006, ¶18.                                     |
| providing a mobile application on a mobile device associated with a patient, the mobile application configuring a user interface for display on the mobile device;                                                                                                                                                                                                                                                | EX1006, ¶¶54-55, 73-75, 94-95, Figs. 5A, 8.      |
| guiding the patient through the user interface to take a first plurality of photos of teeth of the patient, the first plurality of photos corresponding to a first plurality of predetermined views of the teeth, wherein guiding the patient to take the first plurality of photos comprises showing a first plurality of overlays over the teeth in accordance with the first plurality of predetermined views; | EX1006, ¶¶36, 61, 73-75, 85, Figs. 4A-5A, 8, 10. |
| guiding the patient through the user interface to take a second plurality of photos of the teeth, the second plurality                                                                                                                                                                                                                                                                                            | EX1006, ¶¶36, 61, 73-75, 85, Figs. 4A-5A, 8, 10. |

| US20210068923                                                                                                                                                                                                                                                                               | US62/417,985 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| of photos corresponding to a second plurality of predetermined views of their teeth, wherein guiding the patient to take the second plurality of photos comprises showing a second plurality of overlays over the teeth in accordance with the second plurality of predetermined views; and |              |

### C. Maninis

62. Kevis-Kokitsi Maninis et al. published “Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks” (“Maninis,” EX1007) no later than May 2, 2017, on arXiv.org. Copies of Maninis are available to download from links at an abstract page (<https://arxiv.org/abs/1701.04658>), with two versions available: version 2 (EX1007), dated April 28, 2017, and version 1 (EX1021), dated January 17, 2017. The download link for version 2 is <https://arxiv.org/pdf/1701.04658> (or <https://arxiv.org/pdf/1701.04658v2>), and the download link for version 1 is <https://arxiv.org/pdf/1701.04658v1>. The version submitted as EX1007 was downloaded from the version 2 link. The two versions of the paper contain only minor differences, but for purposes of this declaration, I focus on the April 28, 2017, version of Maninis (EX1007).

63. I am familiar with arXiv.org as a user of the system and a registered author of numerous articles available from the system. In May 2017, I submitted several papers to arXiv.org in the same field, so I am familiar with the processes

practiced by arXiv.org for publishing submitted papers at the time that Maninis was submitted. arXiv.org is an openly accessible, moderated repository for scholarly articles, covering an array of scientific fields. Authors submit their articles to arXiv.org to disseminate them freely to interested members of the public. Authors of scientific papers use arXiv.org as a method of publishing their papers that reliably establishes a date of publication, as the arXiv.org repository permanently marks each document with a stamp indicating the date of submission, and arXiv.org consistently publishes submitted documents to the interested scientific community promptly after submission, as I discuss below.

64. The date of submission of a document is reflected in a stamp on the left side of the paper's first page; for example, Maninis indicates a submission date of April 28, 2017. EX1007, 1; EX1022. Authors indicate the field relevant to a paper at the time of submission, which is reflected as part of a stamp on the paper's first page. For example, the stamp for Maninis includes the code "[cs.CV]," indicating that the paper is a computer science paper in the field of computer vision. EX1007, 1.

65. As shown on an archived webpage of the arXiv.org website, it was the established policy of arXiv.org in 2017 to publish submitted papers within about one business day of the date of submission. EX1022 (archived May 3, 2017). As the arXiv.org website explains, a submitted paper will be announced about one day

after being submitted, and it will then be mailed to all arXiv.org subscribers (who would typically be persons interested in the paper's relevant field) shortly thereafter. EX1022. For example, a paper submitted on a Friday would be announced Sunday or Monday and mailed to subscribers no later than Tuesday morning. EX1022. This is consistent with my experience with arXiv.org typical practice in early 2017, as I observed regarding papers I submitted and papers I received as a subscriber during that time. Regarding Maninis, with a submission date of April 28, 2017, it would have been announced on Sunday April 30, 2017, or Monday May 1, 2017, and would have been transmitted to subscribers no later than Tuesday May 2, 2017.

66. Once a paper is announced, it was arXiv.org's stated policy that the paper may not be modified without creating a new version. EX1023 (archived July 18, 2017); EX1026 (archived September 8, 2016). Maninis itself reflects this practice, as the document indicates that it is the second version (with v2 as part of the stamp on the left side of the document), meaning that it was a revision made on April 28, 2017, to an earlier version (that had been submitted on January 17, 2017, *see* EX1021). Thus, under arXiv.org's standard practice, the version of Maninis submitted on April 28, 2017, would not have been changed after its announcement (on or about April 30, 2017). I have reviewed metadata contained in a copy of Maninis downloaded from the download link (<https://arxiv.org/pdf/1701.04658>),

and this metadata indicates that the document was created and last modified on April 30, 2017. This is consistent with Maninis being placed in final form by arXiv.org as of an announcement date of April 30, 2017, with the version of Maninis that was published at that time being the same as the document provided as EX1007. This is also consistent with my experience with arXiv.org regular practice from early 2017 until today.

67. The public availability of Maninis to interested persons prior to July 21, 2017, is also corroborated by archived webpages of both arXiv.org and the authors of Maninis. As of May 3, 2017, the university website featuring the author's Convolutional Oriented Boundaries software included a link (to <https://arxiv.org/abs/1701.04658>) for downloading a PDF copy of Maninis from arXiv.org. EX1024 (archived May 3, 2017). Additionally, as of July 8, 2017, the arXiv.org author search webpage for Prof. Luc Van Gool, the last author listed on Maninis, listed Maninis as one of his recent papers that was available on arXiv.org, providing links to both the abstract page (<https://arxiv.org/abs/1701.04658>) and a link to directly download Maninis (<https://arxiv.org/pdf/1701.04658>). EX1025 (archived July 8, 2017). Thus, Maninis was disseminated to interested persons in the scientific community prior to July 21, 2017. Accordingly, I understand that Maninis may be considered in evaluating the patentability of the claims of the '409 patent.

68. Maninis discloses methods for image classification and contour detection using a Convolutional Neural Network (CNN). EX1007, Abstract. Maninis teaches a “deep learning architecture” that produces significant improvements over state-of-the-art techniques. EX1007, 3, 8-9. Maninis uses a neural network trained on precisely annotated images to detect contours that define regions containing objects. EX1007, 3-4, 5-6, 13, Fig. 2. Maninis accomplishes this contour detection by using a deep learning neural network to generate a “contour probability map,” which is transformed “into a hierarchical boundary map,” with different probabilities translating to different “strengths” of borders between regions containing objects in the image. EX1007, 3, 5. Maninis also demonstrates the ability to label the specific object contained within each identified region, using a process called semantic boundaries, which Maninis shows to also improve over the state of the art. EX1007, 11-12, Tables 4-7, Fig. 16.

## **VII. OVERVIEW OF THE LEVEL OF SKILL**

69. I have been advised that “a person of ordinary skill in the relevant field” (“POSA”) is a hypothetical person to whom one could assign a routine task with reasonable confidence that the task would be successfully carried out. I have been advised further that, in determining the level of ordinary skill in the art that someone would have had at the time the claimed invention was made, I should consider: (1) the levels of education and experience of persons working in the

field; (2) the types of problems encountered in the field; and (3) the sophistication of the technology.

70. In my opinion, a person of ordinary skill in the relevant field as of July 21, 2017 would typically have an advanced degree, such as a Masters' Degree or Ph.D. in computer science or a related field, with at least two years of work in computer vision, image analysis, medical imaging, or related areas. A POSA could have less education but significant professional experience in one or more of these fields. A POSA could be a member of an interdisciplinary team including persons with backgrounds such as electrical engineering, computer science, optical imaging, and dentistry or orthodontics, and could call upon the knowledge of other team members as appropriate.

## **VIII. CLAIM CONSTRUCTION**

71. I have been advised that the claim terms of the '409 patent claim are to be given the plain and ordinary meaning that the terms would have to a person of ordinary skill in the art in question at the relevant priority date. I understand that the analysis focuses on how the patentee used the claim term in the claims, specification, and prosecution history. I also understand that dictionaries or other extrinsic sources may assist in determining the plain and ordinary meaning but cannot override a meaning that is unambiguous from the intrinsic evidence. I have followed these principles in my analysis throughout this declaration.

72. The scope of the challenged claims is discussed throughout my analysis; I provide additional discussion of some claim terms below.

**A. image attribute**

73. Claim 1, as well as some dependent claims, uses the term “image attribute.” The ’409 patent describes an image attribute as an attribute that has a value that “relates to the whole image and not to part of the image.” EX1001, 3:23-37. The specification also provides a list of many different types of image attributes, including attributes related to:

a position and/or an orientation and/or a calibration of an acquisition apparatus used to acquire said analysis image, and/or

a quality of the analysis image, and in particular relating to the brightness, to the contrast or to the sharpness of the analysis image, and/or

the content of the analysis image, for example to the representation of the arches, of the tongue, of the mouth, of the lips, of the jaws, of the gums, of one or more teeth or of a dental, preferably orthodontic, appliance.

EX1001, 3:38-49. Claim 13 recites an almost identical list of image attributes, stating that the image attribute may relate to

a position or an orientation or a calibration of an acquisition apparatus used to acquire said analysis image, or a combination thereof, and/or

a quality of the analysis image, and in particular relating to the brightness, to the contrast or to the sharpness of the analysis image, and/or

the content of the analysis image, in particular to the representation of the arches, of the tongue, of the mouth, of the lips, of the jaws, of the gums, of one or more teeth or of an orthodontic appliance.

EX1001, claim 13. During prosecution, the applicant pointed to this language from the specification in arguing that an image attribute must relate to the image as a whole. EX1004, 858-59.

74. Regarding the “value” of an image attribute, the specification indicates that the value need not be a single numerical value. For example, a list of values for a set of teeth would also be an image attribute, such as a list of all tooth numbers of teeth represented in an image. EX1001, 26:16-26. Based on the description in the specification and the arguments made during prosecution, a POSA would interpret the term “image attribute” to include an attribute having a value (which may be a set of values for multiple teeth) relating to an image as a whole, including but not limited to an attribute related to any one or more of the examples provided the specification and claim 13.

#### **B. tooth attribute**

75. The specification of the ’409 patent distinguishes image and tooth attributes, with image attributes relating to an image as a whole and tooth attribute

relating to a tooth. *See* EX1001, 1:22-28, 3:23-37. However, the specification also explains that a tooth attribute may define an image attribute. EX1001, 3:59-62. For example, a tooth attribute can be a mean separation of a single tooth from an appliance, and an image attribute can be a mean separation of all teeth from the appliance. EX1001, 28:58-65. The '409 patent states that one or more tooth attributes can be “associated with the tooth models as a function of the teeth that they model,” as well as referring to “a tooth attribute relating to a tooth represented on the analysis image.” EX1001, 1:26-27, 8:7-8. The specification also lists examples of tooth attribute categories, although as with image attributes the list is not exhaustive. *See* EX1001, 8:7-50. For example, the specification states that

The tooth attribute is preferably chosen from a tooth number, a tooth type, a tooth shape parameter, for example a tooth width, in particular a mesio-palatine width, a thickness, a crown height, an index of mesial and distal deflection of the incisive margin, or a level of abrasion, a tooth appearance parameter, in particular a translucency index or a color parameter, a parameter relating to the condition of the tooth, for example “abraded”, “broken”, “decayed” or “fitted” (that is to say in contact with a dental, preferably orthodontic, appliance), an age for the patient, or a combination of these attributes. A tooth attribute is preferably an attribute which only relates to the tooth modeled by the tooth model.

EX1001, 8:7-21.

76. Claim 4 recites a subset of these options, indicating that for claim 4, the “tooth attribute is chosen from a tooth number, a tooth type, a tooth shape parameter, a tooth appearance parameter, a parameter relating to the state of the tooth, an age for the patient, or a combination of these attributes.” EX1001, claim 4. Based on the description in the specification and claims, a POSA would interpret the term “tooth attribute” to include any attribute related to a tooth.

**IX. CLAIMS 1-15 WERE OBVIOUS AS ILLUSTRATED BY SALAH, CARRIER, AND MANINIS**

**A. The Combination of Salah, Carrier, and Maninis**

77. Salah teaches methods that are useful for monitoring patient’s teeth, for example, to track the teeth during orthodontic treatment, including steps of acquiring images of the patient’s teeth at various times during treatment. EX1008, Abstract, 8:19-28. Salah teaches that the acquired images can include two-dimensional images, which may be acquired by an image acquisition apparatus such as a mobile phone. EX1008, 13:24-28, 31:4-9. These two-dimensional images can be obtained by users who may not have training in dental arts and who lack specialized stabilization equipment, such as the patient him- or herself, or a relative of the patient. *Id.* To aid such users in acquiring images, Salah teaches that the image acquisition apparatus may assess whether a photo is “the photo is acceptable or that, on the contrary, he/she must take another photo,” and it can also assist a user in positioning the apparatus. EX1008, 27:21-28:2. For example, the image

acquisition apparatus can include marks on a viewfinder or screen to guide alignment, and the device may provide feedback such as arrows or audio cues to indicate how to adjust the apparatus's position to optimize alignment. *Id.*

78. Salah also discloses that acquired images may be analyzed, include determining contours of objects such as teeth, and it further teaches that this process can employ processes such as edge detection, object detection, and deep learning. EX1008, 34:5-36:14. As I discuss below, Carrier discloses further details not expressly taught by Salah involving assessing image alignment and acceptability based on such analysis of an image generating values to be compared to a setpoint, to aid in guiding users to acquire images.

79. Carrier teaches methods to guide a user in acquiring dentition images with an image acquisition apparatus such as a mobile phone. EX1006, ¶7; EX1006, ¶9. Similar to Salah, Carrier teaches methods for “increasing the accuracy of the alignment of the camera lenses with teeth of the patient” to produce “[h]igh quality dental images [that] are usually required to submit a dental case or request a case assessment.” EX1006, ¶57; EX1005, ¶102. In one aspect of this guidance, Carrier provides an overlay, such as an outline of teeth superimposed onto the image. This overlay guides the user to align the image acquisition apparatus (such as a mobile phone) to capture an image of a patient's teeth in an appropriate alignment, as the patient can align the teeth on the screen with the overlay, which is positioned to

represent a correct alignment. EX1006, ¶¶57, 63, Fig. 6B; EX1005, ¶¶102, 108, Fig. 6B. This resembles an aspect of Salah’s teaching, in which “augmented reality” may be used to superimpose a semitransparent image of a reference model to provide similar alignment guidance. EX1008, 30:16-21. Fig. 6B of Carrier (below) shows an example of an overlay of one arch being superimposed and approximately aligned with the teeth of that arch:

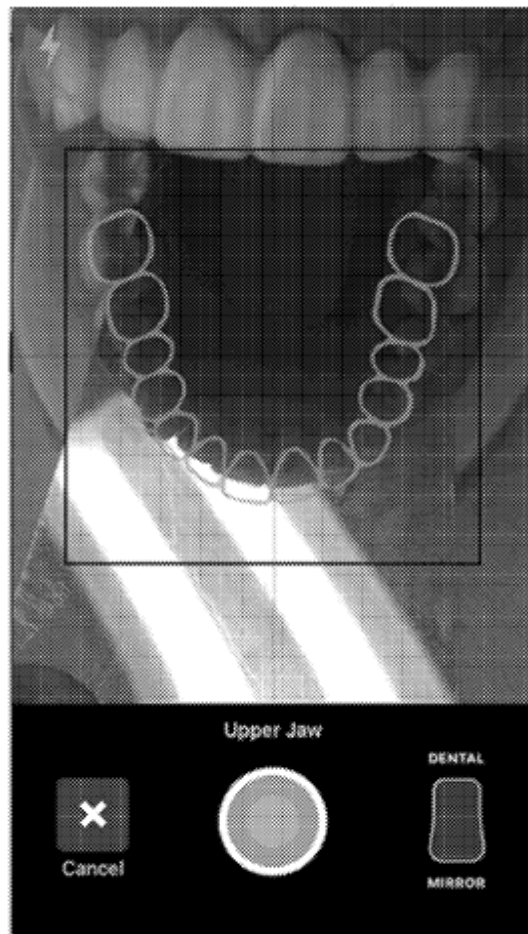


FIG. 6B

80. As part of its alignment assessment, Carrier also teaches using image recognition to detect whether all required teeth are visible, as well as detecting whether the imaged teeth are within a threshold distance of the overlay. EX1006, ¶¶63, 81; EX1005, ¶¶108, 128.

