

**UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

APPLE INC.,  
Petitioner

---

IPR2025-01464  
U.S. Patent No. 9,832,017

**PETITION FOR *INTER PARTES* REVIEW  
UNDER 35 U.S.C. § 312 AND 37 C.F.R. § 42.104**

**TABLE OF CONTENTS**

PETITIONER’S EXHIBIT LIST.....	4
I. INTRODUCTION.....	6
II. GROUNDS FOR STANDING.....	7
III. NOTE.....	7
IV. SUMMARY OF THE ’017 PATENT.....	7
A. Overview of the ’017 patent.....	7
B. Prosecution History of the ’017 patent.....	8
C. Effective Filing Date of the ’017 patent.....	8
V. LEVEL OF ORDINARY SKILL IN THE ART.....	9
VI. CLAIM CONSTRUCTION.....	10
VII. RELIEF REQUESTED AND REASONS FOR THE REQUESTED RELIEF.....	11
VIII. IDENTIFICATION OF HOW THE CLAIMS ARE UNPATENTABLE .	11
A. Challenged Claims and Statutory Grounds for Challenges.....	11
B. Ground 1: Claims 6-17 are obvious in view of Spatharis and Manjunath.....	13
1. Spatharis.....	13
2. Manjunath.....	14
3. Reasons to combine Spatharis and Manjunath.....	14
4. Claim 6.....	18
5. Claim 7.....	37
6. Claim 8.....	37
7. Claim 9.....	38
8. Claim 10.....	39
9. Claim 11.....	42
10. Claim 12.....	42
11. Claim 13.....	43
12. Claim 14.....	45

13.	Claim 15.....	45
14.	Claim 16.....	45
15.	Claim 17.....	45
C.	Ground 2: Claims 6-17 are obvious in view of Fuller and Jain .....	45
1.	Fuller.....	45
2.	Jain.....	46
3.	Reasons to combine Fuller and Jain.....	47
4.	Claim 6.....	48
5.	Claim 7.....	71
6.	Claim 8.....	72
7.	Claim 9.....	73
8.	Claim 10.....	75
9.	Claim 11.....	78
10.	Claim 12.....	78
11.	Claim 13.....	79
12.	Claim 14.....	81
13.	Claim 15.....	81
14.	Claim 16.....	81
15.	Claim 17.....	81
IX.	CONCLUSION.....	81
X.	MANDATORY NOTICES.....	83
A.	Real Party-in-Interest .....	83
B.	Related Matters.....	83
C.	Lead and Back-up Counsel and Service Information .....	83
	CERTIFICATE OF WORD COUNT .....	85
	CERTIFICATE OF SERVICE .....	86

**PETITIONER’S EXHIBIT LIST**

Ex.1001	U.S. Patent No. 9,832,017
Ex.1002	Prosecution History of U.S. Patent No. 9,832,017
Ex.1003	Declaration of Dr. Nathaniel Polish under 37 C.F.R. § 1.68
Ex.1004	<i>Curriculum Vitae</i> of Dr. Nathaniel Polish
Ex.1005	U.S. Patent No. 6,833,865 (“Fuller”)
Ex.1006	U.S. Patent No. 7,295,752 (“Jain”)
Ex.1007	<i>Graphics File Formats: Second Edition</i> , James D. Murray et al. (1996) (“Murray”)
Ex.1008	Reserved
Ex.1009	U.S. Pub. No. 2006/0227995 (“Spatharis”)
Ex.1010	<i>Introduction to MPEG-7: Multimedia Content Description Interface</i> , B.S. Manjunath et al. (editors) (2002) (“Manjunath”)
Ex.1011	U.S. Patent No. 2,618,191 (“Martin”)
Ex.1012	U.S. Patent No. 2,983,793 (“Weber”)
Ex.1013	U.S. Pub. No. 2001/0015759 (“Squibbs”)
Ex.1014	U.S. Pub. No. 2003/0115219 (“Chadwick”)
Ex.1015	U.S. Pub. No. 2002/0184244 (“Hsiao”)
Ex.1016	U.S. Patent No. 6,961,441 (“Hershey”)
Ex.1017	U.S. Patent No. 7,216,232 (“Cox”)
Ex.1018	Markman Order, <i>MyPort, Inc. v. Samsung Electronics Co., Ltd. et al.</i> , 2-22-cv-00114 (EDTX)
Ex.1019	U.S. Pub. No. 2005/0169245 (“Hindersson”)
Ex.1020	U.S. Patent No. 5,875,233 (“Cox-2”)
Ex.1021	<i>The Authoritative Dictionary of IEEE Standards Terms</i> (2000) (“IEEE Dictionary”)

Ex.1022	U.K. Pub. No. 2,430,101 (“Bober”)
Ex.1023	U.S. Patent No. 7,630,545 (“Cieplinski”)
Ex.1024	PCT Pub. No. 2006/076760 (“Yu”)
Ex.1025	U.S. Patent No. 6,930,703 (“Hubel”)
Ex.1026	U.S. Pub. No. 2004/0218080 (“Stavely”)

## I. INTRODUCTION

The Board has already found that the prior art presented in this Petition, Fuller (Ex.1005), likely renders obvious claims 6-17<sup>1</sup> (the “Challenged Claims”) of U.S. Patent No. 9,832,017 (“the ’017 patent”). The Board granted institution based upon Fuller in a previous IPR that settled prior to final written decision. *See Samsung Electronics Co. v. MyPort, Inc.*, Paper 42, IPR2023-00023 (Feb 16, 2024) (“Samsung IPR”).

This Petition additionally presents a new ground based on the Spatharis reference (Ex.1009). Like Fuller, Spatharis shows that the concepts recited in the ’017 patent were well-known. The ’017 patent relates to capturing media such as images, audio, or video and then performing “speech-to-text” and “image recognition” to derive “context tags.” *See* Ex.1001, 5:62-6:14. Spatharis likewise describes a “metadata extraction engine” that “provides analysis” of an “audio/video signal” including “speech to text conversion” of captured digital audio signals, and “visual properties face recognition” of captured digital images to create “tags, annotations, and/or markings.” Ex.1009 ¶¶13, 14, 42.

Therefore, pursuant to 35 U.S.C. §§ 311, 314(a), and 37 C.F.R. § 42.100,

---

<sup>1</sup> Patent Owner previously disclaimed Claims 1-5 of the ’017 patent. Ex.1002, 58-62.

Apple Inc. (“Petitioner”) respectfully requests that the Board institute this IPR and cancel as unpatentable the Challenged Claims under (pre-AIA) 35 U.S.C. §103(a).

## II. GROUNDS FOR STANDING

Petitioner certifies that the ’017 patent is eligible for IPR and that Petitioner is not barred or estopped from requesting IPR challenging the patent claims. 37 C.F.R. § 42.104(a).

## III. NOTE

Petitioner cites to exhibits’ original page numbers whenever feasible.

**Emphasis** in quoted material has been added. Claim terms are presented in *italics*.

## IV. SUMMARY OF THE ’017 PATENT

### A. Overview of the ’017 patent

The ’017 patent describes a “capture device 100” for capturing content such as “[s]till pictures, moving pictures, audio, telemetry or other information.”

Ex.1001, 3:52-56. The content is processed by a “data converter” included in the capture device, which the ’017 patent describes as “any type of device that will capture the information and place it in some type of digitized format.” Ex.1001, 3:54-65. The patent further describes that other “meta data [sic]” is captured in addition to the content, such as a “time and date” or “location” associated with the captured content. Ex.1001, 4:4-12. The content may also be associated with “[c]ontext tags” derived from the content, such as the tags “animal,” or “dog,” or

“Spot,” from a still picture of a dog. Ex.1001, 5:39-46. The captured content is stored in association with the context tags and the additional metadata. Ex.1001, 5:47-58.

**B. Prosecution History of the '017 patent**

The '017 patent was filed on September 21, 2016, and claims priority through a chain of several applications to a provisional application filed September 30, 2002.<sup>2</sup> Ex.1001, Face.

The Office issued a first office action that included only a double patenting rejection. Ex.1002, 135-38. In response, Applicant filed a terminal disclaimer to overcome the double patenting rejection. Ex.1002, 119-20. The Office then allowed all claims and did not provide any reasons for allowance. Ex.1002, 96-7.

**C. Effective Filing Date of the '017 patent**

The '017 patent claims the benefit of two provisional applications and several earlier applications. Ex.1001, Face. One of those applications (11/621,062) is a continuation-in-part and was filed January 8, 2007. *Id.* The '017 patent is entitled to an effective filing date of no earlier than January 8, 2007.

Each independent claim recites “creating an image recognition searchable context tag with image recognition” and this feature is not disclosed in the previous

---

<sup>2</sup> Petitioner does not concede that the '017 patent is entitled to this priority date.

applications. Rather, the concepts related to metadata context tags were added as new subject matter in the continuation-in-part application filed January 8, 2007.<sup>3</sup>

Notably, the Patent Owner did not dispute the previous Petitioner’s priority date arguments in the previous IPR (IPR2023-00023). Rather, the Patent Owner stated that “[f]or purposes of this Response only, MyPort has applied this January 8, 2007 date as the earliest effective filing date for the ’017 patent.” IPR2023-00023, Paper 25, 5.

In this Petition the phrase “effective filing date” and “time of the invention” refers to January 8, 2007.

## **V. LEVEL OF ORDINARY SKILL IN THE ART**

A Person of Ordinary Skill in the Art (“POSITA”) on January 8, 2007, would have had a working knowledge of audio and visual content capture systems that are pertinent to the ’017 patent. Ex.1003 ¶19. That person would have a bachelor’s degree in computer science, computer engineering, electrical engineering, or equivalent training, and approximately two years’ experience

---

<sup>3</sup> Patent Owner bears the burden of production on this issue. *Nat. Alternatives Int’l, Inc. v. Iancu*, 904 F.3d 1375, 1380 (Fed. Cir. 2018) (“[C]laims...are not entitled to priority under § 120 at least until the patent owner proves entitlement to the PTO, the Board, or a federal court.”).

working with audio and visual content capture systems. Lack of work experience can substitute for additional education, and vice versa. Ex.1003 ¶19.

## VI. CLAIM CONSTRUCTION

Claim terms in IPR are construed according to their “ordinary and customary meaning” to a POSITA. 37 C.F.R. § 42.100(b). *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (*en banc*). For the purposes of this proceeding and the grounds presented herein, no claim term requires express construction. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017). Ex.1003 ¶26.

The ’017 patent was the subject of previous litigation in which the following claim terms were construed. *See* Ex.1018 (Markman Order, *MyPort, Inc. v. Samsung Electronics Co., Ltd. et al*, 2-22-cv-00114 (EDTX)).

Claim Term	Construction From Prior Litigation
“ <i>context tag</i> ” (claims 6, 10, 12, 13)	“a searchable element derived from either a data element itself or from the context description element”
“ <i>associating the text and image recognition context tags with the digital image</i> ” (claims 6, 13) “ <i>storing [ . . . ] the digital image in association with the text and image recognition context tags</i> ” (claims 6, 13)	Plain and ordinary meaning
“ <i>transmitting [transmits] the associated stored digital image</i> ”	Plain and ordinary meaning

Claim Term	Construction From Prior Litigation
<i>in association with the text and image recognition context tags”</i> (claims 10, 12)	
<i>“a first digital audio format”</i> (claims 13, 16)	Plain and ordinary meaning
<i>“the internal storage storing the digital image in association with the text and image recognition context tags”</i> (claim 6)	Plain and ordinary meaning
<i>“image source”</i> (claims 13, 16)	“an internal or an external image source”

The prior art presented herein teaches all limitations of the Challenged Claims regardless of whether the Board adopts the district court’s constructions.

**VII. RELIEF REQUESTED AND REASONS FOR THE REQUESTED RELIEF**

Petitioner asks that the Board institute an IPR proceeding and cancel the Challenged Claims in view of the analysis below.

**VIII. IDENTIFICATION OF HOW THE CLAIMS ARE UNPATENTABLE**

**A. Challenged Claims and Statutory Grounds for Challenges<sup>4</sup>**

Ground	Claims	Basis: 35 U.S.C. § 103 (Pre-AIA) over
1	6-17	Spatharis and Manjunath
2	6-17	Fuller and Jain

---

<sup>4</sup> Petitioner relies on the teachings of the cited prior art references, and not on a

U.S. Patent Publication No. 2006/0227995 (Ex.1009, “Spatharis”) was filed April 6, 2006, and published on October 12, 2006. Spatharis is thus prior art to the ’017 patent under pre-AIA 35 U.S.C. 102(a) and 102(e).

The book “Introduction to MPEG-7: Multimedia Content Description Interface,” ISBN 0-471-48678-7, edited by B.S. Manjunath *et al.* (Ex.1010, “Manjunath”) was published in 2002. Ex.1010, 5. Manjunath was a frequently cited resource for POSITAs on MPEG-7. Multiple patents and patent publications published prior to the effective filing date of the ’017 patent (January 8, 2007) include citations and references to Manjunath, indicating that it was publicly accessible and findable by POSITAs prior to the effective filing date. *See* Cieplinski (Ex.1023), 1:48-56 (referring to “Introduction to MPEG-7 Multimedia content description interface Edited by Manjunath, Salembier and Sikora, ISBN 0-471-48678-7”); Bober (Ex.1022), 7:8-12; Yu (Ex.1024), 5:23-27; Ex.1003 ¶35. Manjunath is thus prior art to the ’017 patent under pre-AIA 35 U.S.C. 102(b).

