

Petitioner's Demonstratives

Align Technology, Inc. v. Dental Monitoring SAS

IPR2024-00052 & IPR2024-00110

U.S. Patent Nos. 11,049,248 & 11,109,945

Instituted Grounds

'248 Patent:

Claims	35 U.S.C. §	Prior Art
1, 2, 4, 7-8, 13-16	103	Kim, Invisalign, Salah, and Maninis

'945 Patent:

Claims	35 U.S.C. §	Prior Art
1-2, 4-6	103	Kim, Invisalign, and Salah

The '248 Patent

(57) ABSTRACT

A method for assessing the shape of an orthodontic aligner. Acquisition of at least one analysis image at least partially representing the aligner in a service position in which it is worn by a patient. Analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value. For at least one tooth attribute of an analysis tooth zone representing, at least partially, a tooth on the analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image. For an image attribute of the analysis image, the image attribute relating to a separation between at least one tooth represented on the analysis image, and the aligner represented on the analysis image.

TECHNICAL FIELD

The present invention relates to the field of dental arch image analysis.

The invention proposes a method for analyzing an image, called "analysis image", of a dental arch of a patient, a method in which the analysis image is submitted to a deep learning device, preferably a neural network, in order to determine at least one value of a tooth attribute relating to a tooth represented on the analysis image, and/or at least one value of an image attribute relating to the analysis image.

A "neural network" or "artificial neural network" is a set of algorithms well known to a person skilled in the art.

The neural network may in particular be chosen from: the networks specializing in the classification of images, called "CNN" ("convolutional neural network"), for example

AlexNet (2012)

ZF Net (2013)

VGG Net (2014)

GoogleNet (2015)

Microsoft ResNet (2015)

Caffe: BAIR Reference CaffeNet, BAIR AlexNet

Torch: VGG_CNN_S, VGG_CNN_M,

VGG_CNN_M_2048, VGG_CNN_M_10

24, VGG_CNN_M_128, VGG_CNN_F, VGG

ILSVRC-2014 16-layer, VGG ILSVRC-2014

19-layer, Network-in-Network (Imagenet & CIFAR-10)

Google: Inception (V3, V4).

The networks specializing in the location and detection of objects in an image, the object detection network, for example:

R-CNN (2013)

SSD (Single Shot MultiBox Detector: Object Detection network), Faster R-CNN (Faster Region-based Convolutional Network method: Object Detection network)

Faster R-CNN (2015)

SSD (2015).

The above list is not limiting

EX1001, Abstract, 1:5-28, 16:7-36

The '945 Patent

(57)

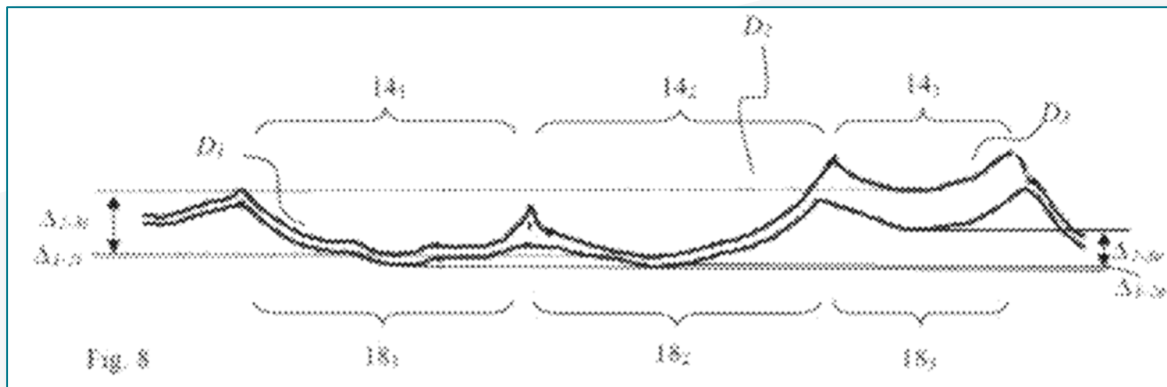
ABSTRACT

A method for evaluating the shape of an orthodontic aligner worn by a patient. Acquisition of at least one two-dimensional image of teeth of the patient as "updated image." At least one updated image as "aligner image," at least partially representing the aligner in a service position. At least one updated image as "dentition image," representing the teeth. If the dentition image is different from the aligner image, conversion of the dentition image so that it represents the teeth as seen under the acquisition conditions used during the acquisition of the aligner image. Determination, for each of a plurality of teeth represented on the dentition and aligner images, of interior and exterior tooth outlines representing the outline of the free end of the tooth on the dentition and aligner images. Comparison of the interior and exterior tooth outlines, followed by determination of at least one score according to the comparison.

A "tooth outline" is understood to mean a line displaying the shape of a free end 22 of a tooth on an image.

Of course, such a line may delimit the free end of this tooth. The tooth outline is then qualified as an "interior tooth outline" 14. A "dentition outline" 16 is made up of the set of interior tooth outlines of a set of teeth (FIG. 4).

Such a line may also delimit part of the aligner. The tooth outline is then qualified as an "exterior tooth outline" 18. In particular, the bottom 20 of the tray has a shape that substantially compliments that of the free ends 22 of the teeth. As a result, in an image depicting the bottom of the tray, the tray-bottom outline defines a set of exterior tooth outlines 18.



EX1001, Abstract, 5:10-20, Fig. 8

'248 Patent Claim 1

- Acquire image by cell phone
- Generically analyze with a computer using known deep learning techniques
- Perform known aligner fit assessment

1. A method for assessing the shape of an orthodontic aligner, said method comprising the following steps:

a") more than 1 week after the start of the treatment with the aligner, acquisition of at least one image at least partially representing the aligner in a service position in which it is worn by a patient, called "analysis image", the analysis image being a photograph, or an image extracted from a film;

b") analysis of the analysis image by means of a deep learning device, trained by means of a learning base, so as to determine a value

for at least one tooth attribute of an "analysis tooth zone" representing, at least partially, a tooth on said analysis image, the tooth attribute relating to a separation between the tooth represented by the analysis tooth zone, and the aligner represented on the analysis image,

in the step a"), a cellphone is used to acquire the analysis image.