81. A POSA would have had reason to use Carrier's technique for assessing alignment of teeth to guide users in an image acquisition apparatus as taught by Salah. Both references are directed toward the same field of endeavor, namely dental imaging. Furthermore, Salah and Carrier provide complementary teachings regarding the same subject within that field: namely obtaining dentition images with users' mobile phones. Both references teach assessing alignment and using the assessment to guide users, with Carrier providing specific details regarding how to effectively assess alignment based on the appearance of the patient's teeth, thereby ensuring high-quality dentition images. A POSA would thus have considered Carrier's teachings directly applicable to the types of imaging methods taught by Salah and would use Carrier's complementary teachings to supplement Salah's disclosure. For example, a POSA would be motivated to use Carrier's measures of image alignment and the presence of all required teeth to provide alignment guidance to users as well as to assess the acceptability of alignment based on images being acquired during the alignment process.

82. In combination, Salah's and Carrier's teachings would each perform substantially the same function as they provide alone. In particular, Carrier's techniques for assessing alignment would be used to determine the feedback to Salah's users when guiding them to acquire dental images, such as indicating whether the patient's teeth were positioned properly in the image, whether all required teeth were visible, and also indicating what types of adjustments should be made to correct unacceptable images. A POSA would thus have had a reasonable expectation of success in combining Salah and Carrier, as the combination involves the use of complementary techniques, applied in understood ways, to produce a predictable outcome.

83. While the combination of Salah and Carrier teaches analyzing an image to produce a value for an image attribute, to the extent that these two references do not expressly teach that this involves using a deep learning device trained by means of a learning base, this is taught by Maninis. Maninis teaches a method referred to as "Convolutional Oriented Boundaries (COB)," which employs a deep learning device to perform image classification, object detection, and semantic segmentation. EX1007, Abstract, 11-12. Maninis provides methods of "contour detection" that produce "multiscale oriented contours and region hierarchies" using "Convolutional Neural Networks (CNNs)." *Id.* To accomplish this, Maninis's device divides images into regions containing different objects, in

both foreground and background, and then classifies the objects into appropriate categories. See EX1007, 1-2, 5-6, 8, 11-12. Maninis accomplished these tasks by training its neural network on an appropriately-labeled training dataset—that is, a “learning base” in the terminology of the ’409 patent. EX1007, 1-2, 7-8.

84. Although Maninis is not restricted to dental imaging, Maninis describes broadly applicable techniques for image classification and object detection using category-specific contour detection based on a deep learning device trained using a learning base. For example, Maninis describes the ability of its system to transfer between different types of neural network architectures and thereby “transfer further improvements ... to perceptual grouping.” EX1007, 13. Maninis also confirms that its process is not narrowly limited to fixed categories, and can in fact generalize concepts of objects even to categories it has not originally seen. EX1007, 10-11. Carrier expressly invites the incorporation of various possible techniques into its apparatus, teaching that its edge detection be achieved using “any appropriate technique (e.g., edge detection image processing, etc.)” and that detecting presence of required teeth may be performed by “image recognition algorithms.” EX1006, ¶¶63, 81; EX1005, ¶¶108, 128. Maninis would be a natural selection for such a technique, as the use of deep learning for object and contour detection represented the established trend in computer vision, as recognized by Maninis. EX1007, 2. Similarly, Salah teaches that image analysis

techniques for detecting contours may employ deep learning, which would further motivate a POSA to apply Maninis's teachings to the imaging apparatus of Salah and Carrier. EX1008, 34:5-36:14. Although Salah notes that certain other image analysis techniques are "preferred," it notes that even if not preferred, deep learning "can also be used." *Id.* A POSA would not interpret Salah's identification of some techniques as preferred over deep learning as a reason not to apply deep learning; rather, a POSA would recognize that Salah had identified deep learning as a useful option that can be applied. Moreover, as I explain below (as well as in §IV), trends developing in computer vision at the time favored increased adoption of deep learning for image analysis, which would further motivate a POSA to employ it in applications like those disclosed by Salah and Carrier.

85. A POSA would view Maninis as providing good reasons to employ deep learning techniques for object and contour detection. Maninis teaches that the "latest wave of contour detectors takes advantage of deep learning to obtain state-of-the-art results." EX1007, 2. Maninis explains that its COB technique is compatible with different CNNs and is applicable in a "modular" fashion that "can incorporate and benefit from future improvements in the base CNN." *Id.* Maninis describes that its technique provides improvements relative to "recent approaches to contour detection using deep learning." EX1007, 3, 10. For example, Maninis demonstrates substantially faster object edge and contour detection and image

classification relative to state-of-the-art techniques. See EX1007, 10 (“Overall, comparing to previous state-of-the-art, we get a significant improvement at a fraction of the computation time (24.37 to 0.79 seconds).”), 11-12 (describing improvements in semantic segmentation to classify specific types of depicted objects).

86. Maninis’s flexibility and prospect to improve performance would motivate a POSA to analyze images using a deep learning device, as taught by Maninis, to improve Salah and Carrier’s image acquisition apparatus. For example, a POSA would be motivated to adapt Salah and Carrier’s image acquisition device to employ a neural network as taught by Maninis, trained on images relevant to Salah’s and Carrier’s dental imaging applications. In that combination, Maninis’s neural network would provide specific functionalities required by Salah and Carrier, including the ability to identify objects (such as the required teeth Carrier proposes using image recognition software to identify) and to detect contours (such as the tooth edges Carrier uses to assess alignments), to aid in guiding image acquisition, as taught by Salah.

87. A POSA would have reasonably expected such a combination to work predictably and successfully. Maninis teaches that its technique is modular, compatible with multiple different architectures, and yields improvements in both speed and accuracy over prior state-of-the-art techniques. EX1007, 8-13. Adapting

Maninis’s techniques to an image acquisition apparatus as taught by Salah and Carrier would also be consistent with trends in computer vision, as I describe above in section IV. Maninis agrees, noting that the use of deep learning constituted the trend for the “latest wave of contour detectors,” allowing them “to obtain state-of-the-art results.” EX1007, 2. Thus, based on Maninis’s promising disclosure and reported improvements on the state-of-the-art, a POSA would be motivated to incorporate a trained neural network as taught by Maninis to accomplish the edge detection and object identification tasks required by Salah and Carrier, with a reasonable expectation of success.

88. Below, I note specific prior art disclosure that satisfies each element of each challenged claim.

### **B. Claim 1**

89. Salah, Carrier, and Maninis teach each and every element of claim 1, as I explain below.

#### **1. Claim 1 preamble**

90. The preamble of claim 1 recites “[a] method for acquiring an image of a dental arch of a patient, said method comprising the following steps.” Salah and Carrier teach the preamble of claim 1.

91. Salah teaches acquiring images of patients’ dental arches. EX1008, Abstract (“[A]cquiring, by means of a first acquisition apparatus, at least one first

two-dimensional image of said teeth....”), 12:20-24 (“The invention also relates to a method for acquiring a two-dimensional image of a part of a dental arch, or of a dental arch or of both dental arches of a patient....”). Carrier similarly teaches methods of obtaining images of patients’ teeth “for therapeutic use.” EX1006, ¶7; EX1005, ¶9. A patient’s dental arch is formed of the patient’s teeth, so by teaching a method to acquire an image of the patient’s teeth, Carrier is teaching a method to acquire an image of the patient’s dental arch.

92. Both Salah and Carrier teach using a mobile device (such as a mobile phone) to acquire dental images, with the device providing guidance for the acquisition of the images. *See, e.g.*, EX1008, 27:21-28:2 (describing guidance by personal device), Figs. 5A-D; EX1006, ¶¶59 (users visually match mobile phone’s on-screen overlay with teeth), 96 (app guides users in taking series of photographs), Figs. 2A-3B, 5C-7C; EX1005, ¶¶104, 173. Carrier illustrates images acquired by this method in the figures. *See, e.g.*, EX1006, Figs. 2A-3B, 5C-7C; EX1005, Figs. 2A-3B, 5C-7C. These images include images of the patient’s dental arch.

93. Salah and Carrier thus teach the preamble of claim 1.

## **2. Step 1(a’)**

94. Salah and Carrier teach step (a’) of claim 1, which recites “activation of an image acquisition apparatus so as to acquire an image, called ‘analysis

image,’ of said arch.” Salah teaches that a user—which may be a non-specialist such as a patient or the patient’s relative—may acquire images of a patient’s dentition using an image acquisition apparatus such as a mobile phone. EX1008, 31:4-7. This image acquisition apparatus can include a photographic or video camera, and it can include a viewfinder or screen display to aid in alignment. EX1008, 27:11-13 (“The image acquisition apparatus preferably comprises a camera or a video or infrared camera, which the user, e.g. the patient or one of their relatives, positions by means of a viewfinder or screen, before operating it”). The use of such an apparatus to acquire images displayed on a viewfinder or screen when aligning the apparatus constitutes the recited “activation” to acquire an “analysis image,” as when such a device displays an image being captured by a camera on a screen, the device has been activated and is acquiring the images. *See also* EX1001, 25:66-26:8 (acquiring real time image being displayed on cellphone constitutes activation of device to acquire analysis image, even if the image is not stored).

95. Carrier similarly teaches acquiring images of a patient’s teeth (including the patient’s dental arch) using a “mobile device such as a smartphone.” EX1006, ¶94; EX1005, ¶171. Carrier teaches that acquiring an image includes “viewing the patient’s teeth, for example, on a screen of a mobile phone” and “displaying, on the screen, an overlay comprising an outline of teeth in a

predetermined view, wherein the overlay is displayed atop the view of the patient's teeth." EX1006, ¶7; *see* EX1005, ¶8 (describing the same in reference to "the screen of a mobile telecommunications device having a camera (e.g., smartphone, mobile phone, etc.)").

96. Carrier's description of displaying an image of the patient's teeth on a mobile phone's screen matches the normal interface used to acquire photographs using modern mobile phones, such as smartphones. That is, when the mobile phone's imaging software is activated, images (e.g., video images) are acquired using the mobile phone's camera and displayed on the screen to allow the user to orient the camera correctly to capture the desired image. Indeed, similar to Carrier, the '409 patent identifies an example of acquiring an analysis image in which a real time image is displayed on cellphone, and the '409 patent further notes that this constitutes acquiring an analysis image regardless of whether the image is stored or discarded.

97. The use of an overlay in Carrier provides an indication to aid in such alignment. *See* EX1006, ¶59 ("The overl[a]y can help to precisely align the phone camera lens with the teeth of the patient and position the mobile telecommunications device at a right distance and a right angle to the teeth of the patient."), EX1005, ¶104 (same). Examples of such mobile phone images, including an overlay superimposed on top of an acquired image of a patient's

dental arch are shown in the figures, such as Figs. 3A-B, reproduced below, with the camera being moved between Figs. 3A and 3B to align the overlay 301 with the patient's upper arch. See EX1006, ¶¶62-63; EX1005 ¶¶107-108.

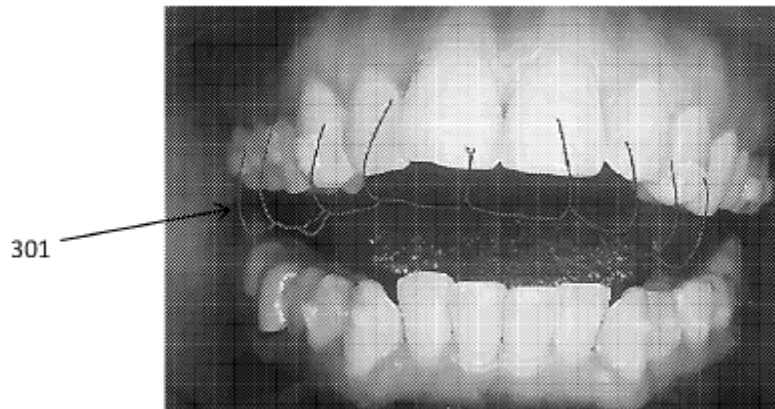


FIG. 3A

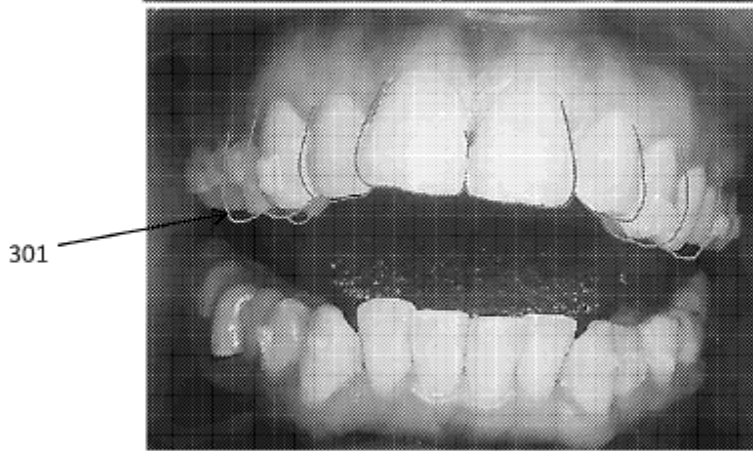


FIG. 3B

EX1006, Figs. 3A-B; see EX1005, Figs. 3A-B (same images as drawings); see also, e.g., EX1006, Fig. 7B (showing overlay of both arches on image acquired by mobile phone); EX1005, Fig. 7B.

98. The mobile phone is thus an image acquisition apparatus, its use to acquire images during alignment is an activation of that image acquisition apparatus, and the image of the arch that is acquired is an “analysis image,” since

Salah and Carrier collectively teach analyzing the images to guide alignment by the user, as I discuss below for the subsequent elements of claim 1. Accordingly, Salah and Carrier teach step 1(a').

### **3. Step 1(b')**

99. Step (b') of claim 1 recites “analysis of the analysis image by means of a deep learning device trained by means of a learning base.” I have been instructed that this claim term should be interpreted to encompass a device comprising a deep learning neural network trained with a learning base, performing a function of analyzing an image (i.e., the analysis image). I apply this interpretation in my discussion throughout this declaration. Salah, Carrier, and Maninis teach this element.

100. Salah teaches that its image acquisition apparatus may guide a user in acquiring images, including determining whether the image is acceptable as well as aiding in positioning via visual or audio feedback indicating how to position the acquisition apparatus. EX1008, 27:21-28:2. Salah further teaches that acquired images may be analyzed, including analysis to detect contours of teeth. EX1008, 34:5-14. A contour map of detected tooth edges is shown in Fig. 6b, reproduced below.

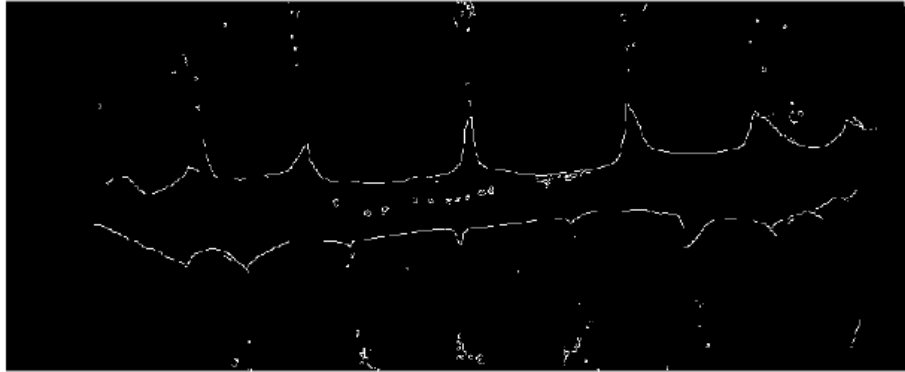


Fig. 6b

EX1008, 34:5-14, Fig. 6b.

101. Salah teaches that its contour detection may employ any of a variety of known methods, including object detection and deep learning. See EX1008, 34:5-36:14 (detecting “contour information,” and identifying a variety of applicable techniques, including using a “MeanShift filter” to “find an object on an image projection,” as well as using “Deepleraning [sic]”). In particular, Salah teaches that an image can be processed using “image processing software” that “can be used, for example, to detect high-contrast regions and determine contours.” EX1008, 34:11-14. More particularly, to detect “discriminant information” (e.g., “contour information”), Salah teaches a list of “well-known and preferred methods,” while noting that other methods, though not necessarily preferred, “can also be used,” including analysis “by Deepleraning.” EX1008, 34:5-36:14. A POSA reading Salah would recognize that “Deepleraning” is a typographical error, and that Salah meant to refer to the use of “deep learning” methods, as the use of

deep learning was well-known for various applications including detection of contours. *See generally* EX1007.