U.S. Patent No. 6,833,865 (Ex.1005, “Fuller”) was filed July 29, 1999 and issued December 21, 2004. Fuller is thus prior art to the ’017 patent under pre-AIA 35 U.S.C. 102(b).

---

physical incorporation of elements. *See In re Mouttet*, 686 F.3d 1322, 1332 (Fed. Cir. 2012); *In re Etter*, 756 F.2d 852, 859 (Fed. Cir. 1985); Ex.1003 ¶57.

U.S. Patent No. 7,295,752 (Ex.1006, “Jain”) was filed February 5, 2002 and issued on November 13, 2007. Jain is thus prior art to the ‘017 patent under pre-AIA 35 U.S.C. 102(e).

**B. Ground 1: Claims 6-17 are obvious in view of Spatharis and Manjunath**

**1. Spatharis**

Spatharis describes a “system for extracting, processing, and sending metadata associated with audio data and/or video data.” Ex.1009, Abstract. Specifically, Spatharis describes a “camera system” that “captures...content” including “sound and/or images” and “extract[s] and/or process[es] metadata from the captured content.” Ex.1009 ¶12. The camera system includes a CPU subsystem.<sup>5</sup> Ex.1009 ¶23. The camera system further includes a “storage device” for “stor[ing] content that is acquired, generated, or processed by any of the subsystems” of the camera system. Ex.1009 ¶17.

Spatharis further describes that the camera system includes a “metadata extraction engine” that “provides analysis of the audio/video signal” including

---

<sup>5</sup> The analysis below relies on distinct functions (e.g., analog-to-digital conversion, image processing, and combining) of the CPU subsystem 270 for different claim elements (converter, camera, and combiner).

“speech to text conversion” of captured digital audio signals, and “visual properties face recognition” of captured digital images. Ex.1009 ¶42. Spatharis describes that this extracted metadata is associated with and stored with the digital image. Ex.1009 ¶42; Ex.1003 ¶¶32-33.

## **2. Manjunath**

Manjunath describes the “MPEG-7” format—also referred to as the “Multimedia Content Description Interface”—which “standardizes the description of multimedia content supporting a wide range of applications.” Ex.1010, 22. Manjunath describes this format as enabling the “search and retrieval” of multimedia content based on metadata (e.g., context tags) extracted from the content. Ex.1010, 22. For example, Manjunath teaches a process of describing “multimedia content using natural language text” called “text annotation,” and describes the “practice of using text annotations to search, catalogue and index multimedia content” as “both longstanding and widespread.” Ex.1010, 127; Ex.1003 ¶¶34-35.

## **3. Reasons to combine Spatharis and Manjunath**

A POSITA would have found it obvious, and indeed would have been motivated, to organize the metadata extracted by Spatharis’ camera system as a text-based searchable file consistent with the MPEG-7 standard as described in Manjunath. Ex.1003 ¶36.

As a threshold matter, Spatharis and Manjunath are analogous art because both references are in the same field of endeavor as the '017 patent. *See In re Bigio*, 381 F.3d 1320, 1325 (Fed. Cir. 2004). The '017 patent relates to capturing and storing audio and image content in association with context tags generated based on the content. Ex.1001, Abstract, 2:15-18, 5:39-46. Spatharis and Manjunath are similarly related to capturing and storing audio and image content in association with context tags generated based on the content. Ex.1009 ¶¶12, 13, 16 (describing a method for “captur[ing]...content,” such as “sound and/or images” and “extract[ing] and/or process[ing] metadata from the captured content...that includes, for example, data, tags, annotations, and/or markings that can be associated with captured content”); Ex.1010, 127-28 (describing the “MPEG-7” format for storing text annotations in association with multimedia content); Ex.1003 ¶37. Thus, Spatharis and Manjunath are analogous art to the '017 patent.

Spatharis describes a system “for extracting, processing, and sending metadata associated with audio data and/or video data.” Ex.1009, Abstract. Spatharis further explains that “[m]etadata extracted from captured content and/or raw (i.e., original) content can be stored in, for example, a storage subsystem (not shown) or in database 470. The data can be stored in **one or more formats.**” Ex.1009 ¶44 *see also id.* ¶26. Spatharis thus mentions data formats generally,

leading a POSITA to look to known formats for metadata storage. Ex.1003 ¶38.

Manjunath is a textbook describing known techniques for metadata storage, including the MPEG-7 standard. Manjunath describes that “[t]he practice of using text annotations to **search**, catalogue and index multimedia content is both **longstanding and widespread**” and “MPEG-7 supports this practice with the TextAnnotation data type that allows free text, keyword, structured and dependency structure annotations.” Ex.1010, 127 (italics omitted). A POSITA would have thus been motivated to organize the metadata in Spatharis according to the “longstanding and widespread” techniques described in Manjunath both to increase the utility of the resulting digital image to users by making the metadata searchable, and to conform the resulting digital image and associated metadata conform to the industry-standard MPEG encoding standard described by Manjunath. Ex.1010, 127; Ex.1003 ¶39.

Implementing Spatharis’ metadata format using the industry-standard MPEG-7 format represents a simple combination of prior art elements (the extracted metadata of Spatharis with Manjunath’s MPEG-7 encoding format), according to known methods to yield predictable results (the extracted metadata of Spatharis formatted and stored according to Manjunath’s MPEG-7 encoding format). *See KSR Int’l v. Teleflex Inc.*, 550 U.S. 398, 416 (2007). A POSITA would have been motivated to perform the combination to achieve the same

benefits and improve the Spatharis system “in the same way” as the similar systems described in *Manjunath. KSR*, 550 U.S. at 417 (finding obviousness when a known technique “would improve similar devices in the same way”); Ex.1003 ¶40. In particular, implementing Spatharis with a tried-and-true industry standard such as MPEG-7 provides the benefit of wide-ranging interoperability. Ex.1010, 22 (“[T]he need for **interoperability between devices** has also been recognized and several standardization activities have been launched. MPEG-7, also called ‘Multimedia Content Description Interface’, standardizes the description of multimedia content supporting a **wide range of applications.**”). Ex.1003 ¶40.

A POSITA would have had a reasonable expectation of success in making such a combination because *Manjunath* describes systems operating in the proposed manner. Ex.1010, 127-128; Ex.1003 ¶41. *Manjunath* “offers a practical step-by-step walk through of the components, from systems to schemas to audio-visual descriptors.” Ex.1010, 22-23. “It addresses the selection of the multimedia features to be described, the organization and structuring of the description, the language to instantiate the description, as well as the major processing tools used for indexing and retrieval of images and video sequences.” Ex.1010, 23.

*Manjunath*’s detailed textbook instructions thus provide predictability, leading to an expectation of success. Indeed, the combination is nothing more than “the predictable use of prior art elements according to their established functions.” *KSR*,

550 U.S. at 417; Ex.1003 ¶41.

#### 4. Claim 6

**[6.0]** *A system for capturing image and audio information for storage, comprising:*

To the extent the preamble is limiting, it is taught, in the combination, by Spatharis.

**First**, Spatharis describes a “camera system” for “**captur[ing]...content,**” such as “**sound and/or images**” (“*system for capturing image and audio information*”), and “extract[ing] and/or process[ing] metadata from the captured content.” Ex.1009 ¶12; Ex.1003 ¶¶42-44.

**Second**, the image and audio information in Spatharis is captured “*for storage*” because the camera system includes a “storage device” for “**stor[ing] content** that is acquired, generated, or processed by any of the subsystems.” Ex.1009 ¶17; Ex.1003 ¶45.

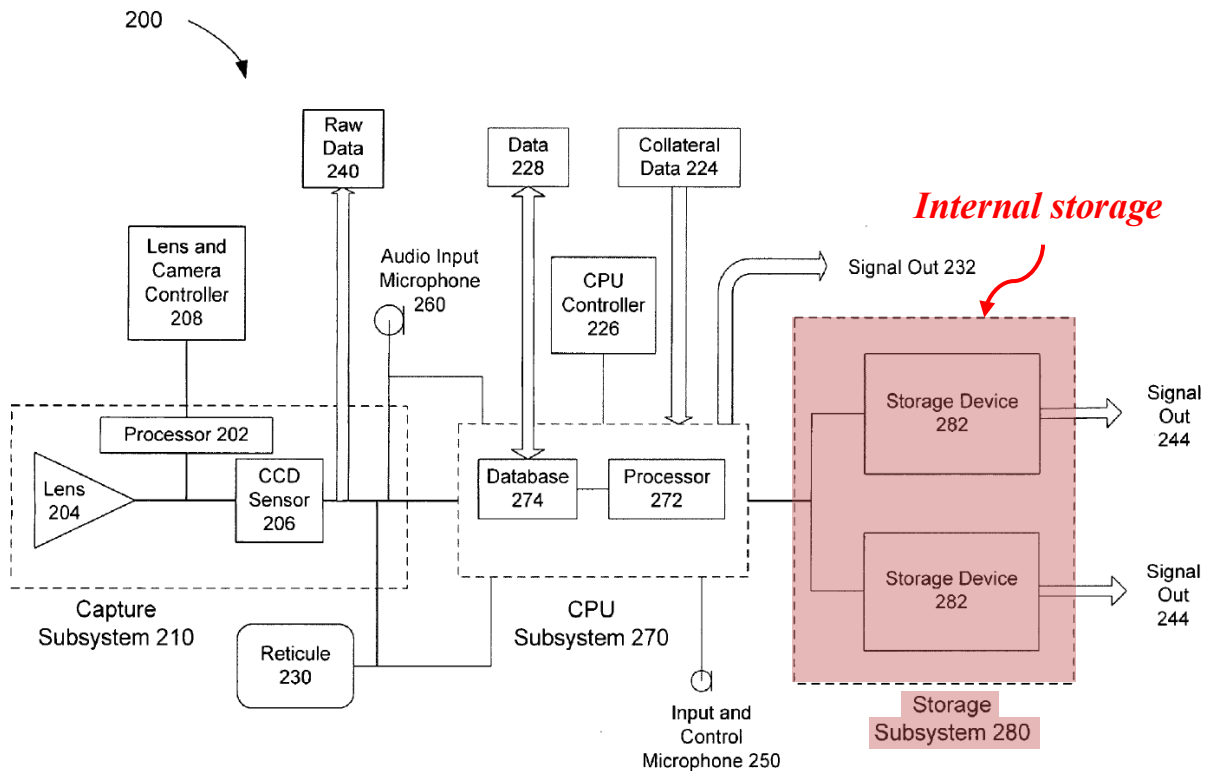
Thus, because Spatharis describes a camera system that captures image and audio data, which are then stored in a storage device, Spatharis in the combination renders obvious “*a system for capturing image and audio information for storage.*” Ex.1003 ¶¶42-46.

**[6.1]** *a capture device having:*

As described in [6.0], in the combination, Spatharis describes a “camera system” (“*a capture device*”) that “**captures...content,**” such as “**sound and/or images.**” Ex.1009 ¶12. Spatharis thus renders obvious “*a capture device.*” Ex.1003 ¶47.

**[6.2] *internal storage;***

As described in [6.0], in the combination, Spatharis’ camera system includes a “storage device” for “stor[ing] content that is acquired, generated, or processed by any of the subsystems.” Ex.1009 ¶17. As shown in Spatharis’ FIG. 2, the camera system 200 includes a “**storage subsystem 280**” (“*internal storage,*” annotated in red) that “store[s] raw content (i.e., unenriched content) or content that has been enriched (e.g., processed and associated with metadata).” Ex.1009 ¶26. Ex.1003 ¶48.



**Ex.1009, Fig. 2 (annotated); Ex.1003 ¶48.**

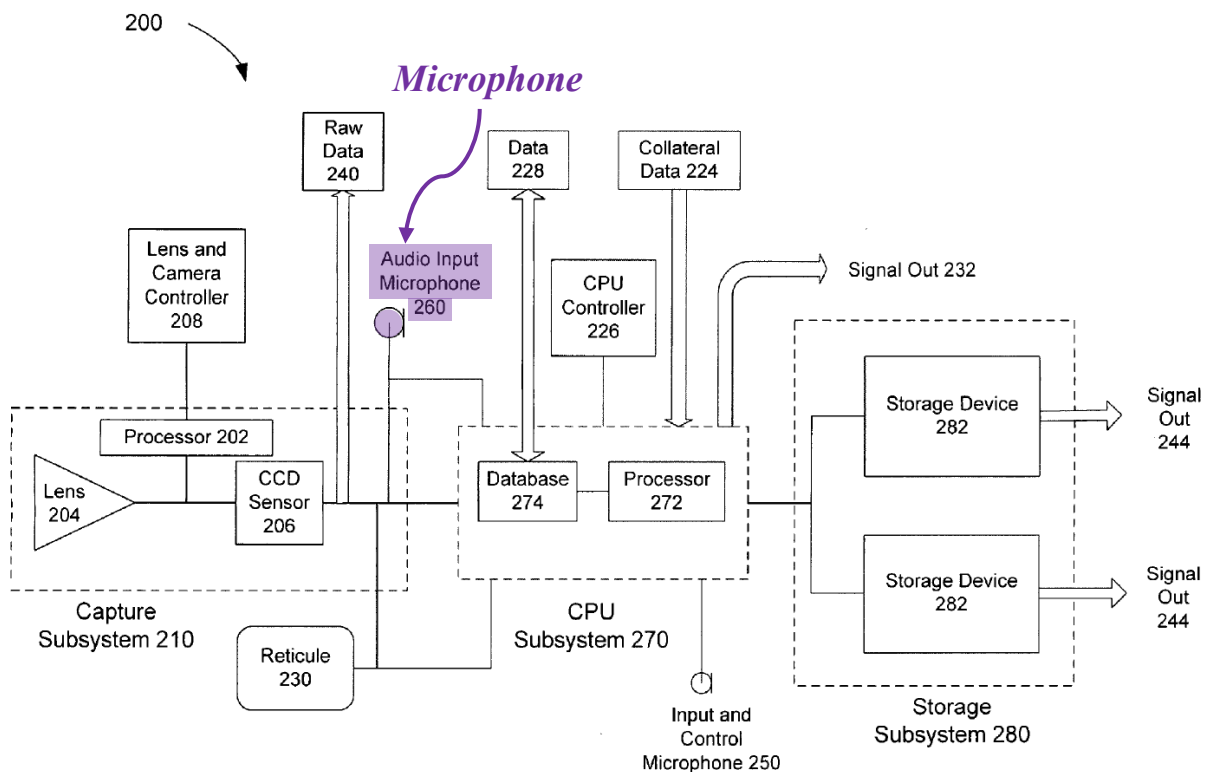
Thus, because Spatharis’ camera system includes storage devices, Spatharis in the combination renders obvious “*internal storage.*” Ex.1003 ¶¶48-49.

