UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

Ex parte KIRAN CHALLAPALI

Appeal 2007-2662
Application 09/961,991
Technology Center 2600

Decided: January 10, 2008

Before KENNETH D. HAIRSTON, JOHN A. JEFFERY,
and CARLA. M. KRIVAK, *Administrative Patent Judges*.

KRIVAK, *Administrative Patent Judge*.

DECISION ON APPEAL

Appellant appeals under 35 U.S.C. § 134 (2002) from a final rejection of claims 1-22. We have jurisdiction under 35 U.S.C. § 6(b) (2002).

We reverse.

STATEMENT OF CASE

The invention is directed to a video processing system, method and program product that processes a stream of frames of video data such that the processed stream contains less data than the inputted video data. This is accomplished by a viseme identification system that determines if frames of inputted data correspond to a predetermined viseme, a viseme library that stores the frames that correspond to the at least one predetermined viseme, and an encoder that encodes each frame that corresponds to the predetermined viseme by utilizing a previously stored frame in the viseme library. A decoder decodes the encoded frames using a frame reference library which employs a previously stored frame for decoding a current encoded frame and a morphing system reconstructs frames of video data that were eliminated during the encoding process.

Independent claims 1, 10, 18, and 21 claim a video processing system, a method of processing, a program product, and a decoder, respectively, for processing a stream of frames of video data.

Independent claims 1 and 21, as reproduced below, are representative of the subject matter on appeal.

> 1. A video processing system for processing a stream of frames of video data, comprising a packaging system that includes:
>
> a viseme identification system that determines if frames of inputted video data correspond to at least one predetermined viseme;
>
> a viseme library for storing frames that correspond to the at least one predetermined viseme; and

an encoder for encoding each frame that corresponds to the at least one predetermined viseme, wherein the encoder utilizes a previously stored frame in the viseme library to encode a current frame, and wherein the processed stream contains less data than the inputted video.

21. A decoder for decoding encoded frames of video data that were encoded using frames associated with at least one predetermined viseme, comprising:

a frame reference library for storing decoded frames, wherein the decoder utilizes a previously stored frame in the frame reference library to decode a current encoded frame, and wherein the previously stored frame belongs to the same viseme as the current encoded frame; and

a morphing system that reconstructs frames of video data that were eliminated during an encoding process.

REFERENCES

Chen	US 5,608,839	Mar. 4, 1997

Jörn Ostermann, *Animation of Synthetic Faces in MPEG-4*, 1087-4844/98, IEEE, 49-55, (1998)

The Examiner rejected claims 1-22 under 35 USC § 103(a) based upon the teachings of Chen and Ostermann.

Appellant contends that the applied references neither teach nor suggest to one of ordinary skill in the art modifying the references or combining their teachings, there is no reasonable expectation of successfully modifying the teachings of the cited references, and the references when combined do not teach or suggest all the limitations of the claims (App. Br. 10-11).

ISSUE

The issue before us is did the Examiner err in combining the sound synchronized video system of Chen with the encoder for animation and arithmetic codings of facial animation parameters (FAPs) taught by Ostermann to render the subject matter of claims 1-22 obvious under 35 U.S.C. § 103(a) to one of ordinary skill in the art at the time of the invention?

FINDINGS OF FACT

The following findings of fact are supported by a preponderance of the evidence.

1. Appellant invented a processing system, method, and program product for processing a stream of frames of video. Generally, the system/method encodes inputted video frames and decimates (deletes) video frames that do not have a corresponding viseme in a viseme library such that only a subset of the total number of frames is actually encoded. The frames are encoded using a previously coded frame and/or a frame from a viseme library (Spec. 2, ll. 12-15).

2. The processing system determines if frames of inputted video data correspond to at least one predetermined viseme (Spec. 4, l. 18 through 5, l. 17) in a viseme library that stores frames corresponding to the at least one predetermined viseme (Spec. 5, ll. 18-21).

3. An encoder encodes each frame that corresponds to the at least one predetermined viseme. That is, the encoder utilizes a previously stored

frame in the viseme library to encode a current frame (Spec. 6, ll. 10-13, 19-20).

4. If an inputted frame does not correspond to a viseme, the frame is neither stored in the viseme library nor encoded by the encoder (Spec. 6, ll. 3-5). Rather, a frame detection system eliminates the frames (cl. 5, Spec. 6, ll. 3-9). This allows the processed stream of data to contain less data than the inputted video.

5. A decoder includes a frame reference library that stores decoded frames utilizing a previously stored frame belonging to the same viseme as the current encoded frame (cl. 21; Spec. 7, ll. 8-11). A morphing system reconstructs frames of video data that were eliminated during an encoding process (cl. 21; Spec. 7, ll. 14-16). The encoded frames can be decoded using an image from the reference frame library and an immediately preceding decoded frame (cls. 21-22; Spec 7, ll. 8-10).