102. Carrier provides further details regarding the image analysis involved in guiding alignment. In particular, Carrier teaches that assessing whether an image of a patient's dental arches is properly aligned can be partially automated by analyzing alignment of the edges of the imaged teeth to an overlay to provide feedback about alignment to the user. To accomplish this assessment, the imaging apparatus (e.g., the "software/application software") determines whether the teeth are aligned with the overlay on the screen, and it then uses that determination to display an indicator when the teeth are in alignment. EX1006, ¶63; EX1005, ¶108. This alignment involves analyzing the image to determine distances between the edges of the patient's teeth and the overlay, with alignment indicated when, for example, "a threshold amount/percentage of the edges of the patient's teeth in the image are within a predetermined minimum distance of the overlay." EX1006, ¶63; EX1005, ¶108.

103. Determining a distance between the edges of the patient's teeth and the overlay requires analyzing the image to detect the edges of the patient's teeth. Carrier discloses that "[t]he edges of the teeth may be determined by any appropriate technique (e.g., edge detection image processing, etc.)." EX1006, ¶63; EX1005, ¶108. Carrier also describes that further analysis of the image may be

performed, including analyzing image quality, image focus, and image brightness, and Carrier further describes that its image analysis “can comprise analyzing the image using a library” and can use “image recognition algorithms to ensure that all required teeth are visible.” EX1006, ¶81; EX1005, ¶128. Carrier thus invites the use of applicable image analysis techniques to be used as part of its alignment analysis.

104. Although Carrier does not expressly disclose that its image analysis techniques, such as edge detection, employ “a deep learning device trained by means of a learning base,” Carrier discloses that “[t]he edges of the teeth may be determined by any appropriate technique (e.g., edge detection image processing, etc.).” EX1006, ¶63; EX1005, ¶108. As one example of an “appropriate technique,” the use of a deep learning device for the types of image analysis taught by Carrier was well known in the art. Indeed, as I noted above, Salah expressly teaches using deep learning for contour detection. EX1008, 34:5-36:14.

105. Maninis provides an example of deep learning techniques a POSA would have found useful for the edge and contour detection taught by Salah and Carrier. Maninis teaches methods of detecting edges (or “contours”) in images using a deep learning device in order to perform image classification, using a process referred to as “Convolutional Oriented Boundaries (COB).” EX1007, Abstract. Maninis discloses methods of “contour detection” that produce

“multiscale oriented contours and region hierarchies” using “Convolutional Neural Networks (CNNs).” *Id.* In particular, Maninis’s deep learning architecture generates a “contour probability map,” which is transformed “into a hierarchical boundary map,” with different probabilities translating to different “strengths” of borders between regions containing objects in the image. EX1007, 3, 5; *see id.*, Figs. 3-5 (depicting contour probability maps used to detect contours between regions).

106. Maninis describes that “latest wave of contour detectors takes advantage of deep learning to obtain state-of-the-art results,” and further notes that “[a]t the core of all these deep learning approaches, lies a *base CNN*,” with Maninis’s COB technique being compatible with different CNNs, applicable in a “modular” fashion that “can incorporate and benefit from future improvements in the base CNN.” EX1007, 2-3 (emphasis original). Maninis describes that its techniques provide improvements over “recent approaches to contour detection using deep learning,” presenting a “CNN architecture to produce contour detection at different scales, ... all in a single CNN forward pass at the image level.” EX1007, 2-3. Maninis’s “deep learning architecture” is illustrated in Fig. 2 (below), showing the generation of contours by using multiple layers of feature maps in a convolutional neural network to detect image contents at different scales (fine and coarse detections) as well as different orientations.

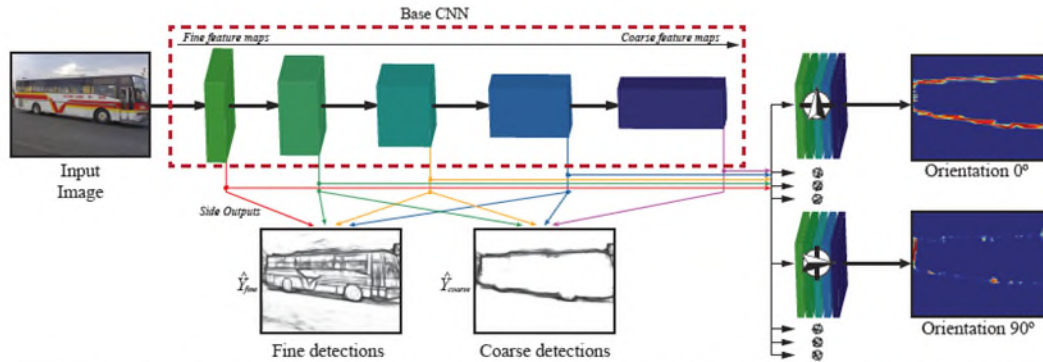


Fig. 2. Our deep learning architecture (best viewed in color). The connections show the different stages that are used to generate the multiscale contours. Orientations further require additional convolutional layers in multiple stages of the network.

EX1007, Fig. 2.

107. Maninis’s use of a computing device comprising a convolutional neural network not only meets the interpretation of step 1(b)’s deep learning device that I discuss above, but it more particularly matches the preferred embodiment of a deep learning device for the ’409 patent. EX1007, 3-5, 9. The ’409 patent repeatedly states that a “deep learning device” is “preferably a neural network.” EX1001, 1:24-25, 1:39-40, 1:51-2:3, 3:14-18, 4:5-6, 8:41-50, 13:64-65, 24:14-15, 24:46-47, 24:66-25:4, 28:27-29, 29:24-25, 29:54-55. Even the specific types of neural networks overlap between the ’409 patent and Maninis: the ’409 patent lists “Microsoft ResNet” and “VGG Net” as two examples of neural networks, matching Maninis’s presentation of “results both using VGGNet and ResNet.” EX1001, 16:13-50; EX1007, 2.

108. Maninis describes that, as part of its neural network’s deep learning, the network is trained on a “training dataset” (i.e., a “learning base” as recited in the ’409 patent). EX1007, 3. Maninis provides an example using a training dataset

of 10,103 images (“the PASCAL Context dataset”) with “localized pixel-wise semantic annotations.” EX1007, 5-6. These annotations are indications, pixel-by-pixel, of the locations of different objects in each image, identifying what object is depicted in each part of the image. To provide “training and optimization,” to the neural network, Maninis divides this set of images into “train, validation, and test sets.” *Id.* These subdivisions of its training dataset (learning base) are then used to train and validate the neural network, demonstrating a substantial improvement over prior classification techniques. *See* EX1007, 5-6, Table 3 (improvement from 24.37 s to 0.79 s), Fig. 6 (showing improved classification accuracy relative to local gradient estimation techniques).

109. In sum, Salah and Carrier teach performing analysis of the analysis image to determine values for tooth presence and tooth edges, and these functions would be performed by means of a deep learning device trained by means of a learning base as taught by Salah and Maninis. Accordingly, Salah, Carrier, and Maninis teach step 1(b’).

#### **4. Step 1(c’)**

110. Step (c’) of claim 1 recites “determination, for the analysis image, as a function of the results of the analysis in the preceding step, of a value for an image attribute.” As I describe above, an attribute having a value related to, among other options, a position, orientation, and/or calibration of the acquisition apparatus or

the content of the analysis image would be an image attribute. EX1001, 3:38-49, claim 13; *supra* section VIII.A. Salah, Carrier, and Maninis teach this element, as I explain below.

111. Salah teaches that the acquisition apparatus can provide feedback that indicates whether an acquired dentition image is (1) acceptable or (2) needs to be retaken, and it can also indicate how the apparatus needs to be moved for correct positioning. EX1008, 27:21-28:2. While Salah does not expressly state that providing this feedback involves determining a value for an image attribute, a POSA would recognize that determining image attribute values would be a natural way to provide this functionality. In order for a computer program to make a determination about whether specified criteria are met (such as criteria for acceptability of an image), POSAs understood that the program should produce one or more values that can differentiate between meeting and not meeting the criteria. Likewise, to provide an indication about how an imaging apparatus must be moved to properly align it, a POSA would naturally seek to produce values related to alignment, such as measures of angle or displacement from correct alignment for the image. Generating values for the image's relevant attributes would be a routine step to include in software that determines whether an image is "acceptable" or properly positioned. For example, as I discuss below, at least two

types of image analysis taught by Carrier involve determining such image attribute values.

112. In one example of an image attribute, Carrier teaches that its image processing produces edges of teeth, which are used to determine whether the teeth are positioned properly in the frame for high-quality imaging. EX1006, ¶63; EX1005, ¶108. Carrier teaches that this alignment may be determined by assessing a distance between the edges of each tooth and the overlay. EX1006, ¶63; EX1005, ¶108. For example, Carrier teaches that the distance may be measured to determine whether “the edges of the teeth are within about 1 mm, about 0.5 mm, about 0.1 mm, about 0.05mm, etc.)” of the overlay. EX1006, ¶63; EX1005, ¶108. Carrier also teaches that, as another option, the measure used by the device to assess image alignment may be an average of the teeth’s deviation from the overlay. EX1006, ¶68; EX1005, ¶114. In relation to the specification of the ’409 patent, the set of tooth distances would be analogous to the set of tooth numbers represented on the image, whereas the average distance from the overlay would be analogous to the average distance between the teeth’s edges and the aligner.

113. Carrier’s tooth-overlay distances relate to several of the examples I highlighted above in section VIII.A. For example, the distances relate to position, orientation, and/or calibration of the acquisition apparatus, with larger distances indicating larger discrepancies in these values. Additionally, the distances would

relate to the content of the analysis image, since the distance of teeth from the overlay relates to the representation of the patient's teeth or arches. Thus, each of Carrier's distances is a value for an image attribute.

114. Furthermore, Carrier's distances are "determine[ed], for the analysis image, as a function of the results of the analysis in the preceding step." The "analysis in the preceding step" is the recited "analysis of the analysis image by means of a deep learning device trained by means of a learning base." Carrier's distances are determined "as a function of the results" of that analysis because the analysis identifies the edges of teeth (the "results" of the analysis), and the distance between teeth and overlay is calculated as a function of the edges (that is, the edges are used as an input to calculate the distances between each tooth and the overlay). Specifically, in the context of Salah and Maninis, these edges would be derived from the deep learning contour detection taught by Salah and Maninis.

115. Carrier also discloses that the apparatus may use "image recognition algorithms to ensure that all required teeth are visible," which involves another example of image attribute. EX1006, ¶81; EX1005, ¶128. A POSA would understand that to ensure all required teeth are visible, each required tooth should be detected by the image recognition algorithm—that is, by applying Maninis's deep learning techniques.

116. Maninis teaches how to use its contour detection algorithm to detect the presence of objects in an image. *See* EX1007, Abstract, 1-2, 11-12. Use of object detection, as taught by Maninis, in the image acquisition apparatus of Salah and Carrier to detect visible teeth would determine an image attribute related to the content of the analysis image—specifically, it would relate to which teeth are visible.

117. Maninis’s deep learning algorithm uses the contours it generates to define boundaries around regions and to identify the objects in each region of the image. *See* EX1007, Abstract, 11-12 (discussing semantic contours and semantic segmentation). Maninis describes this type of image recognition task as involving “image classification, object detection, and semantic segmentation” where detected contours are assigned to categories corresponding to the “ground-truth” (i.e., the identification of the real-life objects present in both the foreground and background of an image). EX1007, 1-2. Objects are identified in regions by “label[ing] each of the detected pixels with the corresponding semantic class.” EX1007, 11-12; *see id.*, Tables 4-12 (identifying specific object types). By labeling sets of pixels with the objects detected to be at the pixels, Maninis produces values (i.e., the values assigned as labels) for the objects in the image.

118. In view of Maninis’s teaching that its deep-learning-based contour detection techniques can be used to identify and classify objects in an image, a

POSA would be motivated to use those techniques as part of Carrier’s “image recognition algorithms” to detect individual teeth corresponding to the contours used by Carrier to assess alignment. Adjusting Maninis’s neural network to the specific task of identifying teeth would be a straightforward task well within the skill of the POSA. To accomplish such an application, a POSA would train Maninis using a training dataset representative of the types of images to be categorized (i.e., images of patients’ dental arches taken in the context of Salah and Carrier’s imaging technique). These images would be labeled with appropriate annotations representing the categories (such as types of teeth or individual tooth identifiers) that would be required to identify required teeth. A POSA would then set the output layers of Maninis’s neural network to provide the required object identifications corresponding to the labels used to identify individual teeth, based on Maninis’s teaching of semantic boundaries, so that the identified teeth in an image could be compared to a set of required teeth.

119. Regarding the specific types of categories to be annotated and identified, a POSA would be motivated to identify teeth at least by type, such as incisor, canine, pre-molar, or molar. Salah and Carrier both show (e.g., in their figures) that these different types of teeth have different shapes. *See, e.g.*, EX1008, Fig. 4d (segmented arch model with different shapes for different tooth types); EX1006, Fig. 6B (showing different tooth shapes across arch for both teeth and

outlines); EX1005, Fig. 6B (same). Indeed, Salah even discusses that to best acquire images of different types of teeth, different acquisition conditions may be desirable. *See* EX1008, 54:23-26 (varying acquisition conditions based on tooth type, such as “incisor” and “molar”). Differences in tooth shape would be detectable using object detection techniques like those taught by Maninis, as the differences would be visible in the image and result in both different edges and different pixel characteristics (such as shades and patterns) within the regions defined by the edges. To identify individual teeth, a POSA would be further motivated to use the long-established system of tooth numbering to annotate each tooth in a training dataset used to train a neural network like that disclosed by Maninis to identify teeth by tooth number, as identification by tooth number would be a straightforward, unambiguous way of identifying required teeth that was well-established in a dental context. *E.g.*, EX1009, ¶¶7-8, 12 (discussing known methods of categorizing teeth by type and/or number). Using different categories for each tooth type or number would be consistent with Maninis, which teaches classifying based on large numbers of different categories. EX1007, 5 (training on dataset annotated with 459 different semantic categories).

120. Because the deep learning techniques of Maninis support detection of objects such as the “required teeth” taught by Carrier, the values indicating

whether the required teeth are all present or not would likewise be determined as a function of the deep learning image analysis taught by Salah and Maninis.

121. Finally, the values generated to assess Carrier's distances and to ensure that required teeth are visible would also be values for attributes of the image as a whole. Regarding distances, a set of all tooth distances or an average of those distances would be an attribute of the image as a whole because each type of value is a measure of overall image alignment, not merely a property of a single part of the image (e.g., one tooth) taken in isolation. *See also* EX1001, 28:58-65, 31:28-37 ("The value of the image attribute can also, for example, be a measurement of the separation, for example a measurement of the maximum separation, or of the mean separation between the teeth represented on the analysis image and the aligner."). Similarly, values (such as tooth types or tooth numbers) that represent whether Carrier's required teeth are present or absent would be values of image attributes because they represent a property of whether the required teeth are present in the image as a whole. Indeed, this latter example matches closely with an example provided in the '409 patent's specification, where a "list of the tooth numbers represented" constitutes an image attribute. EX1001, 26:16-26.

122. Thus, Salah, Carrier, and Maninis teach step 1(c').

## **5. Step 1(d')**

123. Step (d') of claim 1 recites “comparison of said image attribute value with a setpoint.” Salah, Carrier, and Maninis teach this element.

124. As I discussed above for step 1(c'), Salah's image acquisition apparatus assesses whether an image is acceptable and properly positioned. EX1008, 27:21-28:2. While Salah does not expressly disclose that evaluating whether an image is “acceptable” involves comparison to a setpoint, a POSA would consider such a comparison to be a routine option. Speaking generally, a POSA would implement an assessment of whether an image is “acceptable” or “not acceptable” by defining some criteria for acceptability and determining whether those criteria were satisfied. Such criteria would naturally be used to define a setpoint, which would be the value or values differentiating between acceptable and unacceptable images. Carrier, in particular, teaches comparing image attribute values to setpoints for determining whether an image is acceptable, as I explain below.

125. Carrier produces image attribute values in the form of distances from teeth to an overlay, as I explained for step 1(c'). Carrier further teaches that these values may be compared to a setpoint. For example, Carrier teaches that the distances may be compared to a threshold, such as determining “whether the edges of the teeth are within about 1 mm, about 0.5 mm, about 0.1 mm, about 0.05mm,

etc.)” of the overlay. EX1006, ¶63; EX1005, ¶108. If a sufficient percentage of teeth are closer than this threshold, the apparatus can use this as an indication of alignment. EX1006, ¶63; EX1005, ¶108. Carrier also expounds on alternative implementations of this threshold comparison, teaching that it may be implemented “by estimating an indicator of an average deviation of the outline from a contour of the teeth of the patient.” EX1006, ¶68, EX1005, ¶114. Thus, Carrier teaches that the “image attribute value” (distance from teeth to overlay) is compared with a setpoint (threshold distance to indicate alignment for image of the teeth).