**[6.3] a microphone interfacable [sic] with and [sic, an] external audio information source that generates external audio information and**

In the combination, Spatharis’ camera system includes a “capture subsystem 210” that “includes a source audio input microphone 260” that captures “audio” as “analog signals.” Ex.1009 ¶24. Spatharis’ camera system captures the analog audio signals by “interfac[ing]” the microphone 260 with an external environment in which an object (“an external audio information source”) is making a sound (“generates external audio information”) consistent with the conventional

operating details of microphones that were well-known many years prior to the '017 patent. Ex.1003 ¶51; *see, e.g.*, Ex.1011, 7:62-66 (issued in 1952, and describing that the “sound of [an] instrument...is picked up by the microphone 25, and a corresponding electrical signal is obtained”); Ex.1012, 5:70-73 (issued in 1961, and describing a “microphone...in use by a speaker” producing “[s]peech signals” that “are amplified by audio amplifier 14”).

Spatharis' FIG. 2 shows the microphone 260 included in the camera system 200:



**Ex.1009, Fig. 2 (annotated); Ex.1003 ¶51.**

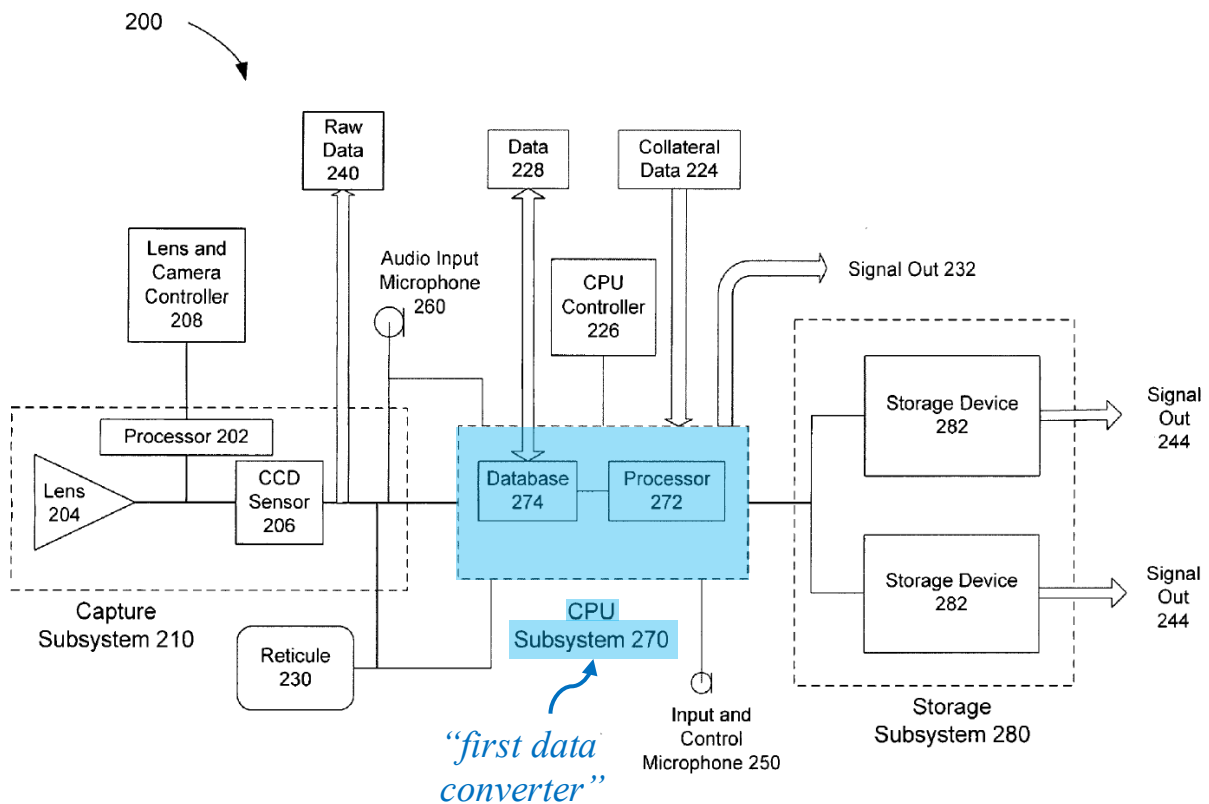
Thus, because Spatharis' camera system includes a microphone for capturing sound from the environment, Spatharis in the combination renders obvious "*a microphone interfaceable with [an] external audio information source that generates external audio information.*" Ex.1003 ¶¶52-53.

To the extent Patent Owner argues that the claimed "*external audio information*" must include an audio description of the captured image (*see* [6.6]-[6.7]), Petitioner notes that although the claims do not recite any such limitation, Spatharis nonetheless teaches this feature. Spatharis describes that the "camera system 200 includes **an input and control microphone 250 that can receive commentary** from, for example, a camera system operator **assisting in the identification of shots and/or describing the content.**" Ex.1009 ¶32. Spatharis teaches that audio captured by the input and control microphone 250 is processed and stored in the same manner as audio captured by the source audio input microphone 260. *See* Ex.1009 ¶32; Ex.1003 ¶54.

**[6.4] *a first data converter for capturing the first external audio information from the microphone,***

In the combination, Spatharis describes that the "**CPU subsystem 270**" includes one or more "[p]rocessors" (e.g., 272) which "can be specialized modules (e.g., ASICS or digital signal processors) dedicated to performing specific functions" such as analog to digital conversion. Ex.1009 ¶39; Ex.1003 ¶¶55-56.

This analog to digital conversion functionality of the CPU subsystem 270 teaches “a first data converter.” Spatharis further teaches that the “**audio** captured by the” source audio input microphone 260 (“*the first external audio information from the microphone*”) “can be analog signals that are **converted into digital signals/data**” (“*captur[ed]*”) by the “**CPU subsystem 270**” (the “*first data converter*”) included in the camera system 200. Ex.1009 ¶¶23-24; Ex.1003 ¶56. Spatharis’ FIG. 2 shows this configuration:



Ex.1009, Fig. 2 (annotated); Ex.1003 ¶56.

Thus, because Spatharis’ camera system includes a CPU subsystem that converts the audio signal from the microphone to digital data, Spatharis in the

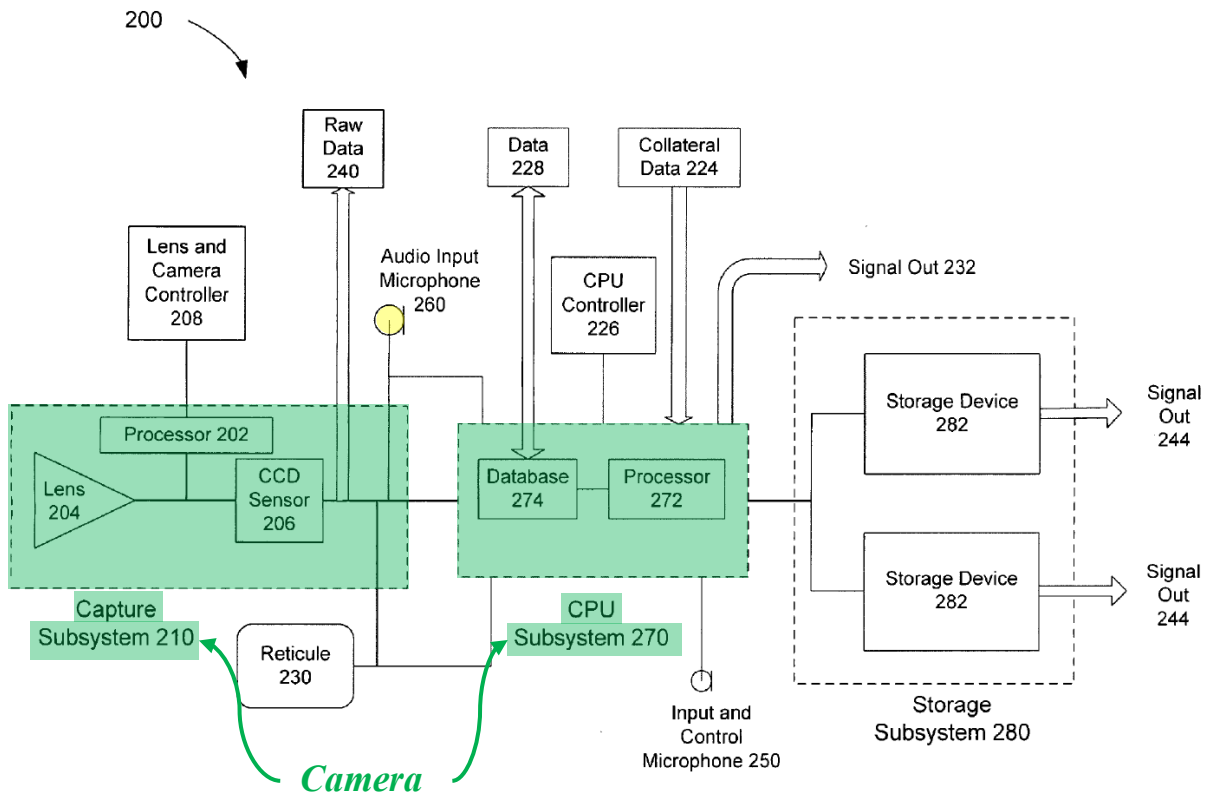
combination renders obvious “*a first data converter for capturing the first external audio information from the microphone.*” Ex.1003 ¶¶55-57.

**[6.5] *a camera interfacing with and [sic, an] external image source to capture an image therefrom;***

**First**, Spatharis teaches a “*camera*” in that the camera system 200 includes a “**capture subsystem 210.**” Ex.1009 ¶23. Spatharis describes that the “capture subsystem 210” includes a “**still camera**” for acquiring the images that “can include any combination of one or more lenses 204” and “CCD 206 sensors.” Ex.1009 ¶24. The capture subsystem 210, including the “still camera,” and the image processing functionality of the CPU subsystem 270, collectively, teach the claimed “*camera.*” Ex.1009 ¶¶23-24; Ex.1003 ¶58.<sup>6</sup> Spatharis’ FIG. 2 shows this configuration:

---

<sup>6</sup> Before the effective filing date of the ’017 patent (and still today), the term “*camera*” in the context of digital cameras referred to both the image sensor and associated processor, including a processor (e.g., CPU subsystem 270 in Spatharis) as part of a camera was a well-known configuration in digital camera systems. Ex.1003 ¶58; *see, e.g.*, Hubel (Ex.1025), 2:51-54 (describing a digital camera “for capturing a plurality of images during a pan of a scene,” and explaining that the “camera may broadly be viewed as including an imager and a processor.”); Stavelly



**Ex.1009, Fig. 2 (annotated); Ex.1003 ¶58.**

**Second**, the capture subsystem “acquires images” that “are processed by the CPU subsystem 270” (“interface[es]...with an external image source to capture an image therefrom.”) Ex.1009 ¶23. The object having its picture taken is an “external image source” as claimed. Ex.1003 ¶59.

(Ex.1026 ¶14) (describing a “digital camera” comprising “a lens 12...an image sensor 13 for receiving images transmitted by the imaging optics 12” and a “processor 14 (or microprocessors 14) is coupled to the image sensor 13.”).

Thus, because Spatharis' system acquires images with a visual capture subsystem, such as a CCD sensor and processor, Spatharis in the combination renders obvious "*a camera interfacing with [an] external image source to capture an image therefrom.*" Ex.1003 ¶¶58-60.