6. A viseme is a basic unit of speech in the visual domain that corresponds to a phoneme (which is the basic unit of speech in the acoustic domain). It describes the particular facial and oral movements that occur alongside the voicing of phonemes (Wikipedia).

7. Phonemes and visemes do not share a one-to-one correspondence. Often several phonemes share the same viseme (Wikipedia).

8. A viseme contains many frames of data.

9. Chen teaches a sound-synchronized video system that processes and decodes a stream of unsynchronized digitized audio and video signals, memorizes a plurality of visemes corresponding to phonemes in the audio

signal, fetches visemes corresponding to the phonemes, and synchronizes the decoded audio and video signals by applying the fetched visemes (Abstract; col. 1, ll. 33-41).

10. Chen uses a codec (coder-decoder, CD1, Fig. 1) to decode signals input to a receiver (RE1, Fig. 1). A phoneme recognition module (PR1, Fig. 1) divides the incoming speech components from the codec into recognizable phonemes. A lookup table (LT1, Fig. 1) stores a number of visemes and maps the incoming phonemes into the visemes in the lookup table that specify a corresponding mouth shape (col. 2, ll. 12-22).

11. A feature extraction module (FE1, Fig. 1) in Chen extracts mouth information visemes from the decoded and delayed video signal. A mouth deformation module (MD1, Fig. 1) receives inputs from the video signal along with information from the feature extraction module FE1, and the visemes from the lookup table LT1. It then uses the mouth location generated by the feature extraction module FE1 and replaces the information concerning the mouth shape from the feature extraction module FE1 with the mouth shape specified by the visemes from the lookup table LT1 to provide lip synchronization (col. 2, ll. 23-39).

12. Ostermann teaches MPEG-4 standardization for animation of synthetic faces by specifying a set of face animation parameters (FAPs), each corresponding to a particular facial action deforming a face model in its neutral state (49, section 2, para. 1).

13. A frame of video data contains many FAP values (52, section 3.1 and 55, section 6).

14. In Ostermann, each FAP set contains two high-level parameters: visemes and expressions (FAP group 1) (50, section 2.2). The encoding of the current FAP value depends only on one previously coded FAP value (52, section 3.1). The value of an FAP at a particular time instant kFAP$(i)_k$ is predicted using the previously decoded FAP value FAP$(i)_{k-1}$ (52, section 3.1).

16. To avoid transmitting all FAPs for every frame, the encoder of Ostermann transmits a mask indicating for which groups of FAP values or for which FAPs within a group of FAP values are transmitted, sending incomplete sets of FAPs to the decoder (52, section 3.1).

17. The data received by Ostermann's decoder is arithmetically decoded and added to the previously decoded value to recover the encoded FAP value (52, section 3.1).

18. The decoder extrapolates values of unspecified FAPs to create a complete set of FAPs. However, the standard is vague in specifying how the decoder is to extrapolate FAP values (53, section 3.1).

PRINCIPLES OF LAW

In rejecting claims under 35 U.S.C. § 103, it is incumbent upon the Examiner to establish a factual basis to support the legal conclusion of obviousness. *See In re Fine*, 837 F.2d 1071, 1073, (Fed. Cir. 1988). In so doing, the Examiner must make the factual determinations set forth in *Graham v. John Deere Co.*, 383 U.S. 1, 17 (1966). "[T]he examiner bears the initial burden, on review of the prior art or on any other ground, of

presenting a *prima facie* case of unpatentability." *In re Oetiker*, 977 F.2d 1443, 1445 (Fed. Cir. 1992).

On appeal, Appellant bears the burden of showing that the Examiner has not established a legally sufficient basis for combining the teachings of Chen with those of Ostermann. Appellant may sustain this burden by showing that, where the Examiner relies on a combination of disclosures, the Examiner failed to provide sufficient evidence to show that one having ordinary skill in the art would have done what Appellant did. *United States v. Adams*, 383 U.S. 39 (1966); *In re Kahn*, 441 F.3d 977, 987-988 (Fed. Cir. 2006); *DyStar Textilfarben GmbH & Co. Deutschland KG v. C.H. Patrick, Co.*, 464 F.3d 1356, 1360-1361 (Fed. Cir. 2006). The mere fact that all the claimed elements or steps appear in the prior art is not per se sufficient to establish that it would have been obvious to combine those elements. *United States v. Adams*, *Id.*; *Smith Industries Medical Systems, Inc. v. Vital Signs, Inc.*, 183 F.3d 1347, 1356 (Fed. Cir. 1999).

Where the claimed subject matter involves more than the simple substitution of one known element for another or the mere application of a known technique to a piece of prior art ready for the improvement, a holding of obviousness must be based on "an apparent reason to combine the known elements in the fashion claimed." *KSR Int'l Co. v. Teleflex, Inc.*, 127 S. Ct. 1727, 1740-41 (2007). That is, "there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness." *Id.*, 127 S. Ct. at 1741 (quoting *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006)).