EX1001, claim 1; see EX1050

'945 Patent Claim 1

- Acquire image
- Detect edges with a computer using known image processing techniques
- Perform known aligner fit assessment based on comparison of edges

1. A method for evaluating the shape of an orthodontic aligner worn by a patient, said method comprising the following steps:

1) acquisition of at least one two-dimensional image of teeth of said patient, referred to as “updated image”, under actual acquisition conditions;

at least one updated image referred to as “aligner image”, at least partially representing the aligner in a service position in which it is worn by said teeth; and at least one updated image referred to as “dentition image”, identical to or different from the aligner image, representing said teeth,

~~2) if the dentition image is different from the aligner image, conversion of the dentition image so that it represents said teeth as seen under the acquisition conditions used during the acquisition of the aligner image in step 1);~~

3) determination, by means of image processing software, for each of a plurality of teeth represented on the dentition and aligner images, of interior and exterior tooth outlines representing the outline of the free end of said tooth on the dentition and aligner images, respectively;

4) comparison of the interior and exterior tooth outlines, so as to determine at least one score according to said comparison.

EX1001, claim 1

Claim Construction: Petition ('248 Patent)

Term	Petition	PO response
tooth attribute relating to a separation ...	any attribute of a tooth that may relate to the tooth's separation from an aligner	attribute of a tooth, but not positions or edges of teeth or aligners
image attribute relating to a separation ...	any attribute of the image as a whole that relates to the separation between one or more teeth and the aligner	attribute of the image as a whole
image extracted from a film	image extracted from a video recording	No response
analysis of the analysis image by means of a deep learning device ...	function: analyzing an image structure: a device comprising a deep learning neural network trained with a learning base	No response

Claim Construction: Petition ('945 Patent)

Term	Petition	PO response
aligner and dentition images	may be a single image, with second element of claim 1 ignored	No response
interior and exterior tooth outlines	“interior tooth outline” refers to a line delimiting a free end of a tooth; “exterior tooth outline” refers to a line delimiting a part of an aligner	No response
determination, by means of image processing software ... of interior and exterior tooth outlines....	function: determining interior and exterior tooth outlines structure: encompasses image processing software, such as any of the examples identified in the '945 patent or Salah	No response

Claim Construction: Patent Owner ('248 Patent)

Term	POR	Effect
deep learning device	device that, through training, is capable of analyzing an image and recognizing patterns therein	No effect on petition grounds
shape of an orthodontic aligner	if limiting, the shape of the aligner relative to the shape of the teeth of the patient	No effect on petition grounds

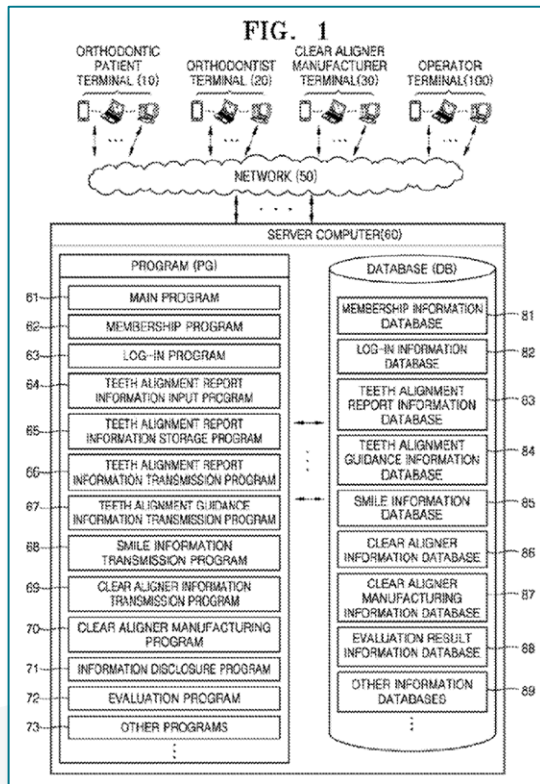
POR 13-16; Reply 5-7

Claim Construction: Patent Owner ('945 Patent)

Term	POR	Effect
for each of a plurality of teeth ... interior and exterior tooth outlines representing the outline of the free end of said tooth	requiring the use of image processing software to determine for each of multiple teeth, an interior tooth outline representing the outline of the free end of each of the multiple teeth and an exterior tooth outline representing the corresponding part of the aligner	No effect on petition grounds
shape of an orthodontic aligner	if limiting, the shape of the aligner relative to the shape of the teeth of the patient	No effect on petition grounds

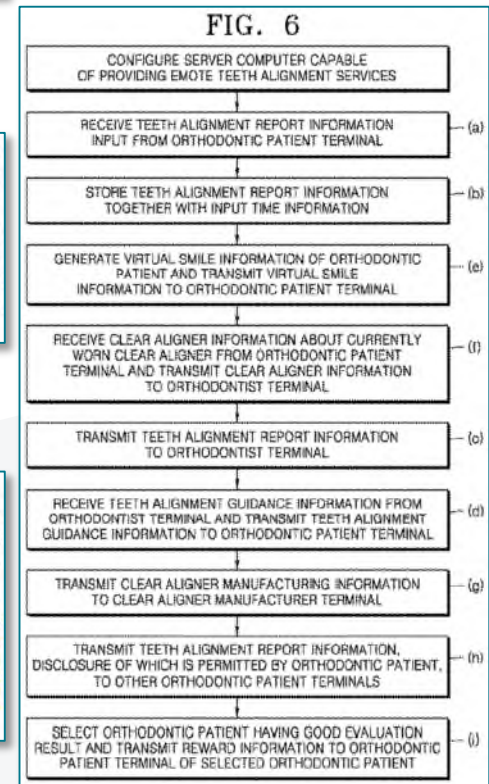
POR 15-19; Reply 5-6

(54) **METHOD AND SYSTEM FOR PROVIDING
REMOTE TEETH ALIGNMENT SERVICES
AND COMPUTER-READABLE MEDIUM ON
WHICH THE METHOD IS RECORDED**



[0042] Herein, the orthodontic patient terminal 10 is not limited to a smartphone and may be one of all information devices capable of transmitting a teeth alignment information. For example, the orthodontic patient terminal 10 may be a personal computer, a mobile device, a personal digital assistant (PDA), a smart pad, a laptop computer, a smart camera, a smart camcorder, an electronic book, a smart scanner, or the like.