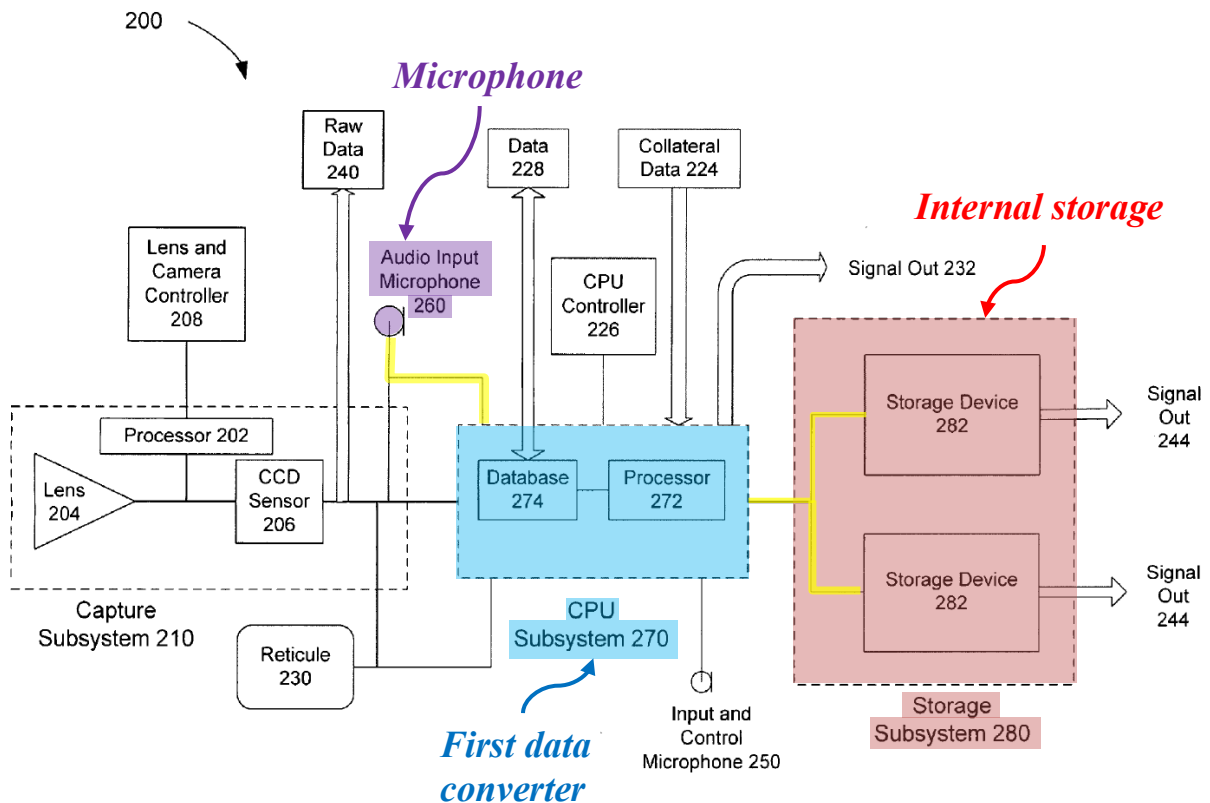
**[6.6] *the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio in internal storage within the capture device,***

**First**, as previously discussed in [6.4], in the combination, Spatharis' "CPU subsystem 270" teaches "*a first data converter.*" Ex.1003 ¶61. In addition, as previously discussed in [6.3], Spatharis teaches that the CPU subsystem 270 converts the first external audio information from the microphone into a "digital signals/data." Ex.1009 ¶24.

**Second**, Spatharis' "CPU subsystem 270" ("*the first data converter*") "**processes**" the digital audio signals ("*the captured external audio information*") and "**transmit[s]**" the processed audio signals "**to the storage subsystem 280**" for storage ("*storing it...as stored digital audio in internal storage within the capture device*"). Ex.1009 ¶23; Ex.1003 ¶62.

**Third**, Spatharis teaches that the processed audio signals "can be stored in the storage subsystem 280 **in many formats**" ("*stor[ed] in a first digital audio format*"). Ex.1009 ¶26; *see also id.* ¶44; Ex.1003 ¶63.

Spatharis' FIG. 2 shows this configuration, with the yellow highlighting showing the path of *captured external audio information* from the *microphone* to the CPU subsystem 270 (*“the first data converter”*) to the storage subsystem 280 (*“the internal storage”*). Ex.1009 ¶¶23, 24, 26; Ex.1003 ¶64:

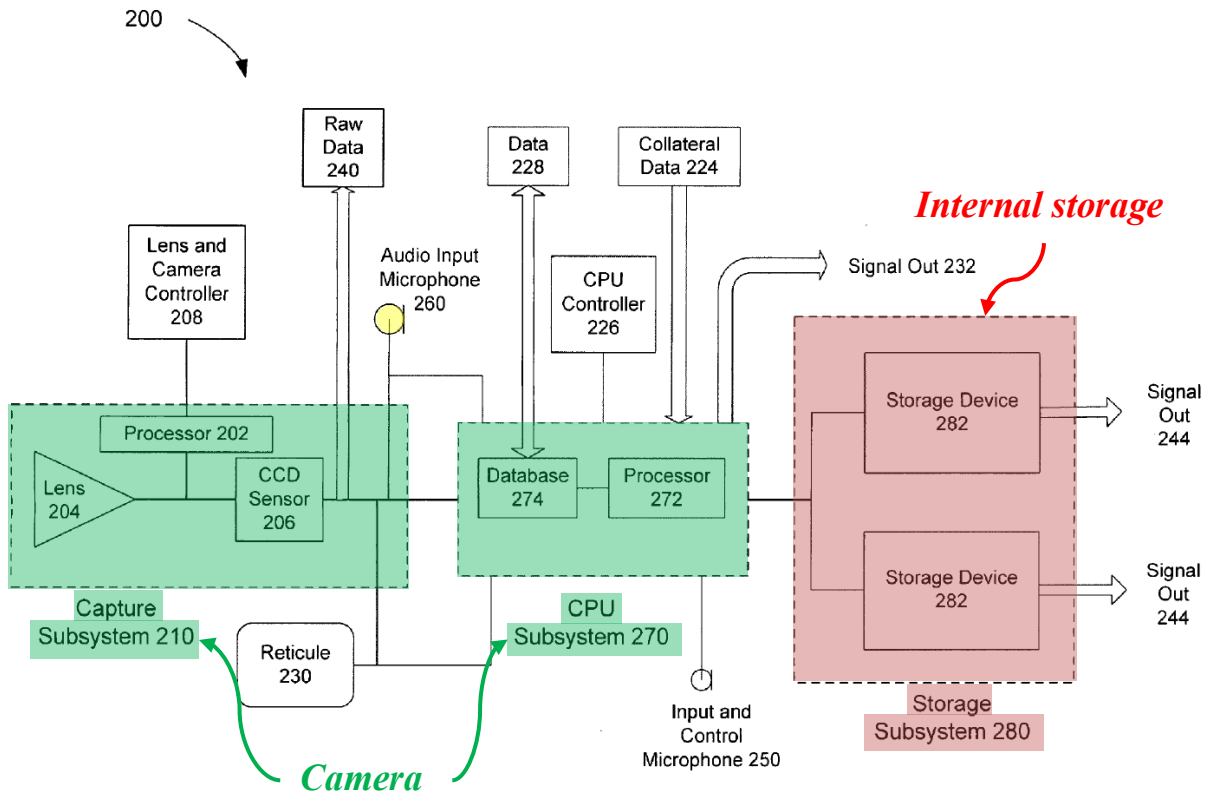


Ex.1009, Fig. 2 (annotated); Ex.1003 ¶64.

Thus, Spatharis describes a CPU subsystem that processes and stores audio data in one or more data formats, which in the combination renders obvious *“the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio in internal storage within the capture device.”* Ex.1003 ¶¶61-65.

**[6.7] the camera for processing the captured image and storing it as a stored digital image in internal storage and**

**First**, as described in [6.5] and shown in FIG. 2, Spatharis' capture subsystem 210, including the "still camera," and CPU subsystem 270, collectively, teach the claimed "camera." In addition, as described in [6.2], Spatharis' camera system 200 includes a storage subsystem 280 ("internal storage"). Spatharis' FIG. 2 shows this configuration:



**Ex.1009, Fig. 2 (annotated); Ex.1003 ¶66.**

**Second**, Spatharis teaches that the capture subsystem 210 “acquires images” that “are **processed by the CPU subsystem 270**” (“*the camera for processing the captured image*”). Ex.1009 ¶23; Ex.1003 ¶67.

**Third**, Spatharis teaches that “[a]fter the CPU subsystem 270 processes the” captured image, “it is transmitted to **the storage subsystem 280**” for storage (“*stor[ed] ... as a stored digital image in internal storage*”). Ex.1009 ¶23; Ex.1003 ¶68.

Thus, because the capture system passes the captured information to the CPU subsystem 270 for processing and storage as a digital image, Spatharis in the combination renders obvious “*the camera for processing the captured image and storing it as a stored digital image in internal storage.*” Ex.1003 ¶¶66-69.

**[6.8] a combiner for generating an association between the stored digital audio and the stored digital image,**

Spatharis’ CPU subsystem additionally functions as a “*combiner for generating an association between the stored digital audio and the stored digital image*” because it stores the audio and video together in storage. *See* Ex.1009 ¶¶24-26, FIG. 2; Ex.1003 ¶70.

In the combination, Spatharis’ CPU subsystem 270 receives both audio and video data: “The captured digital audio and/or video data, which can be **collectively referred to as content**, are made available to the CPU subsystem 270

for processing.” Ex.1009 ¶24. The “capture subsystem 210 can be any type of ordinary or sophisticated still camera or video recorder that can include any combination of one or more lenses 204, CCD 206 sensors and/or microphones 260.” Ex.1009 ¶24.

Spatharis teaches that “[a]fter the CPU subsystem 270 processes” the content, “it is transmitted to the storage subsystem 280” and stored. Because the storage system stores “content,” which Spatharis indicates is collectively audio and video, the stored content as processed by the CPU subsystem is a combination of audio and video data (e.g., a video clip with sound). *See* Ex.1009 ¶¶24-26; Fig. 2. The CPU subsystem 270 thus functions as a “combiner.” Ex.1003 ¶72.

In other words, the functionality of the CPU subsystem 270 that stores the captured audio and video data (“*the stored digital audio and the stored digital image*”) together (in *association*) in the storage subsystem 280 as “stored content,” teaches “*a combiner for generating an association between the stored digital audio and the stored digital image.*” Ex.1003 ¶¶70-73.

**[6.9] a media data converter for converting the received set of captured information to convert the received digital audio to a text based searchable file as a text context tag and**

**First**, Spatharis describes a metadata extraction engine (“*media data converter*”) that provides “speech to text conversion” (“*converting...the received digital audio to a text based...file.*”). Ex.1009 ¶42; Ex.1003 ¶74. Spatharis

describes that the camera system includes a CPU subsystem that “**executes applications for manipulating content captured by the capture subsystem 210.**” Ex.1009 ¶29; *see also id.* ¶15 (the CPU subsystem “can process and be programmed with software applications such as metadata engines that analyze, for example, images in real-time and extract information about content that has been acquired”). Spatharis describes that the CPU subsystem executes a “**metadata extraction engine**” (“*media data converter*”) that “provides analysis of the audio/video signal” including “**speech to text conversion**” of the digital audio signal (“*the received digital audio*”). Ex.1009 ¶42. Spatharis further describes that the metadata extracted by the metadata extraction engine includes “**tags, annotations, and/or markings that can be associated with captured content.**” Ex.1009 ¶13; Ex.1003 ¶74. Spatharis’ FIG. 4 shows this configuration:

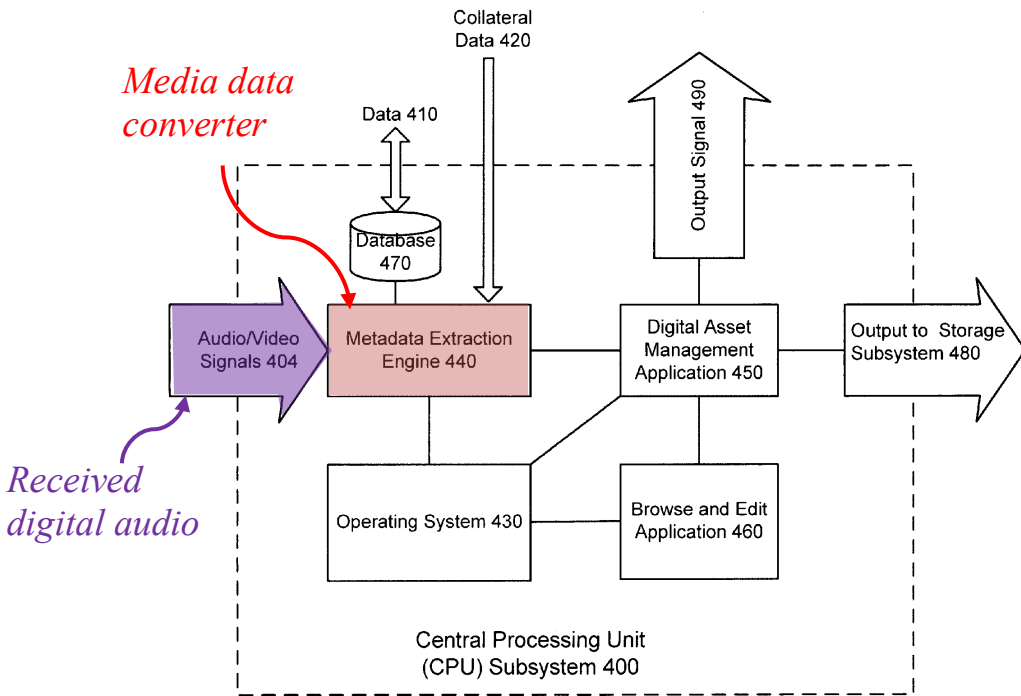


FIG. 4

Ex.1009, Fig. 4 (annotated); Ex.1003 ¶74.