ANALYSIS

The Examiner asserts that Chen teaches all the features of claims 1-22 of Appellant's invention with the exception of the encoder (Ans. 3). The Examiner further asserts that Appellant's "a viseme library for storing frames that correspond to the at least one predetermined viseme" is found in Chen (Ans. 3), wherein the lookup table LT1 of Chen corresponds to the library in Appellant's claim 1. The Examiner cites Ostermann as teaching an encoder and states that the encoder uses "only one previously coded FAP value [to allow] for low-delay communications" (Ans. 3) and thus the data output from the encoder would include less information than the inputted data. Thus, combining Ostermann and Chen would result in the Appellant's invention.

With respect to the Examiner's contention that the lookup table LT1 of Chen corresponds to the viseme library of the invention which stores frames that correspond to at least one predetermined viseme (Ans. 3), Chen's lookup table LT1 stores predetermined visemes for replacing a mouth shape specified by the feature extraction module FE1. Chen does not disclose that frames corresponding to a predetermined viseme are stored in the lookup table LT1 to facilitate encoding each frame as does Appellant.

"Under § 103, the scope and content of the prior art are to be determined; differences between the prior art and the claims at issue are to be ascertained; and the level of ordinary skill in the pertinent art resolved." (*Graham, supra*) First, the scope and content of these references are very different. Chen employs a decoding scheme using visemes and lookup

tables to synchronize audio signals to video signals; whereas Ostermann employs an encoding scheme using MPEG-4 technology and FAPs.

Second, the Examiner acknowledges that Chen does not teach an encoder. Ostermann employs an encoder that encodes FAPs. A value of a FAP at time instant kFAP$(i)_k$ is predicted using the previously decoded value FAP$(i)_{k-1}$. In Ostermann, encoding the current FAP value depends only on the immediately previously coded FAP value (52, section 3.1).

Ostermann also teaches that to avoid transmitting all FAPs for every frame, the encoder transmits a mask indicating for which groups of FAPs values are transmitted. The encoder also specifies for which FAPs within a group values will be transmitted (52, section 3.1).

The Examiner then contends that the encoder of Ostermann could be added to Chen to obtain an apparatus that operates more efficiently by providing low-delay communications (Ans. 3). In this instance there is no suggestion or motivation to add an encoder of this type to Chen. Even if the encoder of Ostermann could be added to Chen, Chen does not employ an MPEG coding scheme which employs FAPs. Chen decodes visemes. Thus, one of skill in the art would not have looked to Ostermann to modify Chen in the manner suggested by the Examiner. A prima facie case of obviousness of claim 1 has not been established by the Examiner.

Claim 21 recites a decoder for decoding encoded frames of video data. The Examiner rejected claim 21 for the reasons applied to claims 7 and 8 (Ans. 5). The Examiner asserts that Chen and Ostermann both disclose a decoder which utilizes a preceding decoded frame to decode a current frame (Ans. 4). The Examiner also asserts that Ostermann discloses a morphing

system that reconstructs frames eliminated by the decimation system: "the decoder is the morphing system, the reconstruction is the extrapolation" (Ans. 5). However, Ostermann does not teach eliminating frames of video data. Ostermann only discloses masking some FAP values within a frame (52, section 3.1) and extrapolating values of corresponding FAPs. Ostermann further states that the "standard is vague in specifying how the decoder is supposed to extrapolate FAP values" (53, section 3.1).

The Examiner has further asserted that the decoder would perform a complimentary operation of the encoder. However, the encoding/decoding of Ostermann cannot be applied to Chen because Chen refers to visemes and not FAP values. Thus, because Chen does not decode in the same manner that Ostermann encodes, the decoding can not be the same.

The Examiner's contention that "it would have been obvious… to take the apparatus disclosed by Chen and add the encoder taught by Ostermann in order to obtain an apparatus that operates more efficiently by providing low-delay communications" (Ans. 3) is not sufficient reasoning for combining these references. The Examiner has failed to provide sufficient evidence to show that one having ordinary skill in the art would have done what Appellant did (*Adams, supra.*). This is more than a simple substitution of one known element for another as asserted by the Examiner. Accordingly, the Examiner has not set forth a prima facie case of obviousness for claim 21.

Based upon the foregoing, it follows that the Examiner has likewise failed to set forth a prima facie case of obviousness for claims 2-20 and 22.

CONCLUSION

We therefore conclude that the Examiner erred in rejecting claims 1-22 under 35 U.S.C. § 103(a).

DECISION

The decision of the Examiner rejecting claims 1-22 is reversed.

REVERSED

KIS

PHILIPS INTELLECTUAL PROPERTY & STANDARDS
P. O. BOX 3001
BRIARCLIFF MANOR, NY 10510