[0066] The clear aligner information transmission program 69 may be a program for receiving the clear aligner information about a currently worn clear aligner from the orthodontic patient terminal 10, storing the clear aligner information, and transmitting the clear aligner information to the orthodontist terminal 20 in such a manner that the orthodontist checks a photo or a video of the clear aligner currently worn by the orthodontic patient, and determines whether the clear aligner is normal and corresponds to an appropriate step.



Invisalign

INVISALIGN® CLINICAL MONITORING GUIDE



*Aligner seating
poorly, >1 mm*



Aligner seating well

General fit issues (Aligners are not seating well, >1 mm)

P

If Aligner discrepancy is <1 mm, this is natural deflection that should occur so the Aligner can move the tooth.

Salah

(54) Title: METHOD FOR MONITORING DENTITION

(72) **Inventors:** **Salah, Philippe**; 18, rue Edouard Vaillant, F-93170 Bagnolet (FR). **Ayache, William**; 18, rue Ancelle, F-92200 Neuilly-sur-Seine (FR). **Debraux, Laurent**; 57, rue des Cascades, F-75020 Paris (FR) **Ghy-Selinck, Guillaume**; 41, rue d'Arleux, F-59169 Cantin (FR). **Pellissard, Thomas**; 5, rue de la famille Solomon, F-95190 Goussainville (FR).

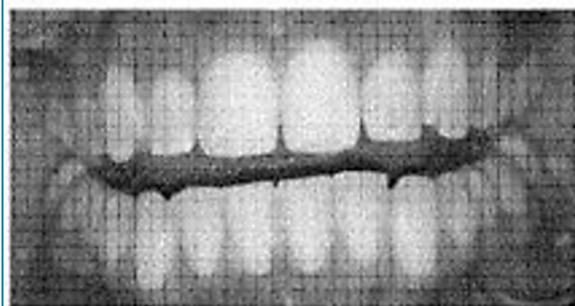


Fig. 6a

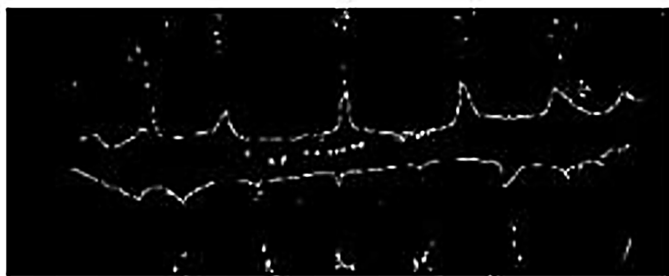


Fig. 6b

Discriminant information is preferably chosen from the group consisting of contour information, color information, density information, a distance information, brightness information, saturation information, reflection information and combinations of these.

A person of skill in the art know how to process an updated image to reveal discriminant information. This processing involves, for example, the application of well-known masks or filters provided by image processing software. Such processes can be used, for example, to detect high-contrast regions and determine contours.

These processes include one or more of the following well-known and preferred methods:

- by applying a Canny filter, in particular to search for contours using Canny's algorithm;
- by applying a Sobel filter, in particular for calculating derivatives using the extended Sobel operator;
- by applying a Laplace filter, to calculate the Laplacian of an image;

...

The following non-limiting methods, although not preferred, can also be used:

- by applying a "MeanShift" filter, in order to find an object on an image projection;

...

- by Active contour modeling (tracing the contour of an object from an image) 2D potentially "noisy";
- by calculating a field of forces, called GVF ("gradient vector flow"), in a part of the image;
- by Cascade Classification;
- by "Deepleraning" [sic].

EX1008, [54], [72], 34:8-36:16, Figs. 6a-b

Maninis

Convolutional Oriented Boundaries: From Image Segmentation to High-Level Tasks

Abstract—We present Convolutional Oriented Boundaries (COB), which produces multiscale oriented contours and region hierarchies starting from generic image classification Convolutional Neural Networks (CNNs). COB is computationally efficient, because it requires a single CNN forward pass for multi-scale contour detection and it uses a novel sparse boundary representation for hierarchical segmentation; it gives a significant leap in performance over the state-of-the-art, and it generalizes very well to unseen categories and datasets. Particularly, we show that learning to estimate not only contour strength but also orientation provides more accurate results. We perform extensive experiments for low-level applications on BSDS, PASCAL Context, PASCAL Segmentation, and NYUD to evaluate boundary detection performance, showing that COB provides state-of-the-art contours and region hierarchies in all datasets. We also evaluate COB on high-level tasks when coupled with multiple pipelines for object proposals, semantic contours, semantic segmentation, and object detection on MS-COCO, SBD, and PASCAL; showing that COB also improves the results for all tasks.

At the core of all these deep learning approaches lies a *base CNN*, starting from the seminal AlexNet [8] (8 layers), through the more complex VGGNet [10] (16 layers) and inception architecture of GoogLeNet [9] (22 layers), to the very recent and very deep ResNets [11] (up to 1001 layers). Image classification results, which originally motivated these architectures, have been continuously improved by exploring deeper and more complex networks. In this work, we present results both using VGGNet and ResNet, showing that COB is modular and can incorporate and benefit from future improvements in the base CNN.

of the art. We have also studied the effects of COB when coupled with well-known pipelines, showing that injecting COB detections to them lead to improvements on Semantic Segmentation and Object Detection. Finally, we report a new state of the art on Semantic Boundary detection.