**Second**, also in the combination, Manjunath describes metadata in the form of searchable text annotations (“a text based searchable file as a text context tag”). Ex.1010, 127. Manjunath explains that “[t]he practice of **using text annotations to search**, catalogue and index multimedia content is both longstanding and widespread.” Ex.1010, 127. And “MPEG-7 supports this practice with the TextAnnotation data type that allows free text, keyword, structured and dependency structure annotations.” Ex.1010, 127. Manjunath provides the following examples of an “**XML file**” for text annotations. Ex.1010, 104, 127-128:

```
<TextAnnotation>
  <FreeTextAnnotation xml:lang="en">
    Spain scores a goal against Sweden.
    The scoring player is Morientes.
  </FreeTextAnnotation>
</TextAnnotation>
```

**Ex.1010, 127 (XML file examples for free  
text annotations)**

```
<TextAnnotation>
  <KeywordAnnotation>
    <Keyword>score</Keyword>
    <Keyword>Sweden</Keyword>
    <Keyword>Spain</Keyword>
    <Keyword>Morientes</Keyword>
  </KeywordAnnotation>
</TextAnnotation>
```

**Ex.1010, 127 (XML file examples for  
keyword annotations)**

```
<TextAnnotation>
  <StructuredAnnotation>
    <Who><Name>Spain</Name></Who>
    <WhatAction><Name>score goal</Name></WhatAction>
    <Where><Name>A Coruña, Spain</Name></Where>
    <When><Name>March 25, 1998</Name></When>
  </StructuredAnnotation>
</TextAnnotation>
```

**Ex.1010, 128 (XML file examples for  
structured annotations)**

As previously discussed in §VIII.B.3, a POSITA would have found it obvious to organize Spatharis' extracted metadata tags as a text-based searchable file consistent with MPEG-7 standards as described in Manjunath. Ex.1003 ¶75.

Thus, because Spatharis describes a metadata extraction engine that performs speech-to-text conversion to create tags and annotations, which would have been searchable as taught by Manjunath, Spatharis and Manjunath render obvious “*a media data converter for converting the [] received digital audio to a text based searchable file as a text context tag.*” Ex.1003 ¶¶74-77.

**[6.10] *creating an image recognition searchable context tag with image recognition of at least a portion of the digital image and***

**First**, as discussed in [6.10], in the combination, Spatharis teaches that the CPU subsystem executes a “**metadata extraction engine**” (“*media data converter*”).

**Second**, the metadata extraction engine “provides analysis of the audio/video signal” including “**extracting information about the**” digital image “such as faces” by performing “**visual properties face recognition**” (“*image recognition*”) on the digital image. Ex.1009 ¶42. Spatharis further describes that the metadata extracted by the metadata extraction engine includes “**tags, annotations, and/or markings that can be associated with captured content.**” Ex.1009 ¶13. That captured content includes “sounds and/or images” and thus Spatharis teaches “*creating an image recognition...context tag with image recognition of at least a portion of the digital image.*” Ex.1009 ¶¶13, 42; Ex.1003 ¶79.

**Third**, Spatharis’ system creates an image recognition “*searchable*” context tag: “[m]etaddata can be used to aid a user and/or a computer system to, for example, **index/organize/classify**, retrieve and/or process captured content.” Ex.1009 ¶13. And as explained above at [6.10], Manjunath explains that “[t]he practice of **using text annotations to search**, catalogue and index multimedia content is both longstanding and widespread.” Ex.1010, 127. As previously discussed in §§VIII.B.3-4, a POSITA would have found it obvious to organize Spatharis’ extracted metadata tags as a text-based searchable file consistent with MPEG-7 standards as described in Manjunath. Ex.1003 ¶80.

Thus, because Spatharis describes extracting metadata such as faces from an image to create tags, which may be searchable as taught by Manjunath, Spatharis and Manjunath render obvious “*creating an image recognition searchable context tag with image recognition of at least a portion of the digital image.*” Ex.1003 ¶¶78-81.

**[6.11] *associating the text and image recognition context tags with the digital image, and***

**First**, as previously discussed in [6.10] and [6.11], in the combination, Spatharis describes generating metadata tags (“*text and image recognition context tags*”) by performing speech and image recognition on the audio and image data, respectively. Ex.1009 ¶¶13, 23, 29, 42.

**Second**, as described above at [6.8] and [6.9], Spatharis teaches obtaining location and time data (“*captured data*”).

Spatharis further describes that the generated metadata tags (“*the text and image recognition context tags*”) are “**permanently associated with the captured content**” (“*the digital image and the captured data*”). Ex.1009 ¶44; Ex.1003 ¶84.

Thus, because both the metadata extracted from the digital audio and video and the collateral metadata are permanently associated with captured content, Spatharis in the combination renders obvious “*associating the text and image recognition context tags with the digital image and the captured data.*” Ex.1003 ¶¶82-85.

**[6.12] *the internal storage storing the digital image in association with the text and image recognition context tags.***

In the combination, Spatharis describes that the “**storage subsystem 280**” (*the internal storage, see [6.2]*) stores “**content that has been enriched (e.g., processed and associated with metadata)** by the CPU subsystem 270.” Ex.1009 ¶26. Thus, Spatharis teaches *storing in the storage subsystem 280 (the internal storage) content including the digital image in association with the generated metadata (the text and image recognition context tags, see [6.10] to [6.12]).* Ex.1009 ¶26; *see also id.* ¶12 (“The captured sound and/or images can be referred to as content”); Ex.1003 ¶86.

## 5. Claim 7

**[7.1] *The system of claim 6, wherein the first data converter captures the first external audio information from the microphone during generation thereof.***

**First**, as described above at [6.3] and [6.4], Spatharis' CPU subsystem converts "*the first external audio information from the microphone*" to a digital form. Ex.1009 ¶¶23-24; Ex.1003 ¶87.

**Second**, this process is performed "*during generation thereof*" because Spatharis' CPU subsystem "can process and be programmed with software applications...that analyze, for example, images **in real-time** and extract information **about content that has been acquired.**" Ex.1009 ¶15; Ex.1003 ¶88.

Thus, because Spatharis converts and extracts audio data "in real-time," Spatharis in the combination renders obvious "*wherein the first data converter captures the first external audio information from the microphone during generation thereof.*" Ex.1009 ¶15; Ex.1003 ¶¶87-89.

## 6. Claim 8

**[8.1] *The system of claim 6, wherein the camera captures the image from the external image source at an instant in time.***

As previously discussed at [6.5], in the combination, Spatharis teaches that the capture subsystem includes a "**still camera**" that "acquires images" ("*capture[s] an image from an external image source*"). Ex.1009 ¶¶23-24.

Spatharis' "still camera" captures a still image representing the state of the external image source at a moment in time. Ex.1009 ¶¶23-24; Ex.1003 ¶90.

Accordingly, because Spatharis' system captures an image at an instant in time using a digital still/video camera, Spatharis renders obvious "*the camera captures the image from the external image source at an instant in time.*" Ex.1003 ¶¶90-91.

## 7. Claim 9

**[9.1] *The system of claim 6, wherein the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.***

As previously discussed in [6.4], Spatharis teaches that the "**audio** captured by the" source audio input microphone 260 ("*the captured external audio information*") "**can be analog signals that are converted into digital signals/data**" ("*process[ed]*") by the "**CPU subsystem 270**" ("*the first data converter*").

Ex.1009 ¶¶23-24; Ex.1003 ¶92. Spatharis further teaches that this processing takes place "in substantially real-time," such that the CPU subsystem begins processing *the captured external audio information* at the time when the capture begins ("*from a start event*") and stops processing at the time when the capture is complete ("*to a stop event*"). Ex.1009 ¶¶12, 15, 22, claim 7. It is also generally understood that

audio segments are not infinite in length—they have a start point and an end point.

Ex.1003 ¶92.

## 8. Claim 10

**[10.1] *The system of claim 6, and further including: a transmitter associated with the capture device for transmitting the associated stored digital image in association with the text and image recognition context tags to the location on the network***<sup>7</sup>;

In the combination, Spatharis teaches that the camera system produces an “output signal 232” that “can be **a processed signal that contains audio and/or video with associated metadata** (i.e., enriched data).” Ex.1009 ¶34. “The output signal 232 can be **transmitted**” to “**allow the downloading of the stored content and/or associated metadata into asset-management...archiving...or library systems.**” Ex.1009 ¶34. Transmitting stored content (“*the associated stored digital image in association with the text and image recognition context tags*”) to a library system makes use of “I/O ports” (“*transmitter*”). Ex.1009 ¶¶19, 34. “I/O ports 180 can be used to couple to outside systems that, for example, load programs, data, executable commands into the camera system 100 and/or its subsystems. I/O ports

---

<sup>7</sup> There is no antecedent basis for the term “*the location on the network.*” For purposes of this proceeding, Petitioner interprets this term as “*a location on a network.*” Ex.1003 ¶93.

180 can also be used to extract signals, content, and/or metadata from the camera system 100 to outside systems.” Ex.1009 ¶19; *see also id.* ¶34; Ex.1003 ¶93. These outside systems/libraries are at a “*location on the network*” at least because Spatharis explains that different components “can be connected using, for example, category-5 cable within an office complex,” thereby teaching that the components are connected to a network. Ex.1009 ¶22; Ex.1003 ¶93.

Thus, because Spatharis’ device includes I/O ports used to transmit data over a network to library archive systems, Spatharis in the combination renders obvious “*a transmitter associated with the capture device for transmitting the associated stored digital image in association with the text and image recognition context tags to the location on the network.*” Ex.1003 ¶¶93-94.

**[10.2] *a system disposed at the location on a network and including: a receiver for receiving the transmitted associated stored digital image in association with the text and image recognition context tags from the transmitter associated with the capture device as a received set of captured information, a database, and the database storing the received associated stored digital image in association with the text and image recognition context tags.***

**First**, as discussed in [10.1], Spatharis teaches that the “output signal 232 can be transmitted” to “**allow the downloading of the stored content and/or associated metadata into asset-management...archiving...or library systems**” (“*system disposed at the location on a network and including: a receiver for receiving the transmitted associated stored digital image in association with the*

*text and image recognition context tags from the transmitter*”). Ex.1009 ¶34; Ex.1003 ¶95. Systems that receive data from a network, such as the asset-management, archiving, and library systems of Spatharis, do so utilizing a “*receiver*” component to interface with the network. Ex.1003 ¶95; *see, e.g.*, Ex.1017, 9:5-12 (describing a “a personal computer...capable of receiving” data from a network through a “cable interface”); Ex.1016, 6:8-18 (describing the use of a “communication interface” for sending and receiving data from a network).

**Second**, Spatharis teaches storing the received data in a “*database*” because the “asset-management,” “archiving,” and “library” systems described in Spatharis store the received data in one or more databases, as such systems were well-known to do. Ex.1009 ¶34; Ex.1003 ¶96; *see, e.g.*, Ex.1014, Abstract, ¶23 (describing a system that “archives” content and extracted metadata in a database); Ex.1015, Abstract (describing an “archival” system that stores information in a database).

Thus, Spatharis teaches asset-management, archiving, and library systems that receive content and associated metadata from its camera system over a network for storage in a database, thereby in the combination rendering obvious “*a system disposed at the location on a network and including: a receiver for receiving the transmitted associated stored digital image in association with the text and image recognition context tags from the transmitter associated with the capture device as a received set of captured information, a database, and the*

*database storing the received associated stored digital image in association with the text and image recognition context tags.” Ex.1003 ¶97.*

## 9. Claim 11

**[11.1] *The system of claim 10, wherein the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.***

See [9.1]. Ex.1003 ¶98.

## 10. Claim 12

**[12.1] *The system of claim 11, wherein the transmitter transmits the associated stored digital image in association with the text and image recognition context tags to the location on the network after at least the stop event associated with the processing of the captured external audio information.***

**First**, as described above at [9.1], Spatharis’ camera system stores content after a “*stop event.*”

**Second**, as described above at [10.1], Spatharis explains that the camera system produces an “output signal 232” that “can be a **processed signal that contains audio and/or video with associated metadata** (i.e., enriched data).”

Ex.1009 ¶34. “The output signal 232 can be **transmitted**” to “**allow the downloading of the stored content and/or associated metadata into asset-management...archiving...or library systems.**” Ex.1009 ¶34; Ex.1003 ¶100.

The “stored content” (“*the associated stored digital image in association with the text and image recognition context tags*”) in Spatharis is not available for

transmission and “downloading” by the external systems until capture has completed (“*after at least the stop event associated with the processing of the captured external audio information*”). Ex.1009 ¶34; Ex.1003 ¶100.

Accordingly, because Spatharis describes transmitting the stored content to an archive or library system, Spatharis in the combination renders obvious “*wherein the transmitter transmits the associated stored digital image in association with the text and image recognition context tags to the location on the network after at least the stop event associated with the processing of the captured external audio information.*” Ex.1003 ¶¶99-101.

## **11. Claim 13**

**[13.0] *A system for capturing image and audio information for storage, comprising:***

*See* [6.0]. Ex.1003 ¶102.

**[13.1] *internal storage;***

*See* [6.2]. Ex.1003 ¶103.

**[13.2] *a microphone interfacable [sic] with an external audio information source that generates external audio information and***

*See* [6.3]. Ex.1003 ¶104.

**[13.3] *a first data converter for capturing the first external audio information from the microphone;***

*See* [6.4]. Ex.1003 ¶105.

**[13.4] *a camera interfacing with an image source to capture an image therefrom;***

*See* [6.5]. Ex.1003 ¶106.

**[13.5] *the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio within the capture device,***

*See* [6.6]. Ex.1003 ¶107.

**[13.6] *the camera for processing the captured image and storing it as a stored digital image;***

*See* [6.7]. Ex.1003 ¶108.

