UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
13/908,855	06/03/2013	Jean-Francois Macq	LUTZ 201663US01	4282

48116 7590 09/24/2019
FAY SHARPE/NOKIA
1228 Euclid Avenue, 5th Floor
The Halle Building
Cleveland, OH 44115-1843

EXAMINER	
RAHMAN, MOHAMMAD J	

ART UNIT	PAPER NUMBER
2487	

NOTIFICATION DATE	DELIVERY MODE
09/24/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

Nokia.IPR@nokia.com
docketing@faysharpe.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Ex parte JEAN-FRANCOIS MACQ, CHRISTOPH STEVENS, PATRICE RONDAO ALFACE, and SIGURD VAN BROECK,

---

Appeal 2017-011397
Application 13/908,855[1]
Technology Center 2400

Before THU A. DANG, BARBARA A. BENOIT, and NORMAN H. BEAMER, *Administrative Patent Judges*.

DANG, *Administrative Patent Judge*.

DECISION ON APPEAL

I. STATEMENT OF THE CASE

Appellants appeal under 35 U.S.C. § 134(a) from the Final Rejection of claims 4–7, and 11–13, which are all of the pending claims. Claims 1–4, 8–10, and 14 have been withdrawn. We have jurisdiction under 35 U.S.C. § 6(b).

We AFFIRM.

---

[1] According to Appellants, the real party in interest is Alcatel Lucent. App. Br. 1.

A. INVENTION

According to Appellants, the invention relates to "the field of 3-dimensional (3D) video image generation," and in particular, to "efficiently storing and transmitting 3-dimensional scenery information." Spec. 1:6–9.

B. ILLUSTRATIVE CLAIM

Claim 4 is exemplary:

> 4. A method, performed by one or more processors, for decoding a multiview video stream representing a plurality of viewpoints of a virtual 3D scenery comprising objects, the method comprising for a first one of said viewpoints:
> - extracting a first combined 2D video stream from said multiview video stream;
> - deserializing said first combined 2D video stream into a first plurality of virtual 2D video streams;
> - extracting color information, depth information, and transparency information pertaining to groups of objects in respective object layers from respective ones of said first plurality of virtual 2D video streams;
> - using said respective color information, depth information, and transparency information pertaining to said groups of objects in said respective object layers to generate a first superimposed stream, said first superimposed stream being consistent with said first one of said viewpoints.

App. Br. 11−12

C. REJECTION

Claims 4–7, and 11–13 stand rejected under 35 U.S.C. § 103(a) over Aliprandi et al. (US 2010/0026712 A1; published Feb. 4, 2010), and Robotham et al. (US 6,160,907; issued Dec. 12, 2000), as supported by well-known teaching of Yun et al. (WO 2006/041261 A1; published Apr. 20, 2010).

II. ISSUES

The principal issues before us are whether the Examiner has erred in finding the *combination* of Aliprandi and Robotham teaches *or suggests* a method for "decoding a multiview video stream," comprising the steps of:

"extracting color information, depth information, and *transparency information pertaining to groups of objects in respective object layers* from respective ones of said first plurality of virtual 2D video streams," and

"using said respective color information, depth information, and transparency information . . . to generate a first superimposed stream." Claim 4 (emphasis added).

III. FINDINGS OF FACT

The following Findings of Fact (FF) are shown by a preponderance of the evidence.

Applicants' Invention

1. The Invention provides encoding of information between a complete 3D model and opaque 2D rendering, wherein the encoded information allows the receiver/displayer to recreate a partial 3D model of the 3D scenery, which in turn allows the production of interpolated viewpoints. Spec. 2:16–23. Figure 3, showing an end-to-end system comprising an encoder and a decoder, is reproduced below:

Figure 3

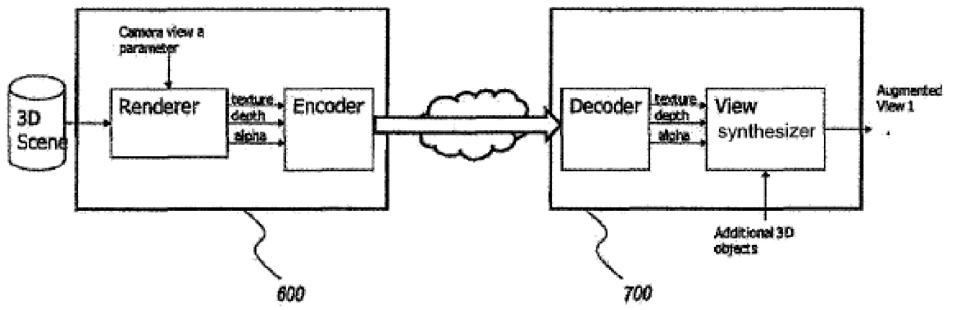

Figure 3 shows an end-to-end system where a single view is rendered and transmitted. *Id.* at 13:5–7. After decoding, the 3D information is inserted and processed (*id.* at 13:6–8), wherein the decoded camera view of the scene and the view containing the objects can then be blended by looking at the depth and alpha values pixel per pixel. *Id.* at 13:19–22.