EX1007, Abstract, 2, 13; *see also id.*, 11-12, Tables 4-7

Maninis

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).

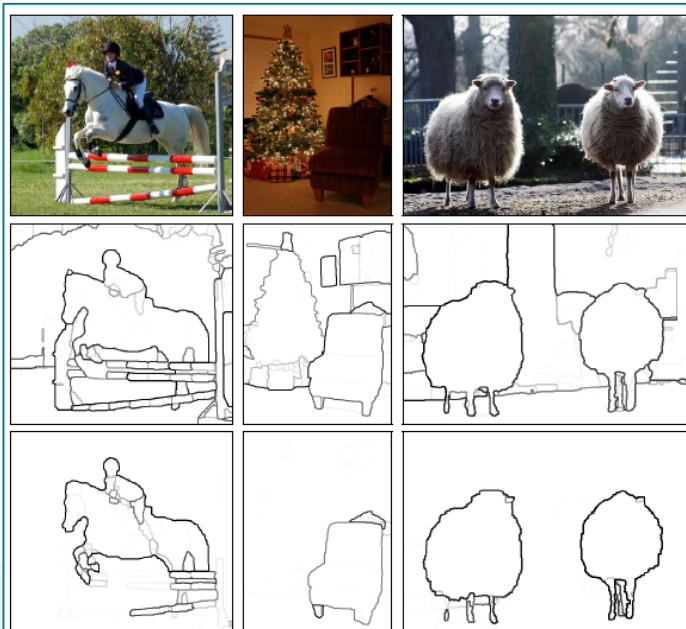


Fig. 10. **Qualitative results for Object Boundaries.** Row 1: original images, Row 2: Generic Image Segmentation results, Row 3: Object Boundary results.

Our technique achieves state-of-the-art performance by a significant margin for contour detection, the estimation of their orientation, and generic (RGB and RGB-D) image segmentation. We show that our architecture is modular by using two different CNN base architectures, which suggests that it will be able to transfer further improvements in CNN base architectures to perceptual grouping. We also show that our method does not require globalization, which was a speed bottleneck in previous approaches. The generalization of COB was further demonstrated when applied to high-level vision tasks (object proposals, object detection, and semantic contours and segmentation) in combination with recent pipelines, where the results are improved in all cases.



Fig. 16. **Qualitative results for Semantic Segmentation.** Row 1: original images, Row 2: Dilated Convolution Network, Row 3: Dilated Network with COB superpixels.

Proposed Combination

- Providing assessment based on images of aligners on teeth, useful for aligner fit evaluation
- Based on combined teachings of Kim and Invisalign of a tooth-aligner separation value in a cellphone image, relative to a threshold
- Identified between edges of teeth and aligner detected by Salah's image processing techniques (945 patent), including deep learning as taught by Salah and Maninis (248 patent)

248:

78. 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 Kim and Salah's image acquisition and analysis procedures. For example, a POSA would have been motivated to adapt Salah and Kim's image acquisition and analysis devices to employ a deep learning architecture as taught by Maninis, trained on images relevant to the dental imaging applications of Salah, Kim, and Invisalign. In that combination, Maninis's deep learning architecture would provide specific functionalities required by Salah, Kim, and Invisalign, including the ability to identify objects (such as the teeth and aligners depicted in each image) and to detect contours (such as the edges of teeth and aligners useful to assess aligner fit as taught by Invisalign).

79. A POSA would have had reason to use the techniques of Salah and Maninis for contour detection to detect the edges of teeth and aligners, as Invisalign teaches that the gap between the surfaces of teeth and an aligner should be below a threshold value (1 mm) when fitting well, aligning with Kim's teaching to assess whether an aligner was "normal and corresponds to an appropriate step." EX1005, ¶66; EX1006, 1. As I explained above, the width of a gap between two depicted objects in an image (such as a tooth end and a surface of an aligner) corresponds to the distance between the contours for each object. Equivalently, the "gap" is a portion of the image between those surfaces, and its width is the distance from one side to the other (i.e., the distance from a contour separating "tooth" from "gap" and a contour separating "gap" from "aligner."). Thus, to apply Invisalign's

945:

applying image processing software to perform contour detection as taught by Salah to detect edges of teeth and aligners, producing a measurement of distance between these surfaces, as taught by Invisalign, to assess whether the imaged aligner fits appropriately for the current step in patient treatment, as taught by Kim.

248-Pet., 20-27; EX1002, ¶¶78-79; 945-Pet., 17-23; EX1002, ¶64

Rationale to Combine

- Kim and Invisalign give reason to analyze of images of aligners on teeth to aid in assessing fit using well-known methods of assessment
 - Kim teaches remotely evaluating cell images to assess aligners currently worn by patient to **reduce time and cost of treatment and facilitate remote assessment**
 - Invisalign teaches **measurements useful to assess aligner**, including threshold (>1mm separation between tooth and aligner edges) indicating whether fit is poor
- POSAs would look to Salah's teaching of image processing techniques to **detect edges of objects in dental photos**, providing respective edges of teeth and aligners to measure separation
- Regarding 248 patent, further combination with Maninis provides **state-of-the-art edge and object detection** to provide accurate edges for measurement
 - Salah teaches use of deep learning for detecting contours in dental images
 - Maninis's deep learning techniques broadly applicable and expected to improve performance
 - Maninis teaches its techniques represent current trend in computer vision

248-Pet., 20-27; EX1002, ¶¶69-81; 945-Pet., 17-24; EX1002, ¶¶56-65

Level of Skill

- Dr. Foroosh meets both parties' definition of level of skill:

Align:

As Dr. Foroosh explains, a person of ordinary skill in the art (POSA) by July 21, 2017, would typically have an advanced degree, such as a Masters' Degree or Ph.D. in computer science, computer engineering, electrical engineering, 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. See EX1002, ¶¶60-61.