126. Regarding determining whether the image contains required teeth, Carrier again teaches a comparison to a setpoint. Carrier indicates that its image quality assessment can involve the device determining whether “all required teeth are visible.” EX1006, ¶81; EX1005, ¶128. Applying this teaching, a POSA would implement such a functionality in the combined device of Salah, Carrier, and Maninis by determining whether every required tooth was “present” or “not present,” which is a comparison of a value for presence (i.e., visibility as detected in the image) against a setpoint of those teeth being required to be visible.

127. Thus, Salah, Carrier, and Maninis teach claim element 1(d’).

## **6. Step 1(e’)**

128. Step 1(e’) recites “sending of an information message as a function of said comparison, the information message being related to the quality of the image

acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof, to check whether the analysis image respects the setpoint and, if it does not respect the setpoint, to guide the operator in order for him or her to acquire a new analysis image.” As I explain below, Salah, Carrier, and Maninis teach this step.

129. To begin with, Salah teaches using “written and/or voice prompts” to guide a user in acquiring images of a patient’s teeth. EX1008, 27:21-28:2. This written or voice message can include “inform the user that the photo is acceptable or that, on the contrary, he/she must take another photo.” EX1008, 27:21-28:2. Salah also teaches guiding the user to acquire a new analysis image; for example, a user can be told “he must redo a photo,” and the device can further “assist in positioning, for example by visual prompts (e.g., by displaying arrows), and/or audio messages (such as succession of beeps, the frequency of which increases as device’s positioning improves), and/or written and/or vocal prompts (‘higher,’ ‘lower,’ etc.).” EX1008, 27:21-28:2. Thus, Salah teaches sending information messages as recited in step 1(e’).

130. Like Salah, Carrier teaches sending information messages regarding whether the image is aligned, and Carrier’s messages specifically check whether the setpoint is respected by the analysis image. For example, Carrier teaches that

the device may “indicate (e.g., by visual and/or audible and/or tactile) image that the teeth are aligned in the frame with the overlay” based on whether a sufficient number of teeth are sufficiently aligned to bring their edges within a threshold distance of the overlay. EX1006, ¶63; EX1005, ¶108; *see also* EX1006, ¶68 (average deviation may be used), EX1005, ¶114. Carrier shows an indicator in Fig. 3B, explaining that “[i]f the overlay nearly matches the teeth, the indicator can be triggered, which indicates the moment to capture the image.” EX1006, ¶67; EX1005, ¶112. In Fig. 3B, the overlay changes color upon alignment, illustrated as a change from a dark gray line to a light gray line as shown below:

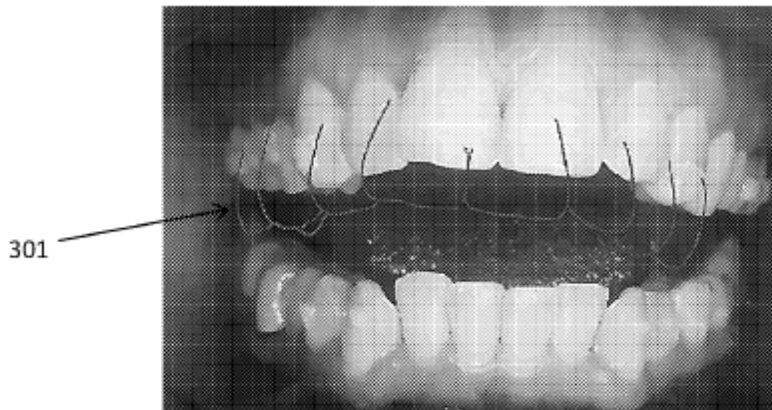


FIG. 3A



FIG. 3B

131. Carrier teaches sending an information message as a function of its comparison indicating whether the analysis image respects the setpoint. As discussed above, if Carrier's setpoint is not "respected" (i.e., if a sufficient number of teeth are not sufficiently aligned to bring their edges within a threshold distance of the overlay), then Carrier's apparatus determines that the patient's teeth are not in alignment. *See* EX1006, ¶63; EX1005, ¶108.

132. Carrier further explains that the overlay may be colored red to indicate misalignment and green to indicate alignment. EX1006, ¶69 ("When the color of the outline changes from red to green, it is indicating the overlay nearly matches the teeth and it is the moment to capture the image."); EX1005, ¶115. As yet another example of an information message indicating whether the setpoint is respected, Carrier teaches that "one or more arrows may be shown on the screen to indicate that the user should move the mobile phone in a particular direction (or by a particular amount) to align the patient's teeth with the overlay." EX1006, ¶70; EX1005, ¶116. This resembles Salah's teaching that arrows may be used to indicate how the position of the image acquisition apparatus should be adjusted. *See* EX1008, 27:21-28:2.

133. In addition to checking whether the setpoint is respected based on the comparison, the messages taught by Carrier and Salah also guide the user to acquire a new analysis image if the setpoint is not respected. The audible or visible

indicators (such as color changes) guide a user because the color indicates whether the alignment is correct, and audio cues like Salah's beeping provide an analogous type of guidance. Further, the arrow indicators taught by Salah and Carrier guide a user directly, as they show how to reposition the image acquisition apparatus. Thus, Salah and Carrier teach analogous indicators that guide a user to acquire a new image when a given image is not acceptable.

134. Carrier's determination of whether all required teeth are visible would similarly involve an information message based on a comparison to a setpoint that guides a user to acquire a new image when the original did not respect the setpoint. Carrier teaches that image recognition should "ensure that all required teeth are visible." EX1006, ¶81; EX1005, ¶128. Where not all required teeth are visible, this would render an image unacceptable. In such a circumstance, a POSA would be motivated to have the device send an information message indicating the unacceptability of the photo as currently acquired, such as the message taught by Salah that the user "must redo a photo." EX1008, 27:21-28:2. This message would relate to the comparison discussed for step 1(d') (because it would be conditioned on that comparison having indicated an unacceptable image) and it would expressly guide the user to acquire a new image by telling the user to redo imaging (such as by adjusting the device or opening the mouth sufficiently to capture the required teeth).

135. Finally, the messages taught by Salah and Carrier relate to the list of properties recited in step 1(e')—that is, “related to the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof.”

136. For example, the arrows and other audio or visual messages relate to “the position of the acquisition apparatus in relation to said arch” because they indicate whether the acquisition apparatus (mobile phone) is properly aligned with the patient’s dental arch or whether it needs to be adjusted, including optionally indicating the direction or amount of adjustment. EX1006, ¶70 (“For example, one or more arrows may be shown on the screen to indicate that the user should move the mobile phone in a particular direction (or by a particular amount) to align the patient’s teeth with the overlay.”); EX1005, ¶116.

137. The messages can also relate multiple categories, such as image quality, apparatus positioning, and opening of the mouth. Salah expressly teaches messages regarding whether the image is “acceptable,” which is an assessment of quality. EX1008, 27:21-28:2. Where acceptability was based on the presence or absence of required teeth, as taught by Carrier, this would relate to quality (lack of visible teeth indicating poor quality), in addition to relating to apparatus positioning (because if the apparatus is mispositioned, required teeth can end up

obscured or out of frame). It would also be related to the opening of the mouth, as a POSA would recognize that, as a matter of common sense, teeth will not be visible in a photograph taken when a patient's mouth is closed. This commonsense notion is reflected in Figs. 9 and 10 of Carrier, which show that a patient closing his or her mouth will make teeth no longer visible. *See* EX1006, Figs. 9-10; EX1005, Figs. 9-10. As an example, an excerpt of Fig. 10 is shown below, with side-by-side images showing a patient with mouth open and teeth visible, and mouth closed and teeth not visible.



EX1006, Fig. 10 (excerpted); *see* EX1005, Fig. 10.

138. Thus, Salah, Carrier, and Maninis teach step 1(e').

139. Accordingly, it is my opinion that claim 1 would have been obvious over Salah, Carrier, and Maninis.

### **C. Claim 2**

140. Claim 2 recites “The method as claimed in claim 1, in which, in the step b'), at least one analysis tooth zone is identified that represents, at least

partially, a tooth on said analysis image, and at least one value of a tooth attribute is determined for said at least one analysis tooth zone, and, in the step c'), the value for the image attribute is determined as a function of said tooth attribute value.”

Salah, Carrier, and Maninis teach this element.

141. As I discuss above for step 1(e'), Carrier teaches that, as part of its process of guiding image capture by users, the apparatus may use “image recognition algorithms to ensure that all required teeth are visible.” EX1006, ¶81; EX1005, ¶128. Moreover, Maninis teaches that its process of contour detection generates boundaries that divide an image into regions, which can then be used for object detection. *See* EX1007, Abstract, 1-2, 11-12. In the context of Salah and Carrier, the contours detected by Maninis would include the edges of teeth, and regions determined by these tooth edges would be regions determined to each contain a tooth. Each such region would constitute an “analysis tooth zone” that has been “identified that represents, at least partially, a tooth on said analysis image,” since the region would be a zone at least partially representing a tooth on the image.

142. Furthermore, the combination of Salah, Carrier, and Maninis teaches “at least one value of a tooth attribute is determined for said at least one analysis tooth zone.” A POSA would be motivated to identify the teeth shown on the image, such as identifying the teeth by type or number, to accomplish Carrier’s

teaching to use “image recognition algorithms to ensure that all required teeth are visible,” as the question of whether each “required” tooth is visible is naturally answered by identifying which specific teeth (e.g., by type or number) are visible and comparing to a set of teeth that are required. EX1006, ¶¶81; EX1005, ¶128. In combination with Salah and Maninis, this would involve detecting which tooth is present in each “tooth analysis zone” (that is, in each region defined by tooth contours), because Maninis teaches that its regions may be used for object detection. See EX1007, Abstract, 1-2, 11-12. Detection of which tooth is present in a contour-defined region would thus be a natural and straightforward application of Maninis’s object-detection process to provide an image recognition algorithm for ensuring required teeth are present, as taught by Carrier.

143. The identification of which particular tooth is present in an analysis tooth zone involves assigning a value of a tooth attribute for the zone. For example, the attribute is the identification of the tooth that present in the zone. In one example, a number could be used to identify a tooth; for example, a POSA would be familiar with the well-known assignment of unique tooth numbers to each tooth as long practiced in dentistry, so each identified tooth might be assigned the tooth number for the tooth that the image recognition algorithm determines to be present. See EX1009, ¶¶7-8, 12. Similarly, teeth could be determined by type, as dividing teeth into different types was similarly well-known. See *id.* The

identification of which tooth is present by number or type in a given zone corresponds to specific examples of tooth attributes recited later in claim 4.

144. Further, using Maninis's edge detection to detect the tooth edges meets another example of a "tooth attribute" as described in the '409 patent: "a tooth shape parameter." EX1001, 8:9-21. The edges of a tooth constitute a specific set of parameters describing the shape of the tooth. In particular, the detected edges constitute the tooth's outline as it appears in the image, and they are thus literally parameters describing the shape of the tooth in the image.

145. Finally, Salah, Carrier, and Maninis teaches the last part of claim 2, which recites "the value for the image attribute is determined as a function of said tooth attribute value." As I discuss above for claim 1, Carrier generates a value for distance based on the edges of each tooth (i.e., a measure of distance from an overlay), and Carrier also motivates generating values representing whether required teeth are present based on recognition of required teeth visible in the image. *See* EX1006, ¶¶63, 70, 81; EX1005, ¶¶108, 116, 128. Thus, for each of the tooth attributes discussed above, the values are generated based on a determined tooth attribute as recited in claim 2, such as tooth type, tooth number, or a tooth shape parameter in the form of a tooth outline.

146. Accordingly, it is my opinion that claim 2 would have been obvious over Salah, Carrier, and Maninis.

### **D. Claim 3**

147. Claim 3 depends from claim 2 and recites multiple elements relating to the formation and use of the “learning base” and the analysis performed in step b’ of claim 1. As I discuss below, each of these elements is taught by Salah, Carrier, and Maninis.

#### **1. Step 3(1)**

148. Step 1 of claim 3 recites “creation of a learning base comprising more than 1000 images of dental arches, or ‘historical images,’ each historical image comprising one or more zones each representing a tooth, or ‘historical tooth zones,’ to each of which, for said tooth attribute, a tooth attribute value is assigned.” Salah, Carrier, and Maninis teach this element.

149. To begin with, a POSA would have understood that a large number of images would be needed to train any deep learning device. Indeed, a POSA would recognize that far more than 1000 images would typically be contained in a learning base (i.e., a training dataset) from which images were used to train such a device. For example, Maninis discloses training its neural network on images selected from a training dataset (corresponding to the ’409 patent’s “learning base”) of over 10,000 images. *See* EX1007, 1 (using “a trainval set with more than 10 000 challenging and varied images”), 5-6 (“We extend the main BSDS benchmarks to the PASCAL Context dataset, which contains carefully localized

pixel-wise semantic annotations for the entire image on the PASCAL VOC 2010 detection trainval set. This results in 459 semantic categories across 10,103 images, which is an order of magnitude (20×) larger than the BSDS.”). Thus, as specifically taught by Maninis, a POSA would consider it natural to train a deep learning device with more than 1000 images.

150. With regard to the images being “images of dental arches, or ‘historical images,’” a POSA would have understood that this would be a routine manner of adapting Maninis’s contour and object detection device to Carrier’s dental imaging apparatus. Maninis discloses using a training set with 459 semantic categories corresponding to the objects depicted in the images. EX1007, 5-6. The semantic category depends on the object depicted in each image; for example, the images shown in Fig. 9 depict images containing animals (a dog, a cat, etc.), foreground objects (a dog toy, an airplane, etc.), and background objects (a fence, trees, the sky, etc.). *See* EX1007, 9, Fig. 10; *see also id.*, Tables 4-7 (listing several object categories used for identifying particular objects in images). Training a neural network on images of such objects lets the network learn to detect the categories of objects it has seen during training. Thus, when training a neural network to perform image recognition and edge detection on images of patient’s teeth, a POSA would understand it to be important to include images of patients’ teeth in the training set. Indeed, to most efficiently train such a neural network, the

training set should include images representative of the types of images likely to be seen during ordinary use of the imaging device, such as real historical photographs taken of patient's teeth with the device in which the trained network will be deployed (e.g., a cell phone as taught by Salah and Carrier).

151. Regarding the requirement that “each historical image compris[e] one or more zones each representing a tooth, or ‘historical tooth zones,’ to each of which, for said tooth attribute, a tooth attribute value is assigned,” a POSA would have performed this as well based on the teachings of Salah, Carrier, and Maninis. As discussed above for claim 2, a POSA would have been motivated to identify teeth in the image by using Maninis's object detection algorithms to assign tooth attributes to imaged teeth, including attributes such as tooth identification, patient age, diagnosis, and tooth shape parameters. In order to train a neural network to detect such attributes, the training set would be labeled with the attributes exhibited by objects depicted in each training image. Maninis teaches this practice, noting that its training set images contain “carefully localized pixelwise semantic annotations for the entire image,” which “results in 459 semantic categories” being assigned to objects depicted within the set of 10,103 images. EX1007, 5-6. Maninis's “pixelwise semantic annotations” are indications of which pixels correspond to each object in a given image, with the shape of a given object indicated by a change from one annotation to another at the boundary between

objects. Applying this same concept to a training set appropriate to the dental imaging application of Salah and Carrier, a POSA would apply annotations to historical dental images to indicate the attributes of each object in the image, including assigning tooth attributes to each tooth in each image.

152. Thus, Salah, Carrier, Maninis teach each part of step 1 of claim 3.

## **2. Step 3(2)**

153. Step 2 of claim 3 recites “training of at least one deep learning device by means of the learning base.” Salah, Carrier, and Maninis teach this element.

154. As discussed above for claim 1, Maninis teaches that its neural network is trained using a training dataset, which satisfies the meaning of “learning base” as used in the ’409 patent. *See, e.g.*, EX1007, 3-4, 7-8. In the context of Salah and Carrier, the learning base discussed above for step 3(1) containing annotated historical images of dental arches would be used to train a neural network like the one disclosed in Maninis. Thus, Salah, Carrier, and Maninis teach training “at least one deep learning device by means of a learning base.”