**[13.7] *a second data converter for converting the received digital audio to a text based searchable file as a text context tag and***

*See* [6.9]. Spatharis' metadata extraction engine teaches both a “*media data converter*” (recited in [6.9]) and a “*second data converter*” (recited in [13.7]).

Ex.1009 ¶¶13, 42; Ex.1003 ¶109.

**[13.8] *creating an image recognition searchable context tag with image recognition of at least a portion of the digital image and***

*See* [6.10]. Ex.1003 ¶110.

**[13.9] *associating the text and image recognition context tags with the digital image; and***

*See* [6.11]. Ex.1003 ¶111.

**[13.10] *the internal storage storing the digital image in association with the text and image recognition context tags.***

*See* [6.12]. Ex.1003 ¶112.

**12. Claim 14**

**[14.1]** *The system of claim 13, wherein the image source is an external image source.*

See [6.5]. Ex.1003 ¶113.

**13. Claim 15**

**[15.1]** *The system of claim 13, wherein the first data converter captures the first external audio information from the microphone during generation thereof.*

See [7.1]. Ex.1003 ¶114.

**14. Claim 16**

**[16.1]** *The system of claim 13, wherein the camera captures the image from the image source at an instant in time.*

See [8.1]. Ex.1003 ¶115.

**15. Claim 17**

**[17.1]** *The system of claim 13, wherein the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.*

See [9.1]. Ex.1003 ¶116.

**C. Ground 2: Claims 6-17 are obvious in view of Fuller and Jain**

**1. Fuller**

Fuller describes a “capture device” such as a “digital video camera 100” for capturing content such as “still images,” “[d]igital video frames,” and “audio.” Ex.1005, 1:25-32, 2:5-62, 7:30-51. The content is then “digitized” by an “analog-

to-digital (A/D) converter 203 and an A/D converter 204” included in the capture device. Ex.1005, 7:30-51. Fuller further describes that other “metadata” is captured in addition to the content, such as a “time/date” or “location” associated with the captured content. Ex.1005, 4:1-23. The content may also be associated with “metadata descriptions” derived from the content, such as “[f]ace identification/recognition” or “Optical Character Recognition (OCR)” of a still image or video. Ex.1005, 2:52-3:9. The captured content is stored in association with the metadata descriptions and the additional metadata. Ex.1005, 8:1-15; Ex.1003 ¶117.

## **2. Jain**

Fuller incorporates Jain (U.S. patent applications Ser. No. 09/134,500) by reference. Ex.1005, 7:50-68. Jain describes a “speech transcription module” which is an example implementation of the audio/video analysis engine of Fuller. Ex.1005, 7:50-68; Ex.1006, 9:33-10:27. This speech transcription module can use “speech recognition” to analyze the “digital audio signal” to produce content-based metadata such as a “full text” transcription of the audio. Ex.1006, 9:33-10:27. The content-based metadata is then stored in association with the captured content and the additional metadata as described in Fuller. Ex.1005, 8:1-15; Ex.1003 ¶118.

### 3. Reasons to combine Fuller and Jain

A POSITA would have found it obvious, and indeed would have been motivated, to implement Fuller's audio/video analysis engine 301 according to Jain's teachings because Fuller explicitly teaches to do so. Ex.1003 ¶119. As noted above, Fuller incorporates Jain by reference. Ex.1005, 7:50-68.

Fuller describes:

[An] audio/video analysis engine 301 which performs metadata extraction. In this example, the Virage audio and Video engines are offered as suitable examples for function 301, and are further described in U.S. patent and Ser. No. 09/134,497 [Jain], entitled "Video Cataloger System with Synchronized Encoders", which are hereby incorporated by reference.

Ex.1005, 7:51-59. Thus, Fuller's disclosure includes Jain's implementation details for the audio/video analysis engine 301. Because Fuller explicitly incorporates Jain by reference, a POSITA would have had a reasonable expectation of success combining the teachings of Fuller and Jain. Ex.1003 ¶120

Accordingly, the combination of Fuller and Jain is merely combining prior art elements (Fuller's audio/video analysis engine and Jain's speech transcription) according to known methods to yield predictable results. *KSR*, 550 U.S. at 416; Ex.1003 ¶¶119-121.

#### 4. Claim 6

**[6.0]** *A system for capturing image and audio information for storage, comprising:*

To the extent that the preamble is limiting, it is taught by Fuller.

**First**, Fuller describes digital video camera 100, which it refers to as a “digital capture system” (“*system for capturing*”), by which “[d]igital video frames are captured sequentially by a CCD sensor 201, while audio is captured by a microphone 202” (“[a] *system for capturing image and audio information*”).

Ex.1005, Abstract, 4:57-59; 7:30-51. In particular, Fuller describes a digital video camera that includes a “CCD sensor 201” to capture images and/or video and a “microphone 202” to capture audio. Ex.1005, 7:30-51.

**Second**, video and audio information captured by Fuller’s digital video camera is “stored in the frame buffer 205 and the sound buffer 206, respectively” and is thus “*for storage.*” Ex.1005, 7:30-51. Additionally, Fuller states that images and audio can be stored “in the form of an MPEG-7 stream” that “passes to an **internal storage unit 702**, which may be a digital tape, a hard disk, or other **storage media.**” Ex.1005, 8:1-15.

Thus, Fuller’s system for capturing image and audio, and storing that image and audio in at least the frame buffer, the sound buffer, and/or the internal storage renders obvious “[a] *system for capturing image and audio information for*

*storage*” as claimed. Ex.1003 ¶¶122-125

**[6.1] a capture device having:**

Fuller describes a digital video camera 100, which it refers to as a “digital capture device” (“*capture device*”), that is used for taking pictures, recording video, and/or recording audio. *See, e.g.*, Ex.1005, 5:9-12. Fuller’s digital video camera is a “*capture device*” because it captures media (e.g., images, video, and audio) as well as metadata (e.g., time and location). Fuller “relates generally to **digital capture devices**, and more particularly, to digital still cameras, digital video cameras, digital video encoders **and other media capture devices.**”

Ex.1005, 1:26-32; *see also id.*, 3:9-18; Ex.1003 ¶126. Fuller’s capture device is shown in Fig. 4 below.

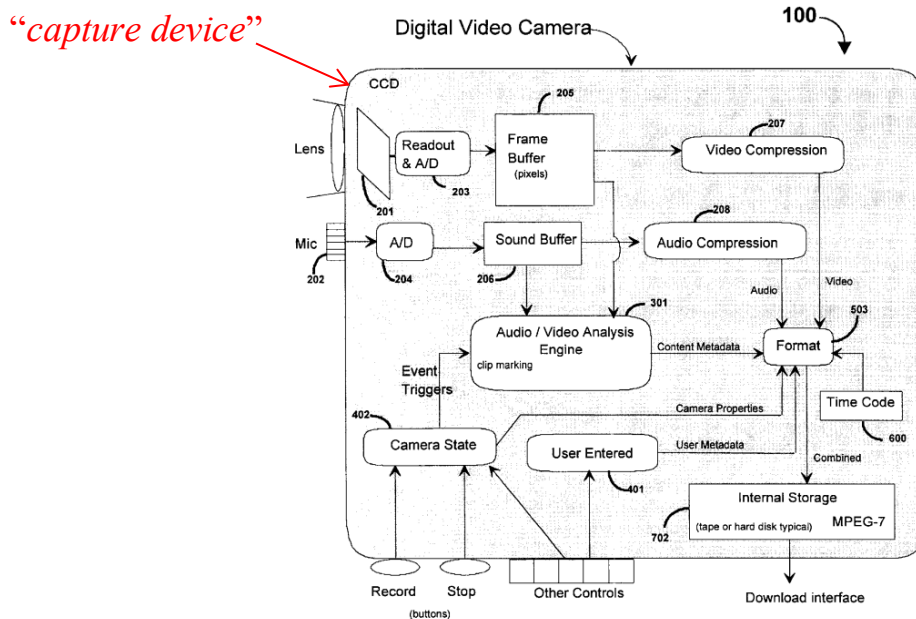


Figure 4

Ex.1005, FIG. 4 (annotated); Ex.1003 ¶126.

Thus, because Fuller describes capturing video and audio with a digital video camera, Fuller in the combination renders obvious “a capture device.” Ex.1003 ¶¶126-127.

**[6.2] internal storage;**

Fuller’s digital video camera includes “internal storage” because it has a frame buffer 205, sound buffer 206, and internal storage 700, 702, 703, 704. Ex.1005, 5:29-42, 7:30-51, 8:1-35, FIGS. 1, 2a, 3, 4, 5. The frame buffer 205 and the sound buffer 206 are “internal” storage because both are located internally within Fuller’s digital video camera, as illustrated below in Fig. 4. Ex.1005, 7:30-51.

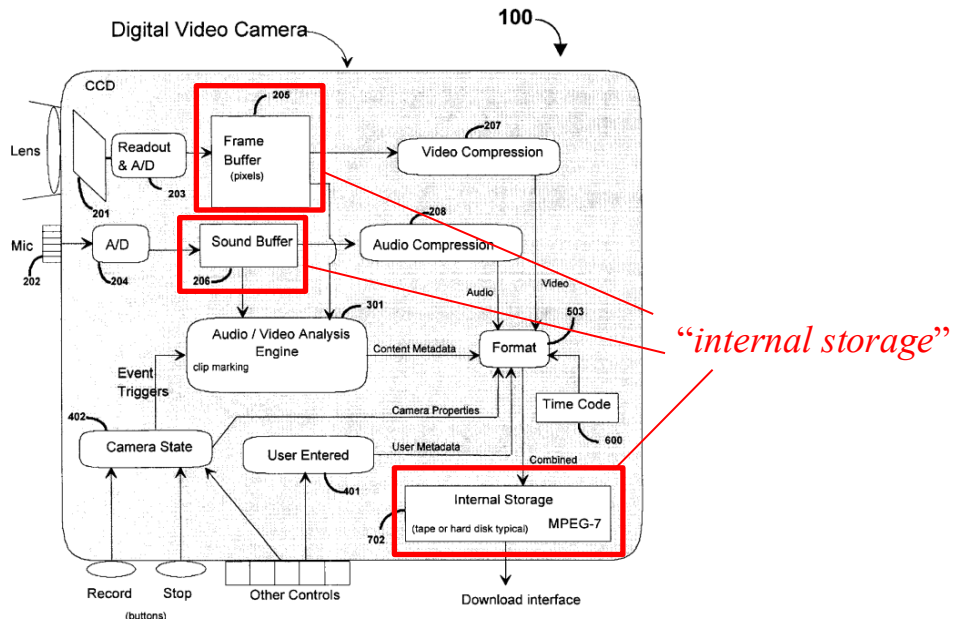


Figure 4

Ex.1005, FIG. 4 (annotated); Ex.1003 ¶128.

Thus, Fuller’s device includes a frame buffer, the sound buffer, and internal storage, which in the combination renders obvious “*internal storage.*” Ex.1003 ¶¶128-129.

**[6.3] a microphone interfacable [sic] with and [sic, an] external audio information source that generates external audio information and**

Fuller’s microphone 202 (“*microphone*”) captures audio information from the environment and is thus “*interfacable with [an] external audio information source that generates external audio information.*” Ex.1005, 7:30-51 (“audio is captured by a microphone 202”); *see also id.*, 5:29-42; Martin (Ex.1011), 7:62-66 (issued in 1952, and describing that the “sound of [an] instrument...is picked up by the microphone 25, and a corresponding electrical signal is obtained”); Weber

(Ex.1012), 5:70-73 (issued in 1961, and describing a “microphone...in use by a speaker” producing “[s]peech signals” that “are amplified by audio amplifier 14”); Ex.1003 ¶130. Fuller’s microphone 202 is shown, for example, in Fig. 4 below.

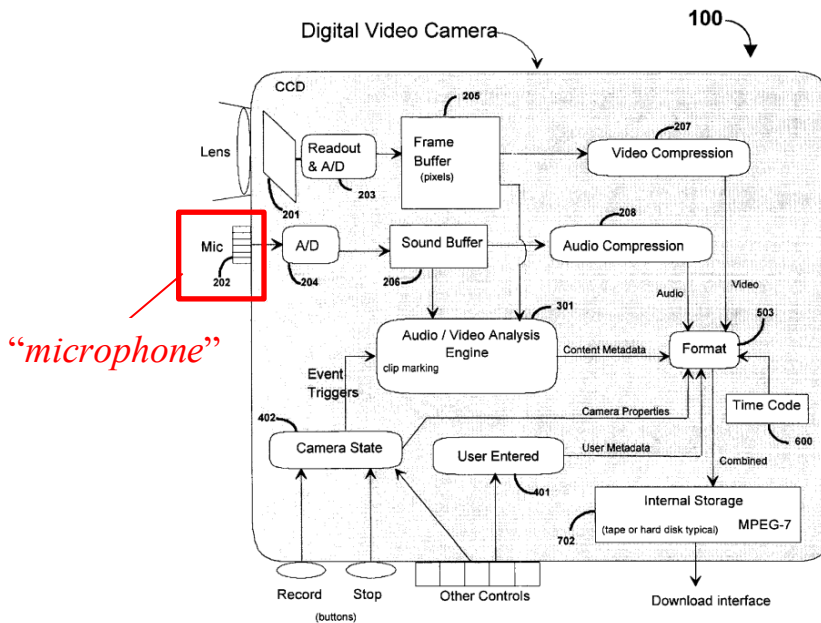


Figure 4

Ex.1005, FIG. 4 (annotated); Ex.1003 ¶130.