PO:

54. I understand that Dental Monitoring has proposed that a person of ordinary skill in the art ("POSA") as of the priority date of the '248 patent would have had a bachelor's degree or higher in computer science, bioinformatics, or a related engineering discipline, and several years of work experience relating to the development of machine learning or artificial intelligence models or algorithms. A POSA would have also worked as part of a multi-disciplinary team, which includes individuals familiar with medical imaging, dental imaging, dentistry, and/or orthodontics.

- Dr. Mongan's opinion on Kim required no dentist input:

Q. So my question again was that in this paragraph you discuss Kim's discussion in Kim's paragraph 66 regarding taking a photo or video of a currently worn clear aligner, correct?

A. This paragraph references Kim 66 and includes a quote about a currently worn, clear aligner and discussion of what that would mean to a POSA.

Q. Did you reach your opinions in these two paragraphs on your own, independent of your discussions with Dr. Kusnoto?

A. I believe I developed an understanding of what this phrase meant prior to talking with Dr. Kusnoto.

248-Pet., 14; EX1002, ¶¶60-61; POR, 3-4; EX2030, ¶54; Reply, 4 n.2; EX1051, 29:15-31:15

Dr. Kusnoto Abandons Criticisms of Dr. Foroosh

- Dr. Foroosh understood all terms as a dental professional would have

- Kim:

Q. So there's no terminology in Kim that you would point me to that you think Dr. Foroosh misunderstood or applied a different understanding than a dental professional would, right?

am glancing through it, I -- I do think Dr. Foroosh is not misunderstanding any of this.

- Invisalign:

little bit, so -- I think with the exception of the illustration of poor fit, okay, for the most part, he's getting the terminology correct.

- Salah:

Q. Are there any terms from the Salah reference Exhibit 1008 that you believe Dr. Foroosh interpreted differently than a dental professional would have?

A. Here, but that's the -- I'm sorry, that's the guy again.

Yeah, I think for the most part he also gets it correct, I mean, in term of the terminology.

- Frey:

sentence kind of suggests that it's -- it's a consistent with methods for orthodontic practice.

It's almost like generalization but, you know, if it's refer only for Frey, there's one -- one instance here then, I mean, this depiction is good.

Reply, 3-4; EX1052, 43:23-45:24, 46:3-47:25,

Kim in view of Invisalign Teaches Aligner Fit Assessment

- Kim teaches assessing cell photos of currently worn aligner to evaluate whether aligner is normal and appropriate for current treatment step:

[0066] The clear aligner information transmission program 69 may be a program for receiving the clear aligner information about a currently worn clear aligner from the orthodontic patient terminal 10, storing the clear aligner information, and transmitting the clear aligner information to the orthodontist terminal 20 in such a manner that the orthodontist checks a photo or a video of the clear aligner currently worn by the orthodontic patient, and determines whether the clear aligner is normal and corresponds to an appropriate step.

- Invisalign teaches fit assessment based on aligner separation and shows separation is visible in digital photos of aligners worn on patients' teeth:



248-Pet., 20-24; Reply, 13-15; 945-Pet., 17-20; Reply, 7-9; EX1005, ¶66; EX1006, 1

Common Practice to Assess Aligner Fit

- Dr. Kusnoto admits checking aligner fit routinely performed along with assessment of teeth

aligner worn by the patient is still suited to the treatment.”). In other words, if a patient was on week 10 of their treatment, an orthodontist would make a visual comparison between the dentition observed on the patient and a visual rendering of expected dentition at week 10 for that patient. *See id.* If the actual dentition deviated from the expected dentition, the orthodontist would examine the areas of deviation and assess what issues were causing dentition to progress differently than expected.

Q. Poor aligner fit could be one thing preventing the treatment from progressing as expected, right?

A. They are somewhat independent because the -- the way that we evaluate them first is we need to make sure that the patient's teeth look similar to what the simulation at the given stage is. Okay? You don't see any growths completely totally out. Okay?

So if that already pass, then we assess the aligner fit secondarily to that, but if that already completely failed then, you know, we don't even bother

48. Second, orthodontists assessed the aligner treatment. Typical issues that might be identified during the assessment included problems with aligner fit, tooth attachments, or dentition complications. *See, e.g.,* EX2026 [Glaser], 39-40, 275-76 (discussing various treatment features to be monitored at each visit); EX1006 (discussing monitoring by the orthodontist of aligner fit, dentition, and attachments).

Q. Here you refer to two primary assessments at a patient's examination, right?

A. Mm-hmm.

Q. And those are assessment of alignment of teeth and assessment of aligner fit, right?

A. Alignment of teeth and aligner fit.

Q. Was it a routine part of every aligner treatment for an orthodontist to assess aligner fit at regular intervals?

A. Assessing the progress of treatment, which include aligner fit, at regular interval, yes.

Frey Corroborates Motivation

- Published prior to earliest claimed priority date
- Describes obtaining photos of aligners worn in service position to perform orthodontic assessment:

Record with a team member some example videos that demonstrate exactly **treatment progress information should be captured on a smartphone**. You should record 3 example videos: **one with teeth apart and aligners in**, one biting down with aligners removed, and one teeth apart and aligners out. Videos work better than individual photos because data capture is continuous and the video can then be paused during viewing. To capture the videos, the patient can

- No showing that any disclosure in Frey had been directly or indirectly obtained from or previously disclosed by any joint inventor
 - Dr. Frey discussing Dental Monitoring service does not equate to obtaining disclosure from 248/945 inventors
 - No evidence showing invention of taking photos of aligners on teeth
- PO's own exhibit indicates obtaining photos like those taught in Frey to track treatment progress of remote patients was already a familiar practice:

The Dental Monitoring service is built on the premise that everyone now carries a smartphone; however, that also means that **photo monitoring** and FaceTime appointments can be implemented without purchasing additional software. **We routinely do this for international and out-of-state patients already** so I don't see the photo monitoring plans offering much value to the practitioner.