According to the Invention, "objects" are "generally surfaces that make up the visual representations of the entities appearing in the scenery," while "object layer" is "a subset of the objects making up a particular visualization of the scenery." *Id.* at 2:11–14. Figure 1, showing the application of the concepts of the Invention to a 3D scenery (*id.* at 7:29–30), is reproduced below:

Figure 1

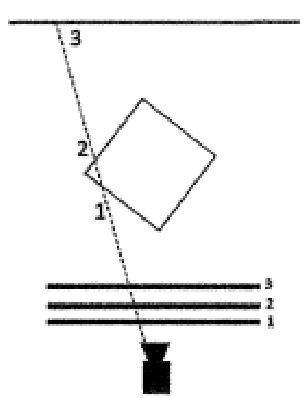

Figure 1 shows an embodiment where, for each pixel location in a 2D camera plane, a ray is considered originating from the camera sensor and passing by that pixel location. *Id.* at 11:13–16. For each object surface intersected in the 3D scene, the color, alpha and depth values are computed and reported. *Id.* at 11:16–19.

Aliprandi

2. Aliprandi discloses rendering of video frames starting from source frames of a scene acquired using a multi-viewing camera system. Aliprandi, Abstract. A three-dimensional impression is obtained by conveying two adjacent views of the same scene related to two different positions. *Id.* at ¶ 13. Pixels of the source frames from the source viewpoints are projected to the target viewpoint as respective contributions to a resulting video frame. *Id.* at ¶ 14. All the contributions of the source views are blended together to obtain a single frame. *Id.* at ¶ 88. In an embodiment, layer blending is used where projected views are used to fill occlusions, layer-by-layer, wherein the cameras are grouped in “layers” before projection with reference to their distance of the target camera, and the contributions are used to fill occlusions layer-by-layer. *Id.* at ¶ 189. Depth information is extracted from the source frames and pixels from the source frames are projected to the target virtual viewpoint by using the depth information. *Id.* at ¶ 14.

Figure 1a, showing a multi-camera capture system, is reproduced below:

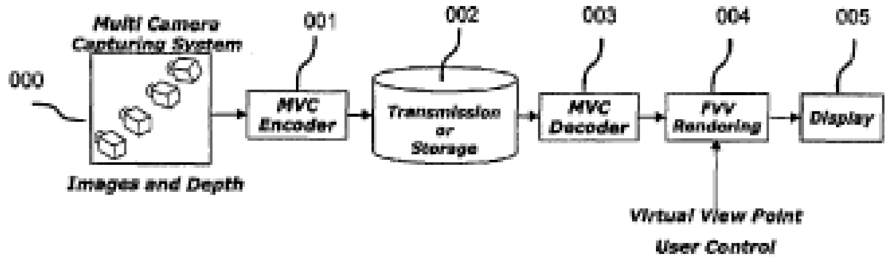

FIG. 1.a

Figure 1a shows a Multi Camera Capture System 000 that provides images and their related depth information to a Multi View Encoder 001. *Id.* at 54. The bit stream produced by the encoder will then be sent through a transmission channel or stored on a storage device 002. *Id.* On the other side, a compliant decoder 003 will receive and decode the stream, the obtained frame and depth information will then be processed by the "FW rendering" block 004, and the obtained frames will be passed to the monitor 005 and shown to the user. *Id.*

Robotham

3. Robotham discloses an integrated three-dimensional process for refining a finish quality media content, wherein elements of a choreography model are iteratively rendered, blended, adjusted, and selected until a desired quality result is achieved. Robotham, Abstract. In Robotham, layered compositing combines multiple visual elements into a single composite montage of images, wherein the individual images are "stacked up" in a series of layers and then "bonded" into a single image sequence. *Id.* at 4:64–5:1.

A human operator is able to refine an image-based or abstraction-based model by adding, changing and deleting an object or group of objects in the image-based (or abstraction-based) model. *Id.* at 22:12–31. A twodimensional finished image representation from the choreography model is generated, and then provided by performing a rendering, blending, and mixing of finished quality, which includes blending multiple image-based objects using associated mattes or depth information. *Id.* at 22:54–61. Each output image frame represents a 2D projection of the 3D virtual stage onto the image plane of the specified virtual camera, wherein this process can include the rendering of abstract objects into image-based objects and blending these with other image-based objects. *Id.* at 22:62–66.