## **3. Step 3(3)**

155. Step 3 of claim 3 recites “submission of the analysis image to the deep learning device for it to determine at least one probability relating to: the presence, at a location in said analysis image, of a zone representing, at least partially, a

tooth, or ‘analysis tooth zone,’ the attribute value of the tooth represented on said analysis tooth zone.” Salah, Carrier, and Maninis teaches this element.

156. As I discussed above for claim 1, a trained neural network of the type taught by Maninis would be used to analyze images of patients’ dental arches as captured by the device of Salah and Carrier. This analysis would involve submitting the images to the trained neural network.

157. Furthermore, Maninis teaches the determination of probabilities related to “the presence at a location in said analysis image, of a zone representing, at least partially, a tooth, or ‘analysis tooth zone.’” For example, Maninis teaches that its contour detection procedure is built “on the concept of Ultrametric Contour Map (UCM), which transforms a contour detection probability map into a hierarchical boundary map, which gets partitions at different granularities when thresholding at various contour strength values.” EX1007, 5 (citation omitted). This framework involves producing “scale-specific contour images (side outputs) for different scales of a network, and combines their activations linearly to produce a contour probability map.” EX1007, 3. In essence, Maninis is describing a process that assigns a probability to locations within an image, with the assigned probabilities indicating the probability that a contour is located along a given set of pixels (a “contour probability map”). Maninis goes on to provide a set of equations by which the “probability  $P(\cdot)$  is obtained.” EX1007, 3. These probability maps are

then used to derive the contours most likely to be present in the image, using techniques adapted from prior techniques to transform “a contour detection probability map into a hierarchical boundary map.” EX1007, 5. Maninis describes the different probabilities as being treated as different “strengths” of boundaries, depicted in the figures as red lines of different thickness. See EX1007, 5, Figs. 3, 4-5 (depicting a simple portion of an image with three boundaries, having strengths of 0.6, 0.2, and 0.6). By comparing these “strengths” derived from the contour probability map, Maninis determines a “hierarchy of regions” dividing the different objects in an image. *Id.*

158. In the context of Salah and Carrier, this contour detection process involves determining a probability related to “the presence at a location in said analysis image, of a zone representing, at least partially, a tooth, or ‘analysis tooth zone,’” because the set of tooth edges define analysis tooth zone that has been determined to contain a particular tooth. Since Maninis’s neural network generates probabilities in the process of defining those edges, it generates probabilities related to the presence of the zone representing the tooth.

159. Lastly, with regard to “the attribute value of the tooth represented on said analysis tooth zone,” as discussed above for claims 1 and 2 (as well as the other steps of claim 3), in the combined system of Salah, Carrier, and Maninis, the detected teeth of a patient in an image would be assigned tooth attributes such as

tooth identification (e.g., by type or number) and shape (determined by the shape of the region containing the depiction of a given tooth). Accordingly, the tooth's attribute value would be "represented on said analysis tooth zone," as the attribute value would be represented both visually as a property of the image itself and by assignment of the value as part of the image recognition process.

#### **4. Step 3(4)**

160. Step 4 of claim 3 recites "determination, as a function of said probability, of the presence of a tooth at a position represented by said analysis tooth zone, and of the attribute value of said tooth." Salah, Carrier, and Maninis teach this element.

161. As discussed above for claims 1 and 2 (as well as the other steps of claim 3), in the combined system of Salah, Carrier, and Maninis, the presence of teeth would be detected within regions identified by a device employing Maninis's contour detection techniques. This corresponds to the application of Maninis's object detection techniques to Carrier's teaching to use "image recognition algorithms to ensure that all required teeth are visible." EX1006, ¶81; EX1005, ¶128. As discussed for previous step of claim 3, the detection of teeth within particular zones (i.e., regions defined by contours) is determined based on the recited probability. Likewise, the assignment of tooth attributes to the teeth within each zone would also occur "as a function of said probability," since Maninis

detects its regions using the probabilities it generates. Thus, Salah, Carrier, and Maninis teach this element of claim 3 as well.

162. Accordingly, it is my opinion that claim 3 would have been obvious over Salah, Carrier, and Maninis.

#### **E. Claim 4**

163. Claim 4 depends from claim 3 and further recites “in which said tooth attribute is chosen from a tooth number, a tooth type, a tooth shape parameter, a tooth appearance parameter, a parameter relating to the state of the tooth, an age for the patient, or a combination of these attributes.” Salah, Carrier, and Maninis teach this element.

164. As discussed above, Salah, Carrier, and Maninis teach assigning tooth attributes including multiple examples recited in claim 4. For example, these attributes would include a tooth number or tooth type (identifying the tooth to “ensure that all required teeth are visible” via the image recognition techniques taught by Maninis, EX1006, ¶81; EX1005, ¶128), as well as a tooth shape parameter (such as the contours defining the shape of the tooth). *See supra* section IX.C (discussing determination of each of these parameters in the combined system of Salah, Carrier, and Maninis). Thus, Salah, Carrier, and Maninis teach detection of tooth attributes chosen from among several categories recited in claim 4.

165. Accordingly, it is my opinion that claim 4 would have been obvious over Salah, Carrier, and Maninis.

#### **F. Claim 5**

166. Claim 5 depends from claim 3 and further recites that “in the step 1) or 1’), a learning base is created comprising more than 10 000 historical images.” Salah, Carrier, and Maninis teaches this element.

167. Claim 5 appears to have an error, as it refers to a step “1’),” yet none of claims 1-3 includes a step “1’).” Regardless, claim 3 includes a “step 1),” and a POSA would have found it obvious to include more than 10,000 historical images in a learning base for the same reasons as discussed above for claim 3. Indeed, Maninis itself expressly describes using more than 10,000 images in its training dataset. *See* EX1007, 1, 5-6. The use of 10,000 historical images of dental arches would simply be a routine and predictable application of Maninis’s teaching to use a sufficient number of training images (e.g., more than 10,000) to the context of Salah and Carrier’s system to train the system to detect tooth contours in images of patient’s oral cavities.

168. Accordingly, it is my opinion that claim 5 would have been obvious over Salah, Carrier, and Maninis.

### **G. Claim 6**

169. Claim 6 depends from claim 1 and further recites that “in the step b’), all said analysis tooth zones are identified, and at least one tooth attribute value is determined for each analysis tooth zone, and, in the step c’), the value for the image attribute is determined as a function of said tooth attribute values.” While claim 6 states that “all said analysis tooth zones are identified,” claim 1 does not recite analysis tooth zones, nor does claim 6 introduce a set of analysis tooth zones such that “all” of the set would be identified. Although claim 6 does not depend from claim 2, claim 2 refers to “at least one analysis tooth zone” being identified that “represents, at least partially, a tooth on said analysis image.” For the purposes of my analysis, I interpret claim 6 to be indicating that analysis tooth zones must be identified for all members of some set of teeth that is depicted in an image. For example, under this interpretation, identifying such zones for all teeth in an image would satisfy claim 6.

170. Salah, Carrier, and Maninis teaches claim 6. As I discussed above for claim 2, these references teach identifying one or more teeth on an image by detecting image contours, as taught by Maninis, to divide the image into distinct regions each corresponding to an object. In the context of Salah and Carrier’s dental images, these regions would each be an “analysis tooth zone” corresponding to a particular tooth that was depicted (at least partially) in the detected region.

Furthermore, Maninis divides the image into regions corresponding to each object. This can be seen, for example, in Fig. 9, where an image of grazing animals, trees, a fence, and other small objects is divided into distinct regions containing each such object:

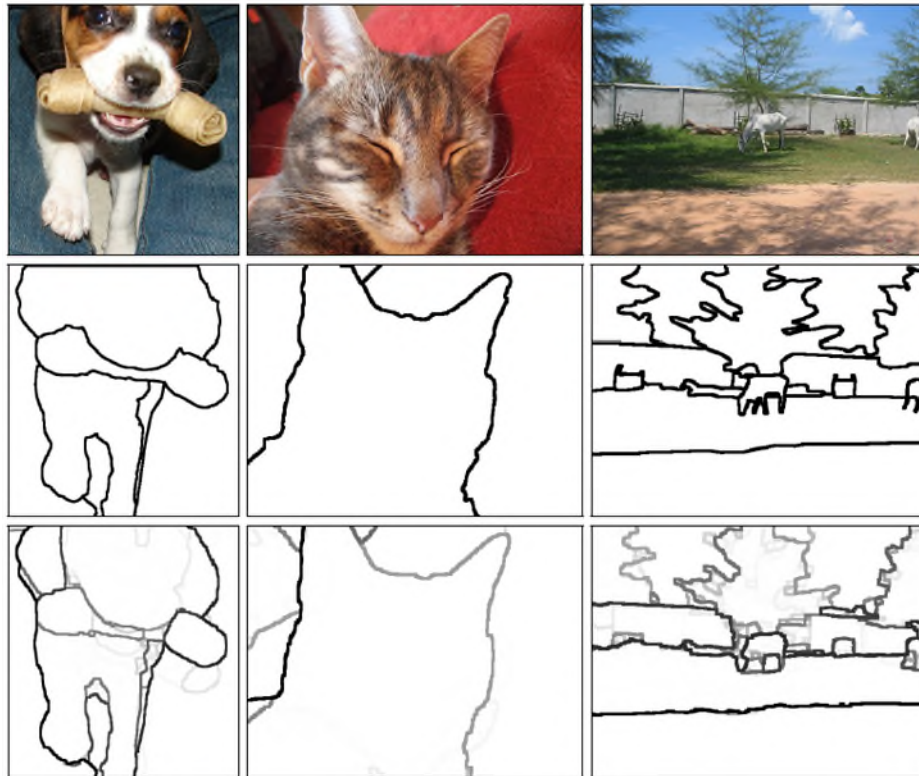


Fig. 9. **Qualitative results on PASCAL - Hierarchical Regions.** Row 1: original images, Row 2: ground-truth boundaries, Row 3: hierarchical regions with COB.

EX1007, Fig. 9. Applying this same mechanism to an imaged arch as taught by Salah and Carrier that contains a number of teeth would result in each tooth being identified in a region (or “analysis tooth zone”).

171. As I also discussed above for claim 2, Salah, Carrier, and Maninis teach assigning various tooth attributes to each detected tooth. *See supra* section

IX.C. Applying this to the multiple teeth that would be detected in a typical image of a dental arch, each tooth-containing region (“analysis tooth zone”) would have at least one tooth attribute determined.

172. Finally, Salah, Carrier, and Maninis teaches that the image attribute value is determined based on the tooth attribute values. *See supra* section IX.C. As I discussed for claims 1 and 2, Carrier teaches detecting edges of each tooth and ensuring that “all required teeth are visible,” which in the context of Salah and Maninis would both involve assigning corresponding tooth attributes to each tooth in its respective zone. EX1006, ¶¶63, 81; EX1005, ¶¶108, 128. As discussed for claim 1, these tooth attribute values are used to determine image attribute values, such as the distance between the teeth and the overlay (which depends on the teeth’s respective edges) and the presence of all identified teeth (which depends on the values indicating of which teeth are visible).

173. Accordingly, it is my opinion that claim 6 would have been obvious over Salah, Carrier, and Maninis.

#### **H. Claim 7**

174. Claim 7 depends from claim 1 and further recites that “the step b’) comprises” three further steps. As I discuss below, each of these further steps is taught by Salah, Carrier, and Maninis.

**1. Step 7(1')**

175. Step 1' of claim 7 recites “creation of a learning base comprising more than 1000 images of dental arches, or ‘historical images,’ each historical image comprising an attribute value for at least one image attribute, or ‘image attribute value.’” Salah, Carrier, and Maninis teach this element.

176. As discussed above for claim 3, a POSA would have been motivated Salah, Carrier, and Maninis to create a learning base of more than 1000 historical images of dental arches to train a neural network. These images would comprise attribute values for at least one image attribute. For example, Maninis teaches that each image in the training set includes pixelwise annotations identifying the objects contained within the image. *See* EX1007, 5-6. In the context of Salah and Carrier, a POSA would include images in a training dataset that were representative of the types of images that would be acquired using Salah and Carrier’s techniques.

177. For example, these images would include dental arch images acquired when following the process described by Salah and Carrier for acquiring such images with a mobile device. Each such image would correspond to a step of the patient dental image acquisition process and, based on Carrier’s teaching to ensure all required teeth are visible (EX1006, ¶81; EX1005, ¶128), each such image would have a set of required teeth. For example, an image taken from a perspective

to capture a single arch from above, like the one shown in Carrier Fig. 6B, might require each tooth in one arch, but not require any teeth in the other arch. Such an image, when annotated according to Maninis’s teachings, would contain annotations identifying whether all of the teeth required for that image were visible in the image. *See* EX1007, 5-6. This annotation of all required teeth would relate at least to the content of the image—i.e., whether the image contained all required teeth as taught by Carrier. *See supra* section IX.D (discussing annotations in the context of claim 3).

## **2. Step 7(2’)**

178. Step 2’ of claim 7 recites “training of at least one deep learning device by means of the learning base.” Salah, Carrier, and Maninis teach this element.

179. This element is the same as step 2 of claim 3, and there is no other difference between claims 3 and 7 to change the analysis I presented above for that claim. To reiterate, Maninis teaches that its neural network is trained using a training dataset, which satisfies the meaning of “learning base” as used in the ’409 patent. *See, e.g.*, EX1007, 3-4, 7-8. In the context of Salah and Carrier, the learning base discussed above for step 1 of claim 7 containing annotated historical images of dental arches would be used to train a neural network like the one disclosed in Maninis, thereby training “at least one deep learning device.”

### 3. Step 7(3')

180. Step 3' of claim 7 recites “submission of the analysis image to the deep learning device for it to determine, for said analysis image, at least one probability relating to said image attribute value.” Salah, Carrier, and Maninis teach this element.

181. As discussed above for step 3 of claim 3, Maninis determines boundaries between regions based on techniques that transform “a contour detection probability map into a hierarchical boundary map.” EX1007, 5. Thus applying Maninis’s technique to produce contours and identify objects involves generating probabilities that relate to boundaries between regions corresponding to the object depicted in each region of an image. *See supra* section IX.D.3 (discussing these probabilities in more detail). Because Maninis’s probabilities relate to the identification of boundaries between regions in the image being analyzed, these probabilities relate to the image attribute value of claim 1. For example, the probabilities relate to edges of each tooth, and thus to the distance between the edges of the tooth and the overlay. Further, because Maninis teaches identification of objects within the boundaries defined by those edges (*see* EX1007, 11-12 (discussing semantic boundaries), the probabilities would also relate to the teeth identified within those boundaries, as discussed above for claim 1.

182. Accordingly, it is my opinion that claim 7 would have been obvious over Salah, Carrier, and Maninis.

### **I. Claim 8**

183. Claim 8 depends from claim 1 and further recites that “the learning base is composed according to” 6 steps. Salah, Carrier, and Maninis teach each of these steps. The steps of claim 8 describe the use of a technique disclosed by Salah for registration of a 2D image to a segmented arch model to map individual teeth to different portions of the image. *See* EX1008, Fig. 1. Claim 8 uses the correspondences generated by the registration of the 2D image to the tooth model to assign attributes from the tooth model to the corresponding parts of the 2D image, which is then added to a learning base. As discussed below, this corresponds to Maninis’s teaching to include annotations reflecting the “ground truth” in training images, because the model (being an actual model of teeth) represents the “ground truth” of the teeth of the patient being imaged.

184. A POSA would be motivated to use Salah’s technique to identify the locations of particular teeth in images to be added to a training library for use in Maninis’s neural network, because doing so would enable the training dataset to contain “carefully localized pixel-wise semantic annotations” by providing a simple mechanism to connect regions of pixels in each image to the “ground-truth” represented by the tooth models that compose Salah’s reference model. EX1007, 2,

5-6. As Salah notes, it is known to create dental models for patients as an aspect of treatment, such as to manufacture orthodontic appliances. EX1008, 1:20-21. Using Salah's image-model registration technique would enable a POSA to leverage information contained in these dental models to label images for training by labeling photographs acquired in conditions similar to those contemplated by Salah and Carrier. *See, e.g.*, EX1008, 31:4-7 (tooth images may be acquired by patients' mobile phones).

185. I describe below in detail how each step of claim 8 is taught by Salah, Carrier, and Maninis.

**1. Step 8(A)**

186. Step A of claim 8 recites “at an ‘updated’ instant, production of a model of a dental arch of a patient, or ‘updated reference model,’ and segmentation of the updated reference model so as to produce, for each tooth, a ‘tooth model,’ and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model.” Salah, Carrier, and Maninis teach this element.