See also, e.g., Ex.1005, FIGS. 1, 5.

Thus, because Fuller’s microphone captures audio information from the environment, Fuller in the combination renders obvious “a microphone interfacable with [an] external audio information source that generates external audio information.” Ex.1003 ¶¶130-131.

**[6.4] a first data converter<sup>8</sup> for capturing the first external audio information from the microphone,**

Fuller's digital video camera 100 includes an analog-to-digital (A/D) converter 204 ("*first data converter*"), which receives "audio...captured by a microphone 202" and converts that analog audio to digital signals ("*for capturing the first external audio information from the microphone*"). Ex.1005, 7:30-51; see also *id.*, 5:30-36. Fuller's A/D converter 204 is a "*first data converter*" because the analog audio is "**digitized by...A/D converter 204,**" for example, "as a sequence of **8-bit or 16-bit waveform samples** at a suitable sampling frequency, such as 44.1 kHz (for CD quality audio)." Ex.1005, 7:30-51; Ex.1003 ¶132. Each time the A/D converter 204 generates a digital "waveform sample[]" from the analog audio signal, it is "*converting*" analog audio data ("*the first external audio information from the microphone*") into corresponding "digital form" data. Ex.1005, 7:30-51,

---

<sup>8</sup> Claim 6 appears to use the term "*first*" data converter to distinguish that data converter from the later-recited "*media data converter.*" Claim 6 and its dependents do not recite a "*second data converter.*" However, independent claim 13 recites a "*first data converter*" similar to that recited in claim 6, and a "*second data converter*" that is similar to the "*media data converter*" of claim 6. Ex.1003 ¶132.

5:31-35; Ex.1003 ¶132. Fuller's A/D converter 204 is shown in Fig. 4 below.

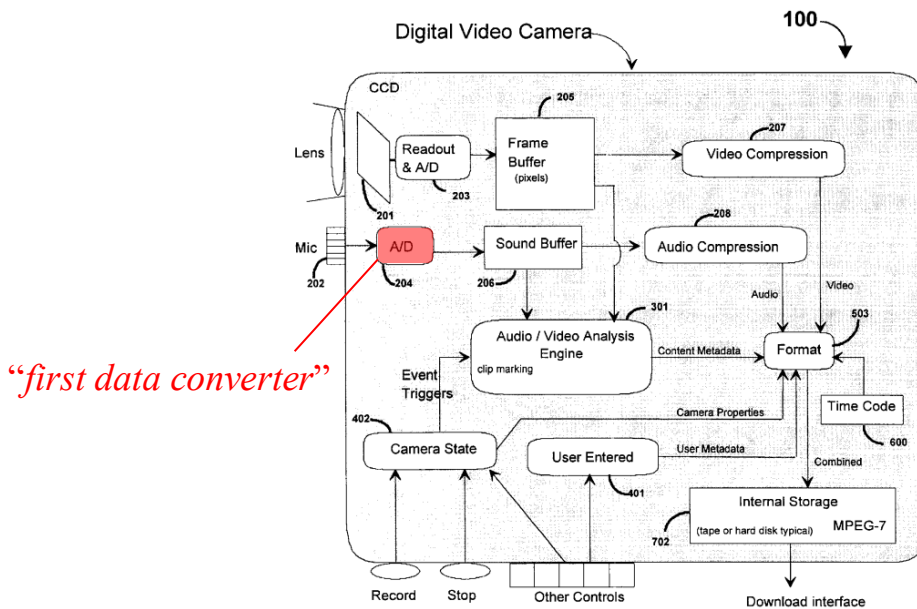


Figure 4

**Ex.1005, FIG. 4 (annotated); Ex.1003 ¶132.**

See also, e.g., Ex.1005, FIGS. 1, 5.

Thus, because Fuller's system includes an A/D converter that converts data in the form of an analog audio signal from a microphone to data in the form of a digital signal, Fuller renders obvious "a first data converter for capturing the first external audio information from the microphone." Ex.1003 ¶¶132-133.

**[6.5] a camera interfacing with and [sic] external image source to capture an image therefrom;**

**First**, Fuller teaches a "camera" in the form of a "a visual sensor 201 such as a CCD chip" and "analog-to-digital unit 203." Ex.1005, 5:29-42, 7:30-51.

**Second**, Fuller's camera takes pictures and records video ("interfacing with

[an] external image source to capture an image therefrom.”). Ex.1005, 5:29-42, 7:30-51. Fuller explains that “[d]igital video frames are captured sequentially by” the visual sensor and then “digitized by” the A/D unit. Ex.1005, 7:30-51; *see also id.*, 5:29-42. Fuller’s visual sensor 201 and A/D unit 203 are shown in Fig. 4 below.

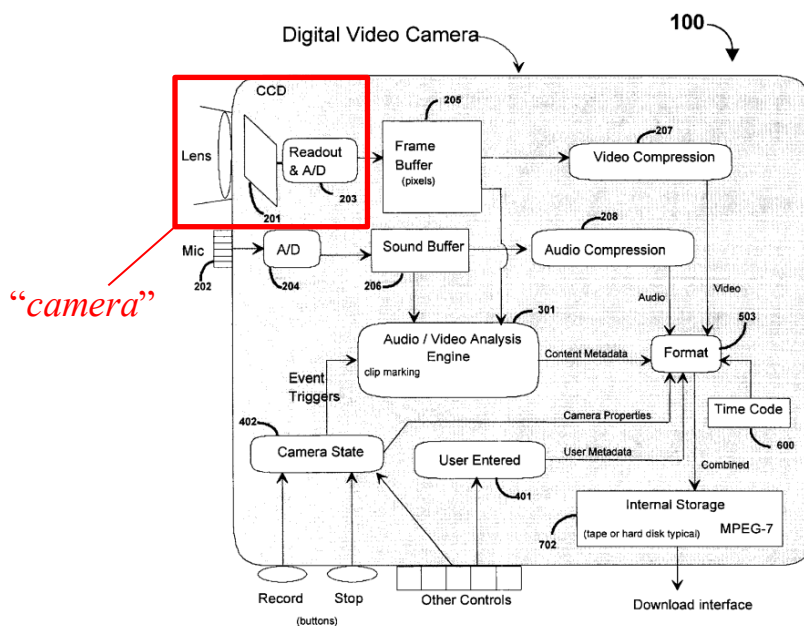


Figure 4

**Ex.1005, FIG. 4 (annotated); Ex.1003 ¶135.**

*See also, e.g., Ex.1005, FIGS. 1-3, 5.*

Thus, because Fuller’s system captures an image using a visual sensor, such as a CCD chip, and an A/D unit, Fuller in the combination renders obvious “a camera interfacing with [an] external image source to capture an image therefrom.” Ex.1003 ¶¶134-136.

**[6.6] *the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio in internal storage within the capture device,***

**First**, as discussed above at [6.4], Fuller’s A/D converter 204 (“*first data converter*”) converts audio data from a microphone 202 into “digital form.”

Ex.1005, 5:30-36, 7:30-51; Ex.1003 ¶137. The A/D converter 204 also

“*process[es]*” the digital form data by “*filter[ing]*” it: “the output of the sensor(s) is converted to digital form and **may be filtered** by an analog-to-digital unit 203 (visual), 204 (audio)” (“*processing the captured external audio information*”).

Ex.1005, 5:30-36, 7:30-51; Ex.1003 ¶137.

**Second**, Fuller’s analog-to-digital converter 204 (“*first data converter*”) then “*store[s]*” the “digital form” audio “in a memory unit...206,” which is also referred to as a “sound buffer 206” (“*the first data converter...storing it in a first digital audio format as stored digital audio*”). Ex.1005, 5:30-39 (the “digital content is then stored in a memory unit 205, 206”), 7:30-51; Ex.1003 ¶138.

The above quotations regarding storing “digital form” audio teach a “*digital audio format*” under the plain and ordinary meaning. *See* Ex.1018, 11-12 (finding that the term “*digital audio format*” should be given its plain and ordinary meaning). In the Samsung IPR, the Board agreed. *See Samsung Electronics Co. v. MyPort, Inc.*, IPR2023-00023, Paper 21 at 41 (Apr. 24, 2024) (“We understand Fuller’s disclosure of ‘digital form’ to be equivalent to a digital format.”).

Furthermore, it was well known, for example, that audio captured by a microphone, converted into digital audio data, and stored in a buffer is in a digital format (e.g., pulse code modulation (PCM) format) with a specific sampling rate and digital encoding scheme. *See, e.g.*, Ex.1019 ¶¶42 (“When sound is received in the microphone 11, an analog sound signal is generated and is A/D converted in the converter AD2 into PCM samples.”), 32 (referring to “PCM format.”); Ex.1020, 11:57-59 (“The PCM audio driver 214 converts the sound data stream from the standard data format into a PCM format for Waveform files.”); Ex.1003 ¶139.

Thus, the “digital form” referred to by Fuller is a “*digital audio format*” such as a PCM format. Ex.1005, 5:22-42, 7:30-51; Ex.1003 ¶140. Accordingly, Fuller renders obvious a “*first digital audio format*” as claimed. Ex.1003 ¶140.

**Third**, as shown, for example, in FIG. 4 below, Fuller’s sound buffer 206 is part of the capture subsystem and is thus “*internal storage within the capture device.*”

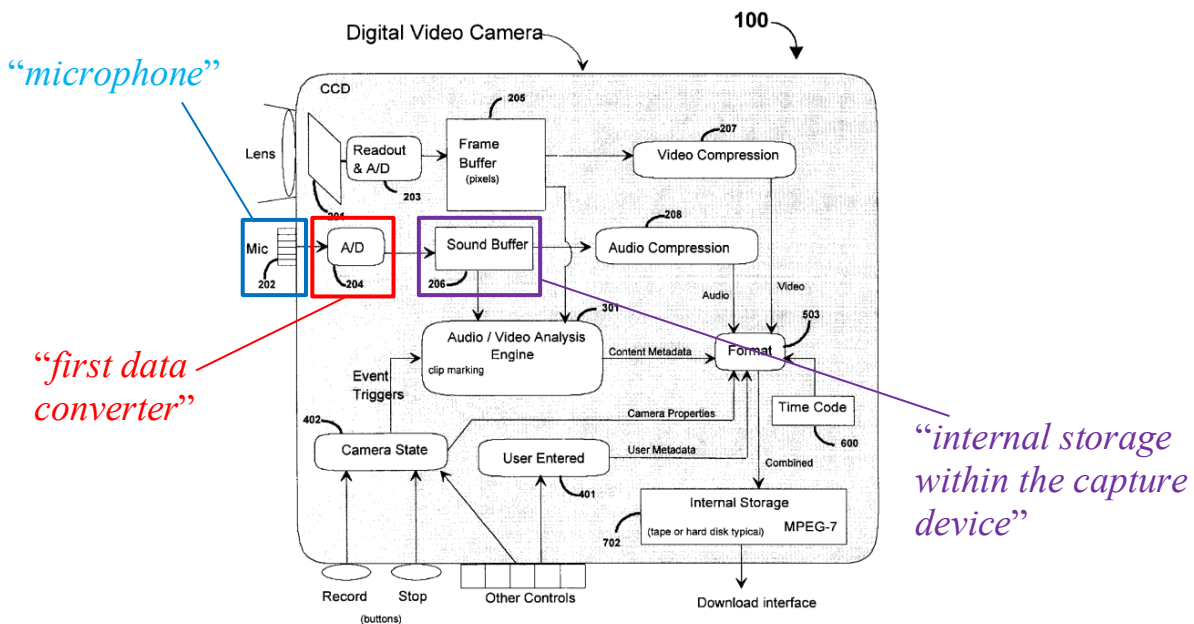


Figure 4

Ex.1005, FIG. 4 (annotated); Ex.1003 ¶141.

See also, e.g., Ex.1005, FIGS. 1, 5.

Thus, Fuller describes an analog-to-digital converter that converts audio data into a digital format as audio data for storage within a sound buffer, which in the combination renders obvious “*the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio in internal storage within the capture device.*” Ex.1003 ¶¶137-142.

**[6.7] the camera for processing the captured image and storing it as a stored digital image in internal storage and**

**First**, as discussed above at [6.5], Fuller’s device includes a visual sensor 201 and an A/D unit 203 for processing the captured image (“*the camera for processing the captured image*”).

**Second**, Fuller’s digital video camera 100 internally stores the digital image “in the frame buffer 205” such as “an RGB frame buffer 205” (“*the camera for...storing [the captured image] as a stored digital image in internal storage*”). Ex.1005, 7:30-51; Ex.1003 ¶144. As shown, for example, in FIG. 4 below, Fuller’s frame buffer 205 is part of the internal capture subsystem and is thus “*internal storage*.”