248-Pet., 23-24; Reply, 17-19; EX1009; EX2053; 945-Pet., 20-21; Reply, 11-12

POSA's Motivated to Combine with Salah

- POSA's motivated to use Salah's image processing techniques (e.g., edge detection, deep learning) to detect and assess gaps between teeth and aligners, as suggested by Kim and Invisalign

EX1005, ¶¶8-10. Additionally, a POSA would have understood that the type of image analysis suggested by Kim and Invisalign could be readily applied on images or videos showing a clear orthodontic aligner being worn on the teeth of a patient. For example, a POSA would have recognized that by using computer-vision techniques to detect the locations of the ends of the patient's teeth and the corresponding surfaces of the aligner, the distance between the two could be measured, corresponding to the assessment taught by Invisalign.

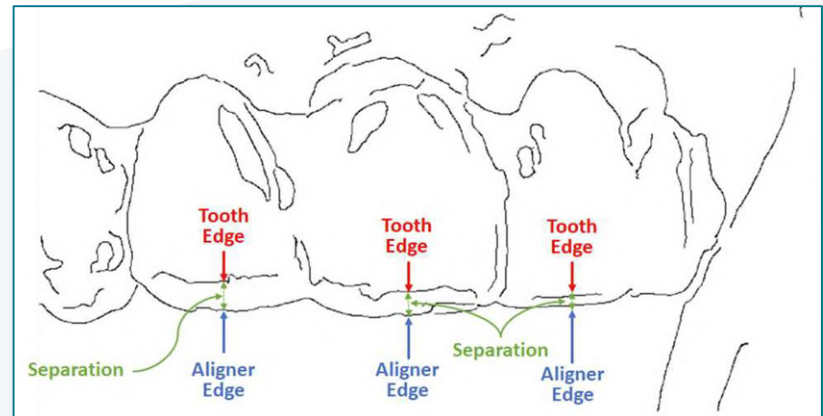
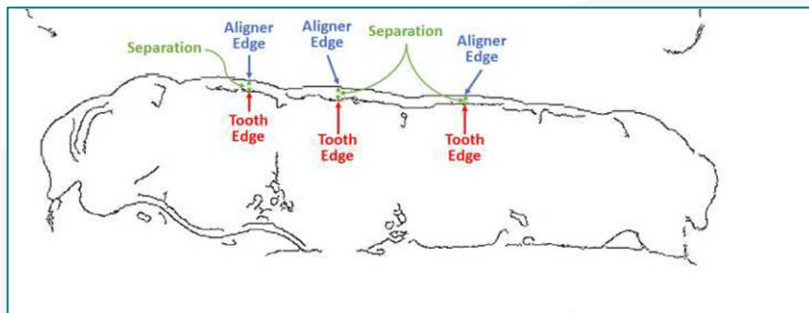
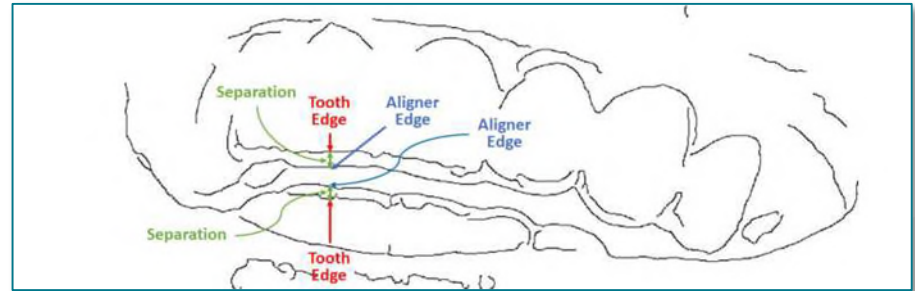
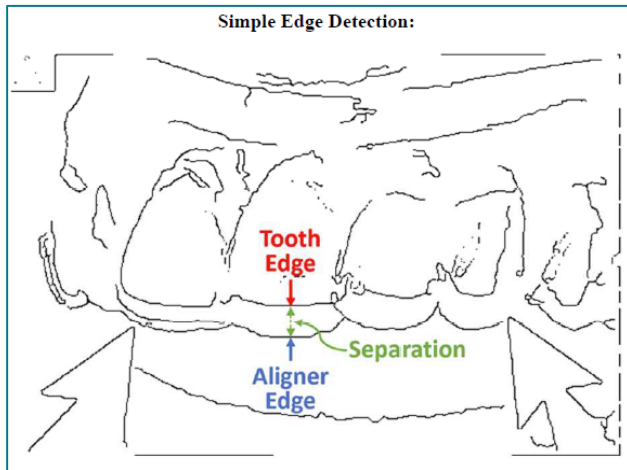
79. A POSA would have had reason to use the techniques of Salah and Maninis for contour detection to detect the edges of teeth and aligners, as Invisalign teaches that the gap between the surfaces of teeth and an aligner should be below a threshold value (1 mm) when fitting well, aligning with Kim's teaching

74. A POSA would have known how to determine the width of such gaps using image processing software. As I discussed briefly above, the distance between depicted teeth and a depicted aligner corresponds to the distance from the edge of one to the edge of the other. Salah provides disclosure useful for determining such edges. Like Kim, Salah teaches application of computer image

24. Salah teaches that analysis of acquired images may include determining contours of objects in an image, including by using deep learning for contour detection. EX1008, 34:5-36:14.