187. In particular, Salah teaches how to produce a model of a patient's dental arch, which is segmented to produce a tooth model for each tooth. Salah describes “creation of a three-dimensional digital reference model of at least part of an arch, preferably at least one of the patient's arches, or an ‘initial reference model’ and, preferably, for each tooth, definition, using the initial reference model,

of a three-dimensional digital reference model of said tooth, i.e. the ‘tooth model.’” EX1008, 1:28-32. Salah illustrates the process of creating the model, segmenting it, identifying each individual tooth, and providing a model for each such tooth. The initial model (prior to segmentation) is shown in Fig. 4a. See EX1008, 22:12-20 (“Figure 4a shows a view of the initial reference model processed to reveal these changes in direction.”), Fig. 4a (reproduced below).



Fig. 4a

EX1008, Fig. 4a.

188. After determining the edges for the teeth in the model in three dimensions, as well as determining gingival edges, Salah teaches that “it is thus possible to determine which parts of the initial reference model correspond to

which teeth ('tooth models')." EX1008, 23:4-8. The determining of which parts of a model correspond to the different teeth constitutes a segmentation of the model (i.e., a dividing of the model into parts (corresponding to teeth, for example)). Further, by assigning tooth models to each of the different teeth, Salah teaches producing "for each tooth, a 'tooth model,'" as recited in claim 8. In addition to describing these tooth models (*see id.*), Salah depicts a collection of tooth models in a segmented model of an entire dental arch in Fig. 4d, reproduced below:

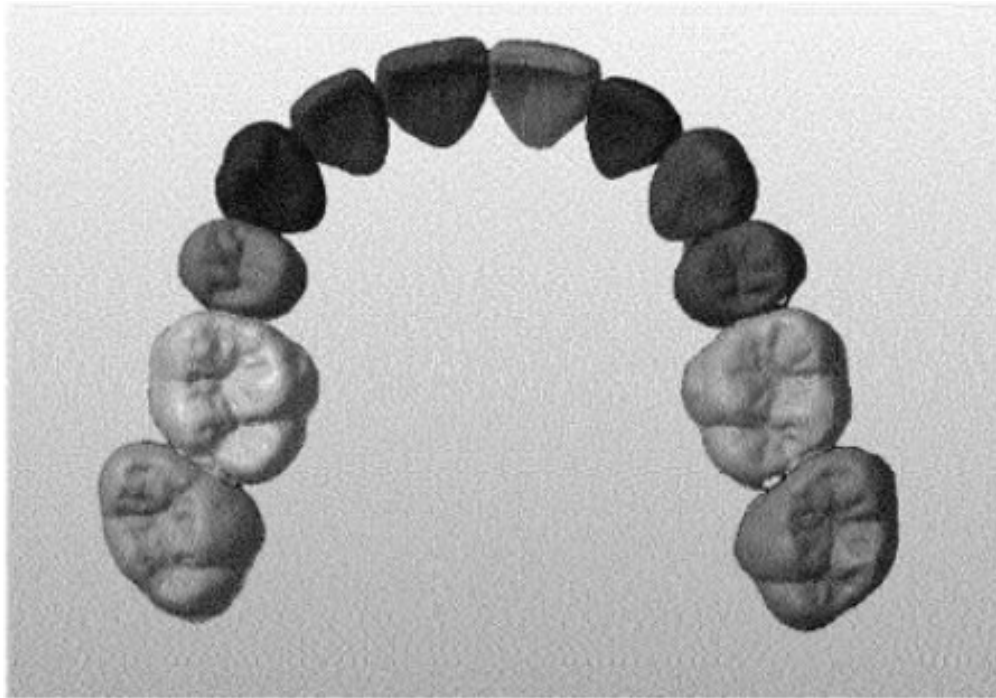


Fig. 4d

EX1008, Fig. 4d. As can be seen in the figure, each tooth has an individual tooth model (discernable via the different shadings used for adjacent teeth), confirming that Salah's process produces a reference model of a patient's dental arch and

segments to produce individual tooth models for each tooth. Indeed, the '409 patent describes Fig. 11D, which shows the same segmented arch as Fig. 4d of Salah, as being a segmented reference model. *See* EX1001, 7:9-16, 19:28-37.

189. Further, to the extent the term “updated” is interpreted to require that the model be a second model, Salah teaches that both an initial and an updated model may be produced. *See, e.g.*, EX1008, 51:22-24 (describing “comparison of the initial and updated reference models”), 9:1-3 (describing generation of updated “reference model to be tested [that] can be modified only by movement of one or more tooth models”), Fig. 1 (showing iterative process for updating and analyzing initial reference model and subsequent reference models to produce final reference model).

190. Lastly, the combination of Salah, Carrier, and Maninis teaches “assignment of a tooth attribute value to each tooth model.” To begin with, Salah’s generation of tooth models itself assigns tooth attribute values to the model. This is evident by inspection of Salah’s depiction of the models in Fig. 4d, showing different models (with different shape parameters) for different types of tooth, with molars (on the ends of the arch) clearly different from other teeth, such as incisors (in the center of the arch) or canines and premolars (between incisors and molars). A POSA would understand that when producing individual models of different teeth, the modeled teeth should be identified (e.g., by tooth number or type), since

different teeth have different shapes, locations, and physical properties. For example, US 2013/0282351 teaches that it was well known to identify three-dimensional tooth models with attributes such as tooth type, tooth number, and criteria such as age and morphological features. *See* EX1009, ¶¶7-8, 12.

## **2. Step 8(B)**

191. Step B of claim 8 recites “acquisition of at least one image of said arch, or ‘updated image,’ in real acquisition conditions.” Salah, Carrier, and Maninis teach this element.

192. As discussed above for claim 1, both Salah and Carrier teach acquiring images of arches in real acquisition conditions. *See, e.g.*, EX1008, 2:1-2, 31:4-7, 32:1-2, 43:5-9, Fig. 1; EX1006, ¶¶7, 59, 62-63, 94, Figs. 3A-B, 7B; EX1005, ¶¶8, 104, 107-08, 171, Figs. 3A-B, 7B; *supra* section 1(a’). Maninis illustrates that its training/validation images (as well as test images for which contour and object detection is performed) are also real images. *See, e.g.*, EX1007, Figs. 9-10 (showing real images drawn from a training set), 5-6 (images have “carefully localized pixel-wise semantic annotations for the entire image”).

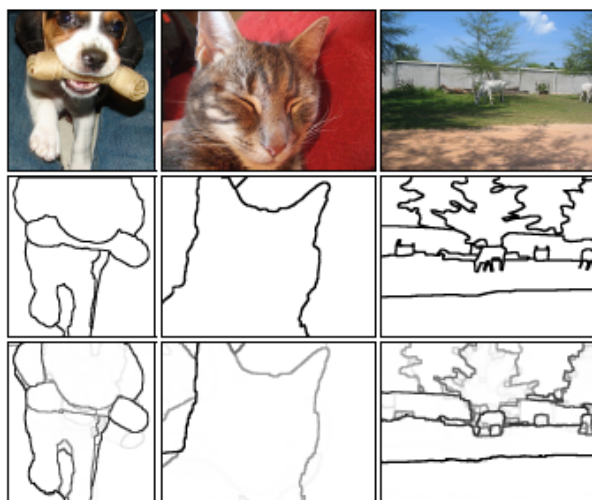


Fig. 9. **Qualitative results on PASCAL - Hierarchical Regions.** Row 1: original images, Row 2: ground-truth boundaries, Row 3: hierarchical regions with COB.

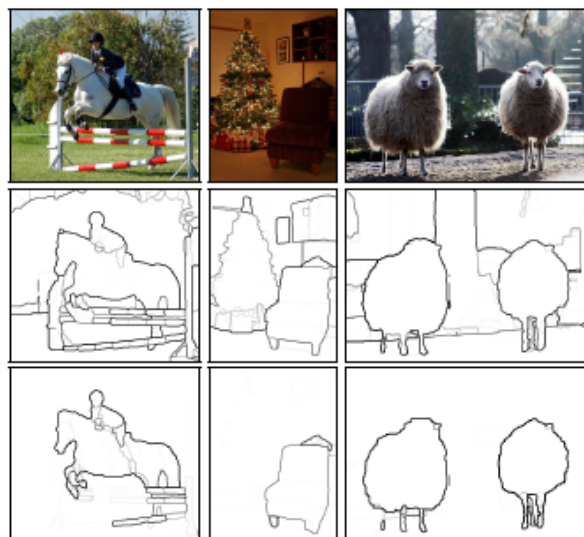


Fig. 10. **Qualitative results for Object Boundaries.** Row 1: original images, Row 2: Generic Image Segmentation results, Row 3: Object Boundary results.

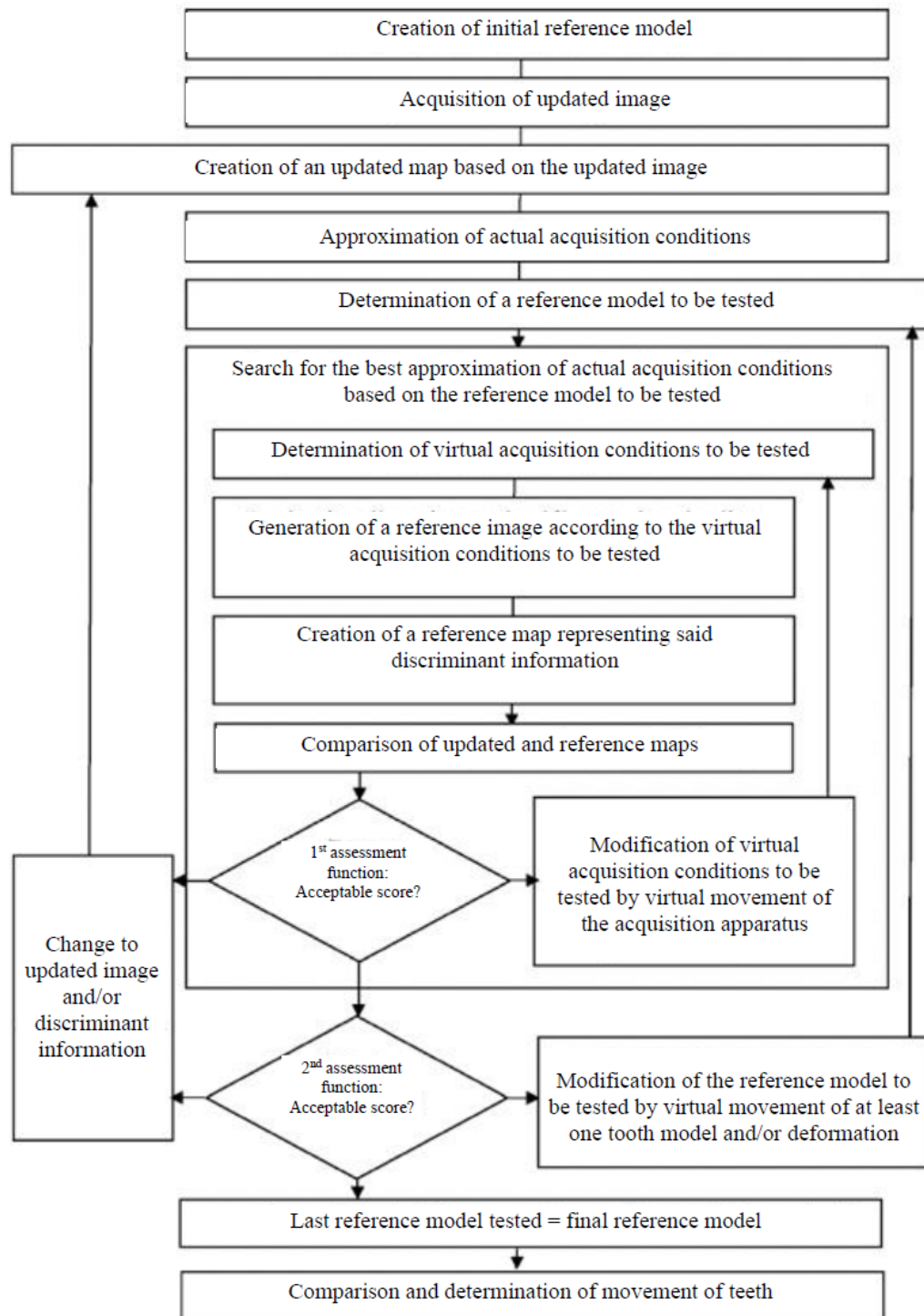
EX1007, Figs. 9-10.

193. A POSA would have recognized it as beneficial to acquire updated images to be included in a learning base for training a neural network as taught by Maninis, as using real acquisition conditions would provide more realistic and effective training. In particular, when applying Salah's teachings to annotate images for inclusion in a training set as taught in Maninis, a POSA would have understood that by obtaining images reflecting real acquisition conditions (e.g., acquired in the way expected to be seen in practice), this would provide for training reflective of the types of images that would need to be classified for real patients.

### **3. Step 8(C)**

194. Step C of claim 8 recites “for each updated image, search for virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called ‘reference image,’ exhibiting a maximum match with the updated image in said virtual acquisition conditions, and acquisition of said reference image.” Salah, Carrier, and Maninis teach this element.

195. In particular, Salah teaches that a 2D image (such as a mobile phone image) may be compared to a 3D model (such as a model of teeth) by “find[ing] the virtual acquisition conditions corresponding as closely as possible to real acquisition conditions in a reference model to be tested, determined based upon the initial reference model.” EX1008, 2:28-3:30. This search involves optimizing “the virtual position of the acquisition apparatus relative to the reference model to be tested, which offers the view of this reference model to be tested, i.e. the reference image, which is closest to the updated image.” EX1008, 48:9-19. This is part of the process of optimization shown in Fig. 1 below (which presents essentially the same optimization process shown in Fig. 16 of the ’409 patent).



EX1008, Fig. 1. Determining a reference image with a “closest” match to an updated image by searching for optimally close virtual acquisition conditions, followed by obtaining the reference image, constitutes “for virtual acquisition

conditions suitable for an acquisition of an image of the updated reference model, called ‘reference image,’ exhibiting a maximum match with the updated image in said virtual acquisition conditions, and acquisition of said reference image,” as recited in step C of claim 8.

#### **4. Steps 8(D)-(F)**

196. Steps 8(D), 8(E), and 8(F) are reproduced below:

D) identification, in the reference image, of at least one zone representing a tooth model, or “reference tooth zone” and, by comparison of the updated image and the reference image, determination, in the updated image, of a zone representing said tooth model, or “updated tooth zone”;

E) assignment, to said updated tooth zone, of the tooth attribute value of said tooth model;

F) addition of the updated image enriched with a description of said updated tooth zone and its tooth attribute value, or “historical image”, in the learning base.

These steps can be summarized as follows: in the updated image (being a real image of teeth) and the reference image (being a virtual image generated from Salah’s arch model to optimally match the updated image), matching zones containing a corresponding tooth model are identified. The tooth model’s tooth attribute value is then assigned to the zone in the updated image, which is then added to the learning base as a “historical image.”

197. Because Salah's dental model is composed of individual tooth models (EX1008, 23:4-8), identifying the zones representing the tooth model in the reference image would be straightforward, as the locations would be known for each tooth model used to generate the reference image from the reference model. Likewise, Salah's teaching that the reference image is one maximally resembling the updated image (derived from real acquisition conditions) implies that the identification of the corresponding tooth zones in the updated model would be straightforward as well. See EX1008, 48:9-19. In particular, the zones containing teeth in the reference image could be used (with manual or automatic adjustment as needed) to identify the tooth zones in the updated image. In fact, Salah teaches that tooth contours may be detected in an updated image, as shown in Fig. 6b, reproduced below

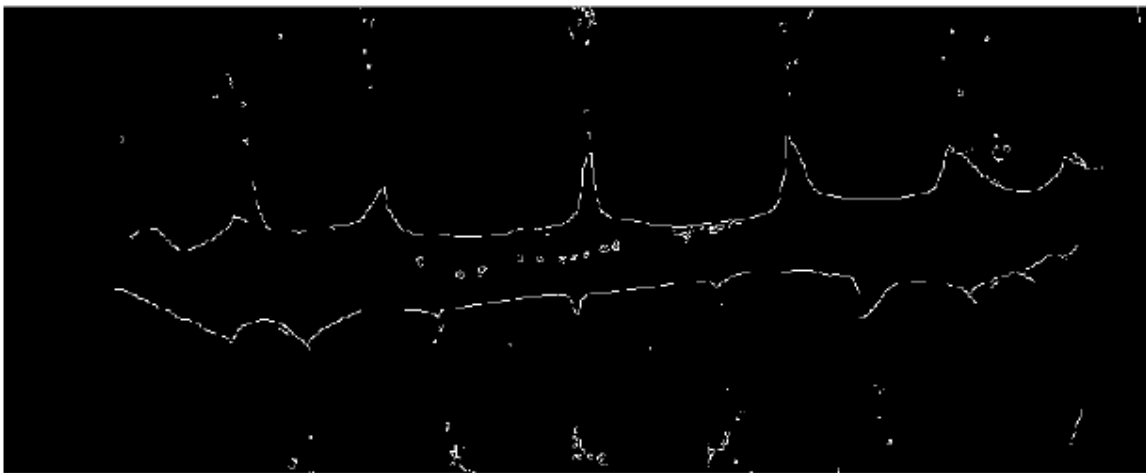


Fig. 6b

EX1008, Fig. 6b; *see id.*, 34:5-14. As Maninis teaches, these edges would correspond to boundaries of regions containing an object that may be detected (in this case, teeth). EX1007, 11-12.