To present a finish-quality rendering and/or image transformation of objects in the choreography model, projections of the objects in the choreography model are calculated from the virtual stage in which they are defined to a 2D viewing space as specified by a camera object or other viewing object. *Id.* at 23:19–29. For each object, additional information is maintained in the choreography model including depth information and transparency "alpha" information, wherein the alpha information is stored with the object and/or represented as a separate matte object to permit proper rendition of inter-object effects such as shadows, reflections and/or refractions, as well as image distorting effects such as atmospheric, spatial, object geometry, morphic, and dynamic texturing. *Id.* at 23:30–42.

Yun

4. Yun discloses a method for encoding/decoding a multi-view video using Layer-Depth Imaging (LDI), wherein the method comprises: generating an LDI (including multiple layers) using color and depth information of each view point image; performing linear decorrelation in each layer of the LDI; performing data aggregation in each linearly-

decorrelated layer; and encoding the aggregated data in each layer to generate an encoded LDI bit stream. Yun, Abstract.

An LDI represents a 3D object with an array of pixels seen from a single camera position, wherein each LDI pixel is represented by its color, depth and other property information supporting LDI rendering such as alpha information. *Id*. at ¶ 3. The LDI is divided into multiple layers, each of which contains a mask indicating the existence of pixel in the layer. *Id*. at ¶ 4. Figure 1, showing an LDI structure, is reproduced below:

[Fig. 1]

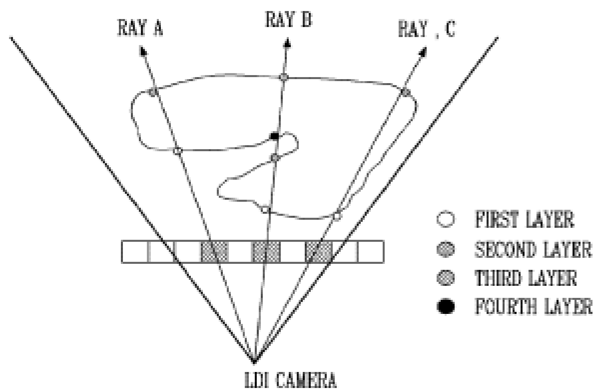

Figure 1 shows a typical LDI structure which includes an array of pixels seen from a single LDI camera position, together with multiple layers based on any viewpoint. *Id*. at ¶ 22. As shown in Figure 1, when rays are shot from an LDI camera position, the rays intersect with an object at a plurality of points, which are ordered from the front to the back, with the first intersections points constituting the first LDI layer, the second points constituting the second layer, and so on. *Id.* Each LDI layer is separated into individual components: luminance, color, transparency, and depth. *Id.*

IV. ANALYSIS

We have reviewed the Examiner's rejection in light of Appellants' arguments presented in this appeal. Arguments which Appellants could have made, but did not make in the Brief are deemed to be waived. *See* 37 C.F.R. § 41.37(c)(1)(iv) (2016). On the record before us, we are unpersuaded the Examiner has erred.

Appellants contend that "Aliprandi is silent with respect to extracting information pertaining to groups of objects <u>in respective object layers</u> from respective a plurality of virtual 2D video streams." App. Br. 7. Although Appellants acknowledge that Aliprandi "mentions blending 'all the contributions of the source view to obtain a single frame,'" Appellants contend "this general disclosure in Aliprandi does not suggest specifically that it would be respective color information, depth information, and transparency information pertaining to objects in respective object layers that is used to generate a first superimposed stream." *Id.* at 8. Similarly, Appellants also acknowledge that Aliprandi "discusses layered blending," but contends that "this is referring to grouping cameras in layers, and is not referring to respective object layers." *Id.*

Further, although Appellants acknowledge that Robotham "discusses transparency information," Appellants contend that Robotham "does not mention respective object layers." *Id.* In particular, Appellants contend that, in Robotham, "[t]he individual images of a visual media element or portions thereof are 'stacked up' in a series of layers and then 'bonded' into a single image sequence," but "this does not suggest respective object layers." *Id.*

Appellants then contend that, even though "the Office Action does not include Yun . . . , Yun does not cure the deficiencies of Aliprandi and

Robotham." *Id.* at 9. In particular, according to Appellants, "Yun does not use any disclosed transparency information to generate a first superimposed stream" as claimed. *Id.*

We have considered all of Appellants' arguments and evidence presented. However, we agree with the Examiner's findings, and find no error with the Examiner's conclusion that claim 1 would have been obvious over the combination of the teachings and suggestions of Aliprandi *and* Robotham, as supported by Yun.