Detecting Edges Straightforward for Each Tooth and Aligner Part

- Edge detection in aligner images straightforward, even using the simplest edge-detectors
- PO seeks to avoid prior art disclosure, but edge detection is straightforward even for PO's allegedly “more representative” images



Maninis is Prior Art

- No dispute on date of submission for publication (months before earliest claimed priority date of July 21, 2017):

arXiv:1701.04658v2 [cs.CV] 28 Apr 2017

- No dispute materially identical version (EX1021) published prior to April 28, 2017. *See Valve Corp. v. Ironburg Inventions Ltd.*, 8 F.4th 1364, 1370-71 (Fed. Cir. 2021).
- Archived policy demonstrates publication no later than May 2, 2017:

Availability of submissions and announcement order

Note: policy change 2017-01 regarding announcement order and the effect of edits.

Submissions to arXiv are normally made public on Sunday through Thursday, with no announcements Friday or Saturday. Check the [current time](#) at the arXiv main site, with the amount of time remaining before the "freeze".

Submissions Received Between (all times Eastern US)	Will Be Announced (all times Eastern US)	Mailed to Subscribers
Monday 14:00 - Tuesday 14:00	Tuesday 20:00	Tuesday Night / Wednesday Morning
Tuesday 14:00 - Wednesday 14:00	Wednesday 20:00	Wednesday Night / Thursday Morning
Wednesday 14:00 - Thursday 14:00	Thursday 20:00	Thursday Night / Friday Morning
Thursday 14:00 - Friday 14:00	Sunday 20:00	Sunday Night / Monday Morning
Friday 14:00 - Monday 14:00	Monday 20:00	Monday Night / Tuesday Morning

- Dr. Foroosh knew policy at the time and saw its regular operation for papers including his own
- Advanced Micro-Devices, Inc.*, 848 F.2d 1560, 1569 (Fed. Cir. 1988): “Evidence of routine business practice can be sufficient to prove that a reference was made accessible before a critical date.”

Pet., 10; EX1007, 1; EX1002, ¶¶53-58; EX1021; EX1022, 3; Reply, 7-8; EX1023; EX1024; EX1025; Paper 35

Combination with Maninis Works as Proposed

- PO argues: “Maninis COB neural network is incapable of determining what [depicted] objects might be”
- Dr. Mongan admits he uses the term “COB neural network” to exclude Maninis’s teachings of using COB in a pipeline for object classification (e.g., in Maninis section 6, pp. 11-12)
- The ’248 petition proposed using Maninis’s deep learning architecture to detect contours and also to classify objects

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

Kim, and Invisalign. In that combination, Maninis’s deep learning architecture would provide specific functionalities required by Salah, Kim, and Invisalign, including the ability to identify objects (such as the teeth and aligners depicted in each image) and to detect contours (such as the edges of teeth and aligners useful to assess aligner fit as taught by Invisalign).

101. 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

between tooth and aligner. A POSA would then have set the output layers of Maninis’s neural network to provide the required object identifications corresponding to the labels used to identify individual teeth, aligner portions, and spaces between them, based on Maninis’s teaching of semantic boundaries, so that the boundaries of each part of the image could be identified by the neural network after training. Applying the deep-learning based image processing approach taught

Pet., 24-27, 34-40; EX1002, ¶¶75, 101-02; Reply, 9-10; EX1048, ¶¶23-28; EX1037, 104:11-19, 107:2-17, 03:14-104:10, 118:13-119:11; POR, 42-46

Maninis Provides Technical Improvement, Unlike '248 Patent

Maninis improves over state-of-the-art:

(Column (3)). Our final technique COB, in which we remove the globalization step, computes the three scales in one pass and add contour orientations, takes 0.79 seconds in mean. 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).

Our technique achieves state-of-the-art performance by a significant margin for contour detection, the estimation of their orientation, and generic (RGB and RGB-D) image segmentation. We show that our architecture is modular by using two different CNN base architectures, which suggests that it will be able to transfer further improvements in CNN base architectures to perceptual grouping. We also show that our method does not require globalization, which was a speed bottleneck in previous approaches. The generalization of COB was further demonstrated when applied to high-level vision tasks (object proposals, object detection, and semantic contours and segmentation) in combination with recent pipelines, where the results are improved in all cases.

'248 patent relies on known, generic deep learning:

A “neural network” or “artificial neural network” is a set of algorithms well known to a person skilled in the art.

The neural network may in particular be chosen from:

the networks specializing in the classification of images, called “CNN” (“convolutional neural network”), for example

AlexNet (2012)

ZF Net (2013)

VGG Net (2014)

GoogleNet (2015)

Microsoft ResNet (2015)

Caffe: BAIR Reference CaffeNet, BAIR AlexNet

Torch: VGG_CNN_S, VGG_CNN_M,

VGG_CNN_M_2048, VGG_CNN_M_10

24, VGG_CNN_M_128, VGG_CNN_F, VGG

ILSVRC-2014 16-layer, VGG ILSVRC-2014

19-layer, Network-in-Network (Imagenet & CIFAR-10)

Google: Inception (V3, V4).

The networks specializing in the location and detection of objects in an image, the object detection network, for example:

R-CNN (2013)

SSD (Single Shot MultiBox Detector: Object Detection network), Faster R-CNN (Faster Region-based Convolutional Network method: Object Detection network)

Faster R-CNN (2015)

SSD (2015).

The above list is not limiting

No Deterrence from Training Costs

- Dr. Kusnoto admits estimates based on unneeded steps:

Q. And were you labeling just the edges of the teeth and the tooth numbers or was there additional labeling?

A. There are some additional parameters like inter-molar width, inter-canine width, the depth of the arch, the contour of the tongue, a few other parameters.

Q. Did that project involve labeling any issue with the teeth, like malocclusion?

A. Could be, yeah. Could be. The one with the American Board of Orthodontics, I think it is, yeah.

Q. Right.

And if it's an image of just of the teeth, not the patient's face, then you don't need to redact the image; right?

MR. BLYTHE: Objection to form.

THE WITNESS: The image itself, no, but the identifier that goes with the image, yes.

Q. And when you talk about the time that it would take to label and your estimate of that and the 20 minutes that you think it would take to label, that's based on your own experience that you just described earlier in that project from 2016; right?