198. A POSA would have understood that identifying the locations of these tooth zones would be important, as Maninis teaches that the training set should include pixelwise annotations of which regions contain which objects, consistent with the “human ground-truth.” EX1007, 2, 5-6. This identification could be done manually, with assistance of edge-detection algorithms (such as a pre-trained version of Maninis’s neural network or using prior edge-detection techniques), or it could be done using additional tools for registering tooth locations. For example, Salah further teaches that by identifying features (such as register marks) in an acquired image it is possible to “it is possible to identify the area of the updated image containing the elements that were the subject of the initial reference model, i.e. the teeth and gums.” EX1008, 33:27-34:2. Identifying the zones containing teeth in both the “reference” and “updated” image would serve this purpose, as it would aid in the accurate annotation of images to be added to a training dataset to be used for training of a neural network, as taught by Maninis, for contour and object detection of dental images for use in Salah and Carrier’s imaging apparatus.

199. This process of applying annotations would merely be an application of Salah’s teaching to register arch models with images to Maninis’s teaching to

provide annotations of all objects in an image to be included in a training dataset, according to known techniques for projecting annotations from a 3D model to a matching 2D image. *See, e.g.*, EX1013, 3689-90; EX1014, 3198, 3200-01. The result of this combination would be predictable: images would be added to a training dataset with annotations matching the ground-truth represented by Salah's model.

200. Accordingly, it is my opinion that claim 8 would have been obvious over Salah, Carrier, and Maninis

#### **J. Claim 9**

201. Claim 9 depends from claim 1 and further recites that "the learning base is composed according to" six steps, with the first three steps being similar to the first three steps of claim 8 and the last three steps being identical to the corresponding steps of claim 8. Salah, Carrier, and Maninis teaches each of these steps, as I discuss below.

##### **1. Step 9(A')**

202. Step A' of claim 9 recites "at an initial instant, production of a model of a dental arch of a patient, or 'initial reference model', and segmentation of the initial reference model so as to produce, for each tooth, a 'tooth model,' and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model." Salah, Carrier, and Maninis teach this element.

203. Step A' of claim 9 differs from step A of claim 8 only with regard to the replacement of the word "updated" in claim 8 with the word "initial" in claim 9. By referring to an "initial" model at an "initial" instant in step A' and an "updated" image at an "updated" instant in step B', claim 9 appears to require that steps A' and B' be performed in order. This is consistent with the teachings of Salah, Carrier, and Maninis. Indeed, Salah teaches that both initial and updated models may be generated and compared with updated images. *See, e.g.*, EX1008, 1:28-2:5 (describing "creation of a three-dimensional digital reference model of at least part of an arch, preferably at least one of the patient's arches, or an 'initial reference model' and, preferably, for each tooth, definition, using the initial reference model, of a three-dimensional digital reference model of said tooth, i.e. the 'tooth model' ... acquisition of at least one two-dimensional image of the patient's arches, the so-called 'updated image,' under actual acquisition conditions ... analysis of each updated image ..."); *see also id.*, 23:4-8, Fig. 4d. The teachings I discuss above for step A of claim 8 thus apply equally to step A' of claim 9, and Salah, Carrier, and Maninis teach this step for similar reasons as discussed above for step A of claim 8.

## **2. Step 9(B')**

204. Step B' of claim 9 recites “at an updated instant, acquisition of at least one image of said arch, or ‘updated image,’ in real acquisition conditions.” Salah, Carrier, and Maninis teach this element.

205. Step B' of claim 9 differs from step B of claim 8 only by reciting “at an updated instant” in step B', whereas the “updated instant” of claim 8 is recited in step A. As discussed above for step B', the recitation of two different instants (initial and updated) does not alter my conclusions regarding the combined teachings of Salah, Carrier, and Maninis. Indeed, Salah describes using an “initial” model in combination with an “updated” image. EX1008, 1:28-2:5. The teachings I discuss above for step B of claim 8 thus apply equally to step B' of claim 9, and Salah, Carrier, and Maninis teach this step for similar reasons as discussed above for step B of claim 9.

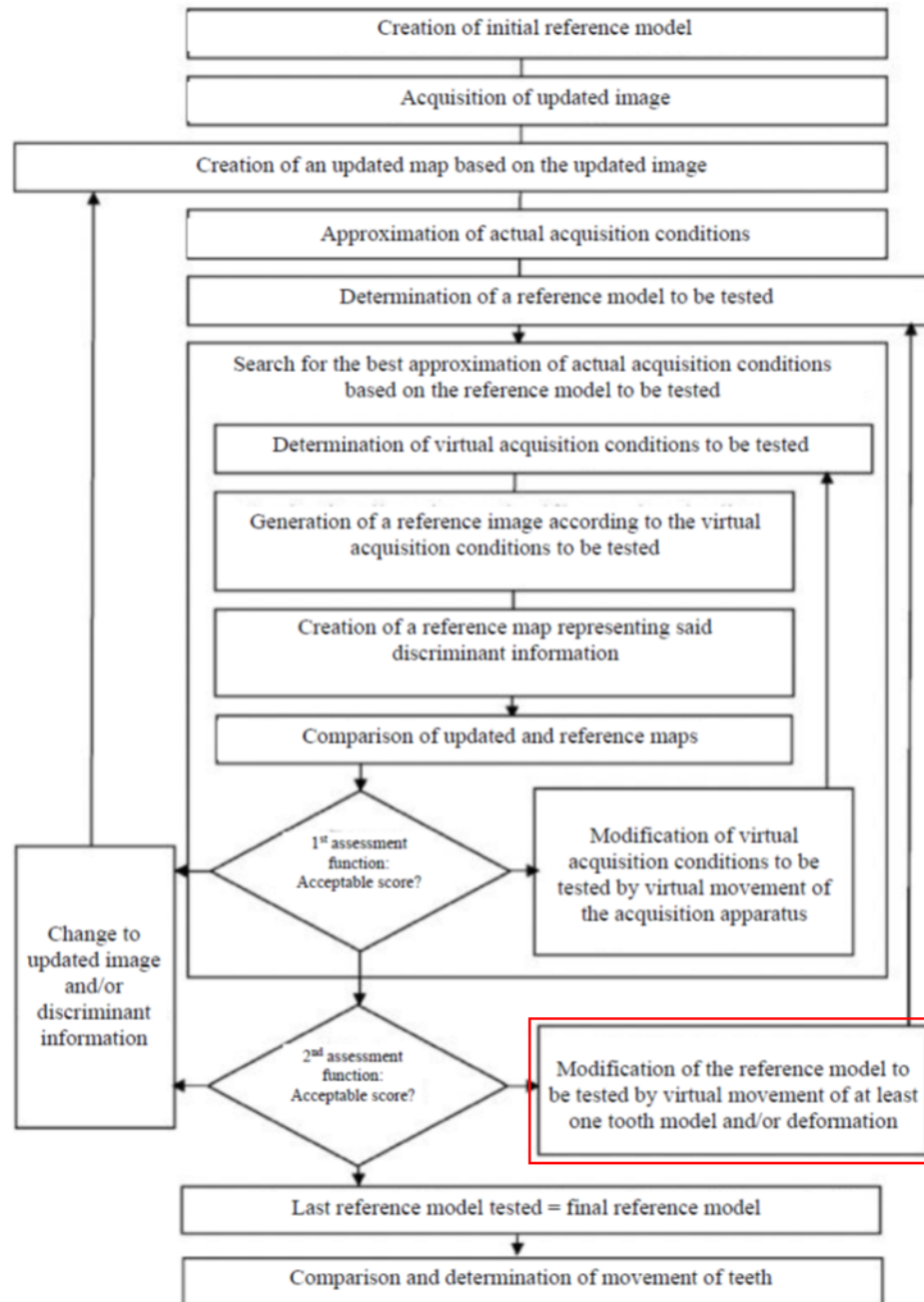
## **3. Step 9(C')**

206. Step C' of claim 9 recites “search, by deformation of the initial reference model, for an updated reference model and virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called ‘reference image,’ exhibiting a maximum match with the updated image in said virtual acquisition conditions.” Salah, Carrier, and Maninis teach this element.

207. Step C' of claim 9 differs from step C of claim 8 in two aspects. First, claim 9 omits the requirements that searching is done “for each updated image” and that “acquisition of said reference image” be performed. Second, claim 9 adds a requirement that search be performed “by deformation of the initial reference model, for an updated reference model.” With regard to the omitted portions, their omission does not affect my analysis as discussed above for step C of claim 8. With regard to deformation, Salah teaches searching for updated reference models by deformation, as recited in claim 9.

208. Specifically, Salah teaches that when searching for virtual acquisition conditions for a reference model to best match an updated image, an updated model may be generated by deformation, such as by adjustment of individual modeled teeth. Salah describes that its sequence of repeated optimizations of a reference model, including the steps discussed above for claim 8, may further include steps in which a “final reference model to be tested is deformed, the first two optimization operations are repeated, and then, depending on the match obtained, the search is stopped or continued by performing a new deformation of the final reference model to be tested, and by running the first two optimization operations.” EX1008, 49:20-24. This step is illustrated in Fig. 1 of Salah as being part of the matching process between an updated image and a reference model that

I discussed above for claim 8. This figure is shown below with the deformation step marked by a red box:



EX1008, Fig. 1 (annotated). Salah explains that performing such deformations as part of an optimization procedure allows generation of “a a final reference model

that takes into account both tooth displacement and tooth deformation.” EX1008, 49:5-7. In other words, Salah teaches that differences between a model and a subsequent state of teeth at a time of an updated image (e.g., due to changes in a tooth’s position) may be accounted for via including its model deformation process in the process discussed above for claim 8, step C.

209. Thus, as Salah expressly teaches that a deformation procedure as recited in claim 9 may be employed as part of its process for generating a reference image in optimized virtual acquisition conditions, Salah, Carrier, and Maninis teach step 9(C’).

#### **4. Steps 9(D)-(F)**

210. Steps 9(D), 9(E), and 9(F) are identical to steps 8(D), 8(E), and 8(F), respectively. And as I discussed above for steps 9(A’)-(C’), those steps have only minimal differences from steps 8(A)-(C), producing a similar correspondence between a model, reference image, and updated image, with the only added element being the use of deformation on the reference model, which Salah teaches may be part of the same process discussed for steps 8(A)-(C). The analysis I provided above with respect to steps 8(D)-(F) above is thus equally applicable to this step of claim 9. Thus, Salah, Carrier, and Maninis teach steps 9(D)-(F).

211. Accordingly, it is my opinion that claim 9 would have been obvious over Salah, Carrier and Maninis.

## **K. Claim 10**

212. Claim 10 depends from claim 1, reciting as follows:

The method as claimed in claim 1, in which, to create a historical image of the learning base,

identifies one or more “historical” tooth zones on an image, called then assigns, to each identified historical tooth zone, a value for at least one tooth attribute, and/or

assigns to an image a value for at least one tooth attribute.

213. Claim 10 does not expressly describe who or what performs the identifying or assigning steps; however, in this analysis, I will assume that these steps are performed by the creator of the historical image. Likewise, the phrase “called then assigns” in “identifies” element of claim 10 is unclear; however, the claim only requires that the “identifies” step “and/or” the “assigns” step be satisfied. Thus, I understand that claim 10 may be obvious if it would have been obvious to practice the “assigns” element, regardless of the meaning of the “identifies” element. In other words, I understand that claim 10 may be satisfied if, to create a historical image of the learning base, the creator of that image “assigns to an image a value for at least one tooth attribute” Salah, Carrier, and Maninis teach this.

214. In particular, Maninis teaches using a training set containing images with “carefully localized pixel-wise semantic annotations for the entire image.”

EX1007, 5-8. As I discuss above in section IX.D, in the context of Salah and

Carrier, these annotations would be indications, for each pixel of the image, of the object contained in that image, such as a tooth identified by tooth attributes tooth number or type. These annotations would thus indicate the region in which each tooth (among other dental structures like lips, tongue, etc.) was present in the image and the identifying attributes of that tooth. The identification (via pixelwise annotations) of the regions containing teeth would thus constitute an identification of historical tooth zones on the image, and the assignment of tooth attributes in these annotations would constitute an assignment to the zone a value for at least one tooth attribute. Further, the assignment to each zone would at the same time constitute an assignment to the image of a value for at least one tooth attribute, since each tooth-containing region would be located in the image. Thus, at a minimum, Salah, Carrier, and Maninis teach “assign[ing] to an image a value for at least one tooth attribute” to create a historical image of the learning base, as recited in claim 10.

215. Accordingly, it is my opinion that claim 10 would have been obvious over Salah, Carrier, and Maninis.

#### **L. Claim 11**

216. Claim 11 depends from claim 1 and further recites that “the acquisition apparatus is a cellphone.” Salah and Carrier teach this element.

217. Salah and Carrier both repeatedly teach that their image acquisition may be a cellphone. For example, Salah teaches that “the acquisition apparatus is a cell phone.” EX1008, 12:7; *see also id.*, 31:4-7 (“In particular, image acquisition can be carried out by the patient himself or by one of their relatives, using a simple cell phone, anywhere, and especially outside a medical, dental or orthodontic practice.”). Carrier also teaches that its image acquisition apparatus is a cellphone. Each of Figs. 2A-2C and 5A-7C shows the image of a smartphone screen being used as an image acquisition apparatus performing Carrier’s imaging methods. Carrier describes Fig. 2A as showing “an example of a screen of an apparatus (such as a smartphone running the app as described herein).” EX1006, ¶30; *see id.*, ¶62 (“It may be particularly helpful to adapt a traditional handheld consumer electronics device, such as phone (e.g., smartphone, smartwatch, pad, tablet, etc.) to take one or, more preferably, a series of images of the teeth....”); EX1005, ¶97 (same).

218. Accordingly, it is my opinion that claim 11 would have been obvious over Salah, Carrier, and Maninis.

### **M. Claim 12**

219. Claim 12 depends from claim 1 and further recites that “the information message is sent by the acquisition apparatus.” Salah and Carrier both teach that the information message is sent by the acquisition apparatus.

220. Each information messages disclosed by Salah and Carrier in the examples discussed above for claim 1(e') is sent by the acquisition apparatus. For example, Salah discloses that the apparatus can provide “written and/or vocal prompts” to guide the user, as well as displaying arrows or beeping to indicate positioning. EX1008, 27:21-28:2. Carrier similarly teaches that the device can “indicate (e.g., by visual and/or audible and/or tactile) image that the teeth are aligned in the frame with the overlay.” EX1006, ¶63; EX1005, ¶108. Carrier also provides examples involving a message displayed on the acquisition apparatus’s screen, including arrows like Salah’s shown on the screen to indicate a direction or amount to move the mobile phone, as well as changing the color of its overlay. EX1006, ¶¶69-70; EX1005, ¶¶114-15.

221. Accordingly, it is my opinion that claim 12 would have been obvious over Salah, Carrier, and Maninis.

#### **N. Claim 13**

222. Claim 13 depends from claim 1 and further recites that “said image attribute relates to” one of several categories of image attribute. These recited image attributes are as follows:

a position or an orientation or a calibration of an acquisition apparatus used to acquire said analysis image, or a combination thereof, and/or

a quality of the analysis image, and in particular relating to the brightness, to the contrast or to the sharpness of the analysis image, and/or

the content of the analysis image, in particular to the representation of the arches, of the tongue, of the mouth, of the lips, of the jaws, of the gums, of one or more teeth or of an orthodontic appliance.

Salah, Carrier, and Maninis teaches an image attribute relating to one or more of these categories.