Although Appellants contend that "Aliprandi does not suggest *specifically* that . . . respective color information, depth information, and transparency information pertaining to objects in respective object layers . . . is used to generate a first superimposed stream," and that Aliprandi's "layered blending" is "not referring to respective object layers" (App. Br. 8 (emphasis added)), the Examiner rejects the claim as being obvious over the *combination* of Aliprandi and Robotham. Thus, the issue here is not what Aliprandi "specifically" suggests individually, but rather, the test for obviousness is what the combination of Aliprandi and Robotham teaches or would have suggested to one of ordinary skill in the art. *See In re Merck & Co.*, 800 F.2d 1091, 1097 (Fed. Cir. 1986). As our reviewing court determined, "one cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references." *In re Keller*, 642 F.2d 413, 208 (CCPA 1981).

Here, the Examiner acknowledges that "Aliprandi does not appear to explicitly disclose transparency information pertaining to groups of objects" (Final Act. 4), but relies on "Robotham from the same or similar endeavor" for "teach[ing] transparency information pertaining to groups of objects."

Id. at 5 (citing Robotham 23:34–37). Robotham discloses presenting a finish-quality rendering and/or image transformation of objects in a choreography model, wherein, *for each object*, additional information is maintained in the choreography model *including* depth information and *transparency "alpha" information*, wherein the alpha information is stored to permit proper rendition of inter-object effects such as shadows, reflections and/or refractions, as well as image distorting effects. FF 3. Thus, we find no error with the Examiner's reliance on Robotham to teach and suggest using color information, depth information, and transparency information to generate a first superimposed stream. Final Act. 5.

We are unpersuaded by Appellants' contention that Robotham "does not mention respective object layers." App. Br. 8. As set forth in Appellants' Specification, an "object layer" is "a subset of the objects making up a particular visualization of the scenery." FF 1, Spec. 2:11–14. Here, Robotham teaches and suggests extracting transparency information pertaining to groups of objects making up a particular visualization, which would include subsets of the objects. FF 3.

Further, in Robotham, *layered compositing* is applied, in which the individual images are "stacked up" in a series of layers and then "bonded" into a single image sequence, and wherein, according to an embodiment, a human operator is able to refine an image-based model by adding, changing and deleting an object or group of objects in the model. *Id.* Given the broadest reasonable interpretation of "object layers" consistent with the Specification and claims, we are unpersuaded of error in the Examiner's reliance on Robotham for at least suggesting extracting information including "transparency information pertaining to groups of objects in respective object layers." Final Act. 5.

The Examiner also points to Yun for supporting the finding that the claimed invention is known to an ordinarily skilled artisan. In particular, according to the Examiner, "Yun is used to demonstrate the state of the similar technology known to the ordinary skill in the art." Ans. 4.

Yun discloses a method for encoding/decoding a multi-view video, which includes generating an LDI representing a 3D object with an array of pixels, wherein each LDI pixel is represented by its color, depth and other property information supporting LDI rendering such as alpha/transparency information. FF 4. Similar to the object shown in Figure 1 of Appellants' Invention (FF 1), as shown in Figure 1 of Yun, Yun's LDI is divided into multiple layers, wherein, each LDI layer is separated into individual components: luminance, color, transparency, and depth. FF 4.

Here, similar to Appellants' invention (FF 1, Fig. 3), Aliprandi discloses a method for decoding a multiview stream representing a plurality of viewpoints of a 3D scenery which comprises extracting 2D video streams, blending the views, and using extracted information for generating a superimposed stream. FF 2, Fig. 1a. As Appellants concede, Aliprandi "mentions blending 'all the contributions of the source view to obtain a single frame.'" App. Br. 8. We find no error with the Examiner's combination of: 1) Aliprandi's method of blending of contributions to generate a superimposed stream (FF 2), with 2) Robotham's method of blending of objects to the views using associated mattes or depth information, including transparency information (FF 3), for teaching and suggesting the contested limitations of claim 4. As the Examiner points out, Yun supports the Examiner's finding that extracting transparency information pertaining to groups of objects would have been obvious. FF 4.

Accordingly, Appellants have not shown the Examiner erred in rejecting claim 4, and claims 5–7, and 11–13 (hich are not separately argued and thus falling therewith), over Aliprandi and Robotham.

V. DECISION

We affirm the Examiner's rejection of claims 4–7, and 11–13 under 35 U.S.C. § 103(a).

No time period for taking any subsequent action in connection with this appeal may be extended under 37 C.F.R. § 1.136(a)(1)(iv).

AFFIRMED