A. Yeah, when the -- as I said, there are multiple times we are doing similar things. So the number always came up would be about the same per image.

- Dr. Foroosh:
Costs not an issue regardless:

random identifiers to the image.” EX2030, ¶88. Dr. Kusnoto does not explain what facial features would need to be redacted in the types of images at issue here. While an image showing the upper part of a person's face, including their eyes, might be identifiable, that does not mean that images of a mouth would be identifiable. For example, the images shown in Salah Figs. 5a-6a show only a patient's dental arches and the lower parts of the face. EX1008, Figs. 5a-6a; see also EX1005, Figs. 8-13 (similar images in Kim). Regarding removing patient identifying information and metadata from image files, as well as assigning random identifiers, these processes can easily be performed automatically at practically no cost, using programs or scripts that would be elementary for a POSA to write.

31. Regardless, a POSA would not have been deterred even by substantial development costs. The production of a learning base is essentially a one-time, upfront expense as part of the development process of a deep learning device. A

Reply, 22-24; EX1038, 81:20-25, 88:9-17, 91:9-18; EX1048, ¶¶25-31

'945 Patent Combination Recognizes Edges of Teeth and Aligners

- PO argues some edge detection (e.g., Canny) does not, on its own, distinguish teeth and aligner edges, but does not respond to petition's case
- Salah's image processing techniques enable identifying respective edges for plurality of teeth/aligner parts:
- POSAs motivated to assess gaps for a plurality of teeth based on teachings of Invisalign:

teeth and an aligner in real life. The ability to see a gap and visually assess it in real life naturally maps over to the ability to detect and measure the same gap using computer vision techniques like Salah's various image processing software options for detecting objects and the contours defining the objects' edges, and applying such known methods of contour detection would produce contours in images like those discussed in Kim and Salah that corresponded to the free ends of teeth and the aligner depicted in such images.

is smaller in that image. Applying more advanced techniques, such as the remaining well-known image processing techniques taught by Salah, would have been expected to resolve the locations of each tooth's edge and the corresponding aligner edges with even greater precision and accuracy, allowing the detection, identification, and measurement of separations between each tooth and the corresponding edges of the aligner in a straightforward and predictable manner.



EX1006, 1. As can be seen above, there are two arrow pointing at the surfaces of an aligner corresponding to distinct teeth also depicted on the image. Invisalign thus indicates applying its distance assessment to a plurality of teeth. Invisalign also urges that practitioners should "be tooth/teeth specific" when indicating necessary adjustments, and it includes a section handling "tooth-specific movements." EX1006, iv, v (table of contents), 2 (cross-referencing to sections handling tooth-specific issues such as incomplete extrusions and unintended intrusions). In light of this disclosure, a POSA would have understood that the distance measurements disclosed by Invisalign applied to a plurality of teeth and the corresponding aligner surfaces for each tooth, not just a single tooth. Thus, when applying Invisalign's teachings, a POSA would have measured the distances

Pet., 31-32, 39-40; Reply, 10-11, 15-17 & n.2; EX1002, ¶¶83, 103; EX1048, ¶14

'248 Claims 7 and 14

...creation of a learning base comprising more than 1000 images of dental arches, or “historical images” with “historical tooth zones,” “to each of which, for at least one tooth attribute relating to a separation between the tooth represented by the historical tooth zone considered, and the aligner represented, a tooth attribute value is assigned;”

- No dispute that measurement of distance between edge of tooth and edge of aligner is a tooth attribute of a tooth zone that relates to the recited separation
- Labeled locations of respective edges (such as the location of a tooth’s edge) are tooth attributes of historical tooth zone
- Labeled gap separating tooth from aligner also a tooth attribute of tooth zone
- Dr. Mongan admits his distinction, requiring tooth attribute value to be a singular number, was mistaken:

I'm just asking, the term "image attribute value" uses value in the singular, just like tooth attribute value, correct?

A. I'd just like to take a minute to review the claim construction that I have here on tooth attribute and image attribute.

Yes, the word "value" is used in the singular in both of those. So my statement that that would have informed a POSA's understanding that the tooth attribute doesn't represent a set of values may not have been correct.

Pet., 44-50, 54-55; Reply, 24-25; EX1002, ¶¶120-37, 149-52

'248 Claim 8

...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... relating to a separation between the tooth or teeth and the aligner represented on the historical image”

- Presence or absence of a gap in an image is an image attribute relating to separation
- Labeled locations of respective edges (e.g. the location of edges of teeth and aligner) are tooth attributes of historical tooth zone, and image attribute value can be a set of all tooth attribute values, as can presence or absence of something in the image, such as an aligner or a gap

also relate to the presence of a separation. See EX1001, 26:11-14, 29:30-53. By annotating the images with an aligner (where present) or gap (where present), the image would be annotated with (and thus comprise) image attributes relating to the recited separation. EX1002, ¶140. Further, the set of locations of each tooth and

In the step c'), an attribute value of the analysis image is determined as a function of the tooth attribute values. The attribute value of the analysis image may relate to its general orientation and may for example take one of the following three values: “right photo”, “left photo” and “front photo”. The attribute value of the analysis image may also be the list of the tooth numbers represented, for example, “16, 17 and 18”. The attribute value of the analysis image may also be, for example, the “presence” or “absence” of a dental, preferably orthodontic, appliance, or the state of opening of the mouth (“mouth open”, “mouth closed”).

Pet., 50-54; Reply, 23-24; EX1002, ¶¶138-45; EX1001, 26:16-26

Motion to Exclude Should be Denied

- EX1022, EX1023, EX1025, EX1026:
 - Moot in view of EX1031
 - Not hearsay
 - Subject to hearsay exceptions under FRE 803(3), 803(6), and 807
- EX1002, paragraphs 62-67:
 - Based on personal knowledge
 - Corroborated by admissible exhibits
 - Supported by computer science expertise (EX1002, ¶66; EX1029, ¶3)