223. With regard to “a position or an orientation or a calibration of an acquisition apparatus used to acquire said analysis image,” Carrier teaches that edges of teeth may be detected and that a distance may be measured between the edges of teeth and an overlay. EX1006, ¶¶63, 68; EX1005, ¶¶108, 114. As discussed above for claim 1, this measurement constitutes the determination of a value for an image attribute—the relative position of the tooth to the overlay. This attribute reflects the degree of alignment of the image acquisition apparatus, and it therefore relates to a position or an orientation of the apparatus. Indeed, Carrier uses this measurement to determine whether the apparatus is aligned. *See* EX1006, ¶¶63, 68; EX1005, ¶¶108, 114. Thus, Carrier itself confirms that its image value attribute relates to the position or orientation of the acquisition apparatus.

224. Carrier also teaches detection of values the content of the image, including representation of, among other options, “the arches” or “one or more teeth.” As I discussed above for claim 1, Carrier discloses that the apparatus may

use “image recognition algorithms to ensure that all required teeth are visible” (EX1006, ¶81; EX1005, ¶128), which a POSA would accomplish by using Maninis’s object detection to generate an image attribute value representing whether the required teeth (such as an arch, or a subset of specified teeth) were all visible. *See supra* section IX.B.4.

225. Accordingly, it is my opinion that claim 13 would have been obvious over Salah, Carrier, and Maninis.

#### **O. Claim 14**

226. Claim 14 depends from claim 1 and further recites that “said analysis image is analyzed and checked in real time.” Salah, Carrier, and Maninis teach this element.

227. Carrier discloses that its image analysis and the subsequent checking (e.g., comparison to setpoints) is conducted in real time. Carrier discloses that its methods “include the use of an ‘overlay’ on a real-time image of the screen, providing immediate feedback on each of the desired orientations.” EX1006, ¶54; EX1005, ¶99. Carrier further discloses that the device provides “automatic detection of alignment in real-time with a patient’s teeth and an overlay.” EX1006, ¶63; EX1005, ¶108. The providing of a real-time image and overlay, in conjunction with immediate feedback based on automatic, real-time detection of

alignment, indicates that Carrier's image analysis and checking is performed in real time.

228. The combination of Salah and Maninis with Carrier would be consistent with Carrier's real-time analysis and feedback. Like Carrier, Salah provides real-time feedback on alignment, so performing real-time analysis as taught by Carrier would be consistent with Salah's feedback techniques. *See* EX1008, 27:21-28:2. As for Maninis, it discloses deep learning contour detection algorithms that produce fast, efficient, and highly-accurate results. EX1007, 1. For example, even when classifying images selected to be "challenging and varied," Maninis discloses that its image analysis required an average of just under 800 ms. EX1007, 1; *see id.*, 9 (reporting "significant improvement" over state-of-the-art techniques, reducing total computation time from "24.37 to 0.79 seconds"), Table 1 (comparing results). Thus, a POSA would reasonably expect, based on the combined teachings of Salah, Carrier, and Maninis, that when using a neural network as taught by Maninis as part of Carrier's image analysis, the resulting device would conduct its image analysis and checking in real time.

229. Accordingly, it is my opinion that claim 14 would have been obvious over Salah, Carrier, and Maninis.

## P. Claim 15

230. Claim 15 depends from claim 1 and further recites that “the deep learning device is a neural network.” Salah, Carrier, and Maninis teach this element.

231. As discussed above for claim 1, Maninis teaches that its image analysis for contour and object detection is performed using a convolutional neural network (CNN). EX1007, Abstract. Indeed, Maninis teaches that the use of a CNN is the core of “deep learning approaches” like its own, and it depicts using a single-pass CNN as the first element of its analysis procedure. EX1007, Abstract, Fig. 1 (reproduced below); *see also id.*, 1-2, 13, Fig. 2.

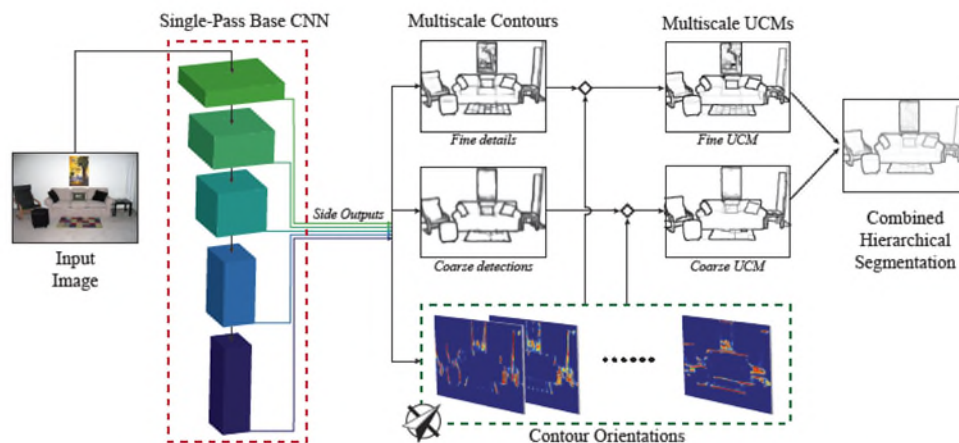


Fig. 1. **Overview of COB:** From a single pass of a base CNN, we obtain multiscale oriented contours. We combine them to build Ultrametric Contour Maps (UCMs) at different scales and fuse them into a single hierarchical segmentation structure.

EX1007, Fig. 1.

232. Accordingly, it is my opinion that claim 15 would have been obvious over Salah, Carrier, and Maninis.

233. For the reasons I have provided above, it is my opinion that each of claims 1-15 of the '409 patent would have been obvious over Salah, Carrier, and Maninis.

**X. CONCLUDING STATEMENTS**

234. In signing this declaration, I understand that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I acknowledge that I may be subject to cross-examination in this case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

235. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: August 31, 2023

By: / Hassan Foroosh, Ph.D. /  
Hassan Foroosh, Ph.D.

## **XI. APPENDIX – LISTING OF CLAIMS**

1. A method for acquiring an image of a dental arch of a patient, said method comprising the following steps:

a') activation of an image acquisition apparatus so as to acquire an image, called “analysis image”, of said arch;

b') analysis of the analysis image by means of a deep learning device trained by means of a learning base;

c') determination, for the analysis image, as a function of the results of the analysis in the preceding step, of a value for an image attribute;

d') comparison of said image attribute value with a setpoint;

e') sending of an information message as a function of said comparison, the information message being related to the quality of the image acquired or to the position of the acquisition apparatus in relation to said arch or to the setting of the acquisition apparatus or to the opening of the mouth or to the wearing of a dental appliance, or to a combination thereof,

to check whether the analysis image respects the setpoint and, if it does not respect the setpoint, to guide the operator in order for him or her to acquire a new analysis image.

2. The method as claimed in claim 1, in which, in the step b'), at least one analysis tooth zone is identified that represents, at least partially, a tooth on said analysis image, and at least one value of a tooth attribute is determined for said at least one analysis tooth zone, and, in the step c'), the value for the image attribute is determined as a function of said tooth attribute value.

3. The method as claimed in claim 2, in which the step b') comprises the following steps:

1) creation of a learning base comprising more than 1000 images of dental arches, or “historical images”, each historical image comprising one or more zones each representing a tooth, or “historical tooth zones”, to each of which, for said tooth attribute, a tooth attribute value is assigned;

2) training of at least one deep learning device by means of the learning base;

3) submission of the analysis image to the deep learning device for it to determine at least one probability relating to:

the presence, at a location in said analysis image, of a zone representing, at least partially, a tooth, or “analysis tooth zone”,

the attribute value of the tooth represented on said analysis tooth zone,

4) determination, as a function of said probability, of the presence of a tooth at a position represented by said analysis tooth zone, and of the attribute value of said tooth.

4. The method as claimed in claim 3, in which said tooth attribute is chosen from a tooth number, a tooth type, a tooth shape parameter, a tooth appearance parameter, a parameter relating to the state of the tooth, an age for the patient, or a combination of these attributes.

5. The method as claimed in claim 3, in which, in the step 1) or 1'), a learning base is created comprising more than 10 000 historical images.

6. The method as claimed in claim 1, in which, in the step b'), all said analysis tooth zones are identified, and at least one tooth attribute value is determined for each analysis tooth zone, and, in the step c'), the value for the image attribute is determined as a function of said tooth attribute values.

7. The method as claimed in claim 1, in which the step b') comprises the following steps:

1') creation of a learning base comprising more than 1000 images of dental arches, or “historical images”, each historical image comprising an attribute value for at least one image attribute, or “image attribute value”;

2') training of at least one deep learning device by means of the learning base;

3') submission of the analysis image to the deep learning device for it to determine, for said analysis image, at least one probability relating to said image attribute value.

8. The method as claimed in claim 1, in which the learning base is composed according to the following steps:

A) at an “updated” instant, production of a model of a dental arch of a patient, or “updated reference model”, and segmentation of the updated reference model so as to produce, for each tooth, a “tooth model”, and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model;

B) acquisition of at least one image of said arch, or “updated image”, in real acquisition conditions;

C) for each updated image, search for virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called “reference image”, exhibiting a maximum match with the updated image in said virtual acquisition conditions, and acquisition of said reference image;

D) identification, in the reference image, of at least one zone representing a tooth model, or “reference tooth zone” and, by comparison of the updated image and the reference image, determination, in the updated image, of a zone representing said tooth model, or “updated tooth zone”;

E) assignment, to said updated tooth zone, of the tooth attribute value of said tooth model;

F) addition of the updated image enriched with a description of said updated tooth zone and its tooth attribute value, or “historical image”, in the learning base.

9. The method as claimed in claim 1, in which the learning base is composed according to the following steps:

A') at an initial instant, production of a model of a dental arch of a patient, or “initial reference model”, and segmentation of the initial reference model so as to produce, for each tooth, a “tooth model”, and, for at least one tooth attribute, assignment of a tooth attribute value to each tooth model;

B') at an updated instant, acquisition of at least one image of said arch, or “updated image”, in real acquisition conditions;

C') search, by deformation of the initial reference model, for an updated reference model and virtual acquisition conditions suitable for an acquisition of an image of the updated reference model, called “reference image”, exhibiting a maximum match with the updated image in said virtual acquisition conditions;

D) identification, in the reference image, of at least one zone representing a tooth model, or “reference tooth zone” and, by comparison of the updated image and the reference image, determination, in the updated image, of a zone representing said tooth model, or “updated tooth zone”;

E) assignment, to said updated tooth zone, of the tooth attribute value of said tooth model;

F) addition of the updated image enriched with a description of said updated tooth zone and its tooth attribute value, or “historical image”, in the learning base.

10. The method as claimed in claim 1, in which, to create a historical image of the learning base,

identifies one or more “historical” tooth zones on an image, called then assigns, to each identified historical tooth zone, a value for at least one tooth attribute, and/or

assigns to an image a value for at least one tooth attribute.

11. The method as claimed in claim 1, in which the acquisition apparatus is a cellphone.

12. The method as claimed in claim 1, in which the information message is sent by the acquisition apparatus.

13. The method as claimed in claim 1, in which said image attribute relates to

a position or an orientation or a calibration of an acquisition apparatus used to acquire said analysis image, or a combination thereof, and/or

a quality of the analysis image, and in particular relating to the brightness, to the contrast or to the sharpness of the analysis image, and/or

the content of the analysis image, in particular to the representation of the arches, of the tongue, of the mouth, of the lips, of the jaws, of the gums, of one or more teeth or of an orthodontic appliance.

14. The method as claimed in claim 1, in which said analysis image is analyzed and checked in real time.

15. The method as claimed in claim 1, in which the deep learning device is a neural network.

## XII. APPENDIX – EXHIBITS

| Exhibit No. | Description                                                                                                                                                           |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001        | U.S. Patent No. 10,755,409 to Salah et al.                                                                                                                            |
| 1003        | Curriculum Vitae of Hassan Foroosh, Ph.D.                                                                                                                             |
| 1004        | U.S. Patent Application No. 16/031,201 File History                                                                                                                   |
| 1005        | U.S. Patent Pub. No. 2021/0068923 to Carrier et al.                                                                                                                   |
| 1006        | Provisional Patent Application No. 62/417,985 to Carrier et al.                                                                                                       |
| 1007        | Maninis et al., “Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks,” <i>arXiv.org</i> (Apr. 28, 2017)                                    |
| 1008        | Translation of PCT Pub. No. WO 2016066651 to Salah et al.                                                                                                             |
| 1009        | U.S. Patent Pub. No. 2013/0282351 to Martin Tank                                                                                                                      |
| 1010        | U.S. Patent Pub. No 2020/0229901 (Application No. 16/827,594)                                                                                                         |
| 1011        | U.S. Patent Pub. No. 2018/0125610 (Application No. 15/803,718)                                                                                                        |
| 1012        | PCT Pub. No. WO 2016066651 to Salah et al.                                                                                                                            |
| 1013        | Xie et al., “Semantic Instance Annotation of Street Scenes by 3D to 2D Label Transfer,” <i>2016 IEEE Conference on Computer Vision and Pattern Recognition</i> (2016) |
| 1014        | Chen et al., “Beat the MTurkers: Automatic Image Labeling from Weak 3D Supervision,” <i>2014 IEEE Conference on Computer Vision and Pattern Recognition</i> (2014)    |
| 1015        | Krizhevsky et al., “ImageNet Classification with Deep Convolutional Neural Networks,” <i>Advances in Neural Information Processing Systems</i> 25 (2012)              |
| 1016        | Patrick Cavanaugh, “Visual Cognition” <i>Vision Research</i> 51 (2011)                                                                                                |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1017 | Prasad et al., “Learning Class-specific Edges for Object Detection and Segmentation,” <i>Computer Vision, Graphics and Image Processing: 5th Indian Conference, ICVGIP 2006</i> (2006)                                                                                                                                                                                                                                       |
| 1018 | Wang et al., “Simultaneous Image Classification and Annotation,” <i>CVPR 1</i> , No. 2 (2009)                                                                                                                                                                                                                                                                                                                                |
| 1019 | Tao et al., “Similarity Voting based Viewpoint Selection for Volumes,” <i>Computer Graphics Forum</i> 35, No. 2 (2016)                                                                                                                                                                                                                                                                                                       |
| 1021 | Maninis et al., “Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks,” <i>arXiv.org</i> (Jan. 17, 2017)                                                                                                                                                                                                                                                                                           |
| 1022 | <i>arXiv.org Submission Help Page</i> , <a href="https://arxiv.org/help/submit">https://arxiv.org/help/submit</a> (as archived May 3, 2017, at <a href="https://web.archive.org/web/20170503071322/https://arxiv.org/help/submit">https://web.archive.org/web/20170503071322/https://arxiv.org/help/submit</a> )                                                                                                             |
| 1023 | <i>arXiv.org Article Replacement Help Page</i> , <a href="https://arxiv.org/help/replace">https://arxiv.org/help/replace</a> (as archived July 18, 2017, at <a href="https://web.archive.org/web/20170718173822/https://arxiv.org/help/replace">https://web.archive.org/web/20170718173822/https://arxiv.org/help/replace</a> )                                                                                              |
| 1024 | <i>COB - Convolutional Oriented Boundaries</i> , <a href="http://people.ee.ethz.ch/~cvlsegmentation/cob/">http://people.ee.ethz.ch/~cvlsegmentation/cob/</a> (as archived May 3, 2017, at <a href="http://web.archive.org/web/20170503233424/http://people.ee.ethz.ch/~cvlsegmentation/cob/">http://web.archive.org/web/20170503233424/http://people.ee.ethz.ch/~cvlsegmentation/cob/</a> )                                  |
| 1025 | <i>arXiv.org Author Search Page for Luc van Gool</i> , <a href="https://arxiv.org/find/cs/1/au:+Gool_L/0/1/0/all/0/1">https://arxiv.org/find/cs/1/au:+Gool_L/0/1/0/all/0/1</a> (as archived July 8, 2017, at <a href="https://web.archive.org/web/20170708002617/https://arxiv.org/find/cs/1/au:+Gool_L/0/1/0/all/0/1">https://web.archive.org/web/20170708002617/https://arxiv.org/find/cs/1/au:+Gool_L/0/1/0/all/0/1</a> ) |
| 1026 | <i>arXiv.org Article Replacement Help Page</i> , <a href="https://arxiv.org/help/replace">https://arxiv.org/help/replace</a> (as archived September 8, 2016, at <a href="https://web.archive.org/web/20160908205209/https://arxiv.org/help/replace">https://web.archive.org/web/20160908205209/https://arxiv.org/help/replace</a> )                                                                                          |