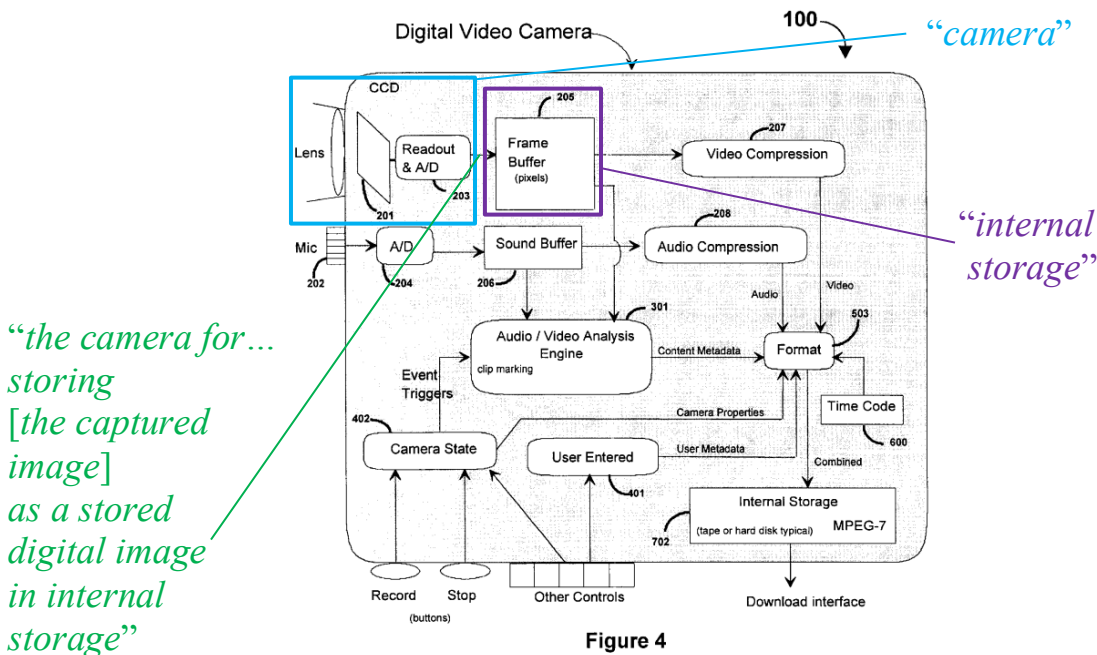


Figure 4

**Ex.1005, FIG. 4 (annotated); Ex.1003 ¶144.**

See also, e.g., Ex.1005, FIGS. 1-3 and 5.

Thus, because the visual sensor passes the captured information to the A/D unit for processing and storage as a digital image in the frame buffer, Fuller renders obvious “*the camera for processing the captured image and storing it as a*

*stored digital image in internal storage.”* Ex.1003 ¶145.

**[6.8] *a combiner for generating an association between the stored digital audio and the stored digital image,***

Fuller’s digital video camera 100 includes a formatting unit 503 (“*combiner*”) for generating an “[o]utput...in the form of an MPEG-7 stream, which functions as a data container that packages the compressed **audio/video** stream with the metadata” (emphasis added) (“*a combiner for generating an association between the stored digital audio and the stored digital image*”). Ex.1005, 8:1-15; Ex.1003 ¶146. Moreover, Fuller states that “[a]ll metadata and **audio/video** data may be formatted into a **combined** MPEG-7 container format” (emphasis added). Ex.1005, 5:2-3. As shown by annotated FIG. 4 below, the stored digital audio from the sound buffer 206 and the stored digital video from the frame buffer 205 are combined into a “[c]ombined” audio/video stream by the formatting unit 503. Ex.1005, 7:30-8:15.

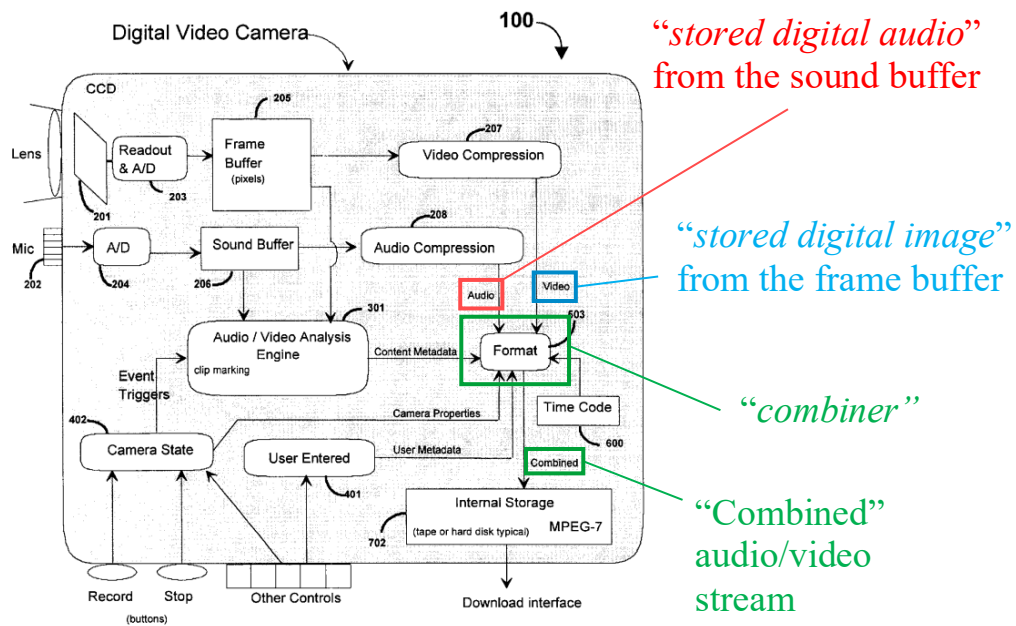


Figure 4

Ex.1005, FIG. 4 (annotated); Ex.1003 ¶146.

See also, e.g., Fuller FIG. 5.

Thus, because the formatting unit generates a combined audio/video stream from the stored digital audio from the sound buffer and the stored digital video from the frame buffer, Fuller renders obvious “a combiner for generating an association between the stored digital audio and the stored digital image.”

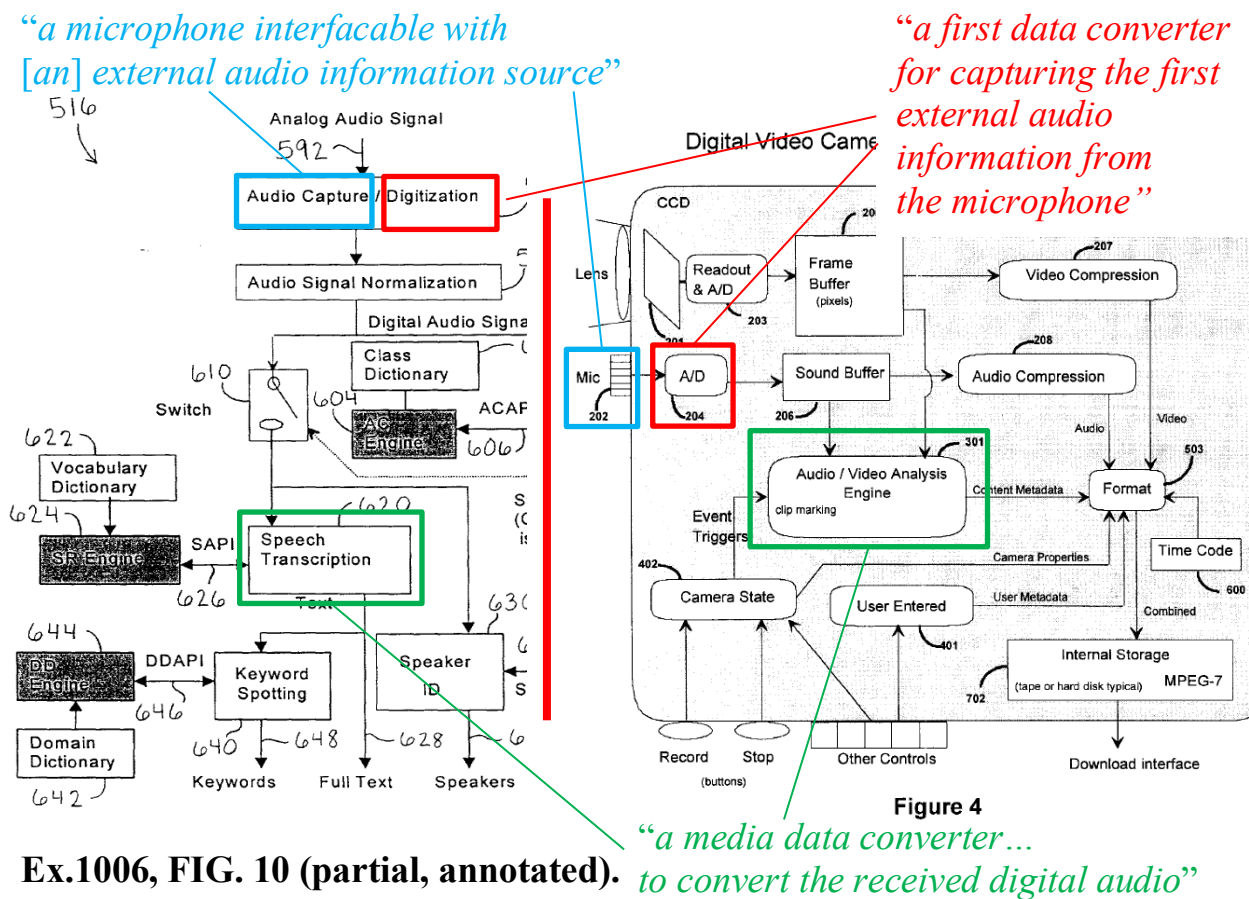
Ex.1003 ¶¶146-147.

**[6.9] a media data converter for converting the received set of captured information to convert the received digital audio to a text based searchable file as a text context tag and**

**First**, Fuller’s digital video camera 100 includes an audio/video analysis engine 301 (“a media data converter for converting the received set of captured information”) that converts media, such as audio and video signals, into extracted

metadata. Ex.1005, 7:30-8:15. While Fuller generally explains that the audio/video analysis engine 301 “performs metadata extraction,” Fuller cites to and incorporates Jain by reference for the implementation details. Ex.1005, 7:50-67.

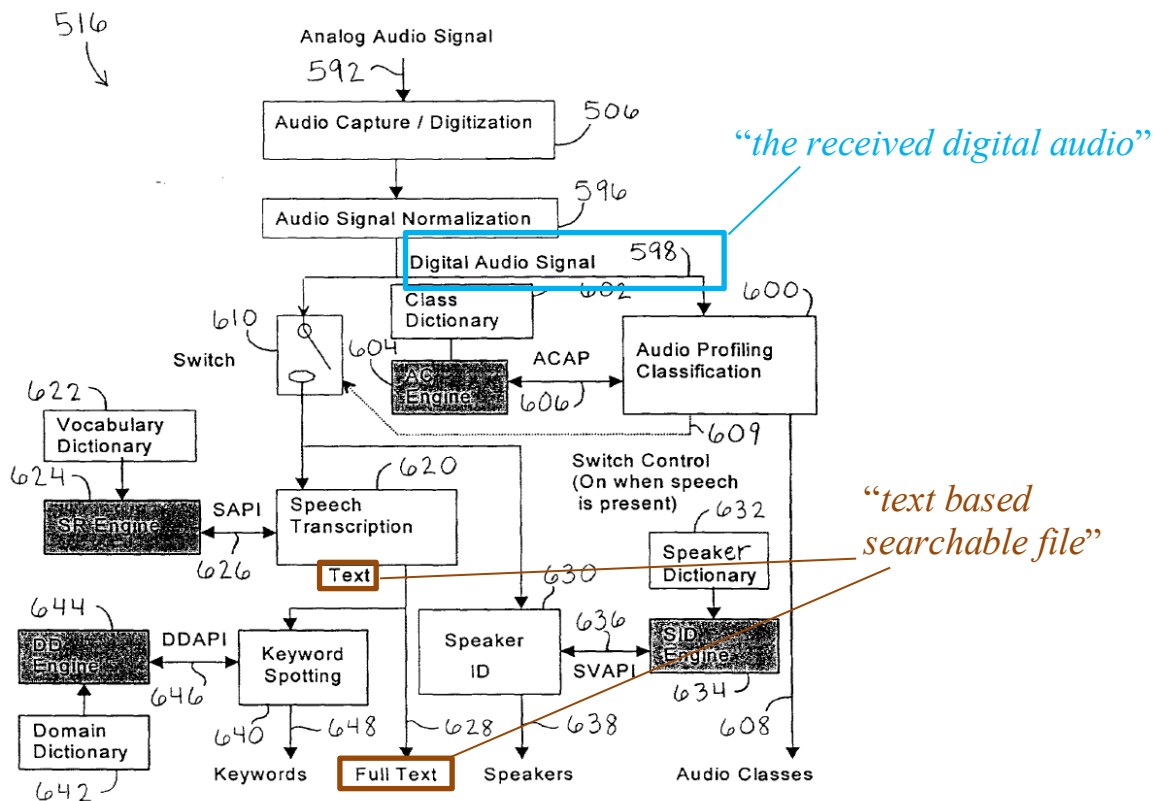
As shown by annotated FIG. 10 of Jain below (alongside annotated FIG. 4 of Fuller), Jain illustrates a speech transcription module 620 as an example implementation of performing metadata extraction. Ex.1003 ¶149.



Ex.1006, FIG. 10 (partial, annotated). Ex.1005, FIG. 4 (annotated); Ex.1003 ¶149.

The audio/video analysis engine 301—implemented according to Jain’s

incorporated teachings—includes a “speech transcription module 620” that uses “speech recognition” to convert the “digital audio signal 598” (“*received digital audio*”) into a “full text 628” (“*text based searchable file*”) transcription of the speech (“*a media data converter for converting the received set of captured information to convert the received digital audio to a text based searchable file as a text context tag*”). Ex.1006, 9:33-10:27; Ex.1003 ¶150.



**Ex.1006, FIG. 10 (annotated); Ex.1003 ¶150.**

The audio/video analysis engine 301—implemented according to Jain’s incorporated teachings—can output the full text 628 as a file including text (“*text based...file*”). Ex.1006, 9:33-10:27. The full text 628 is a “*file*” under the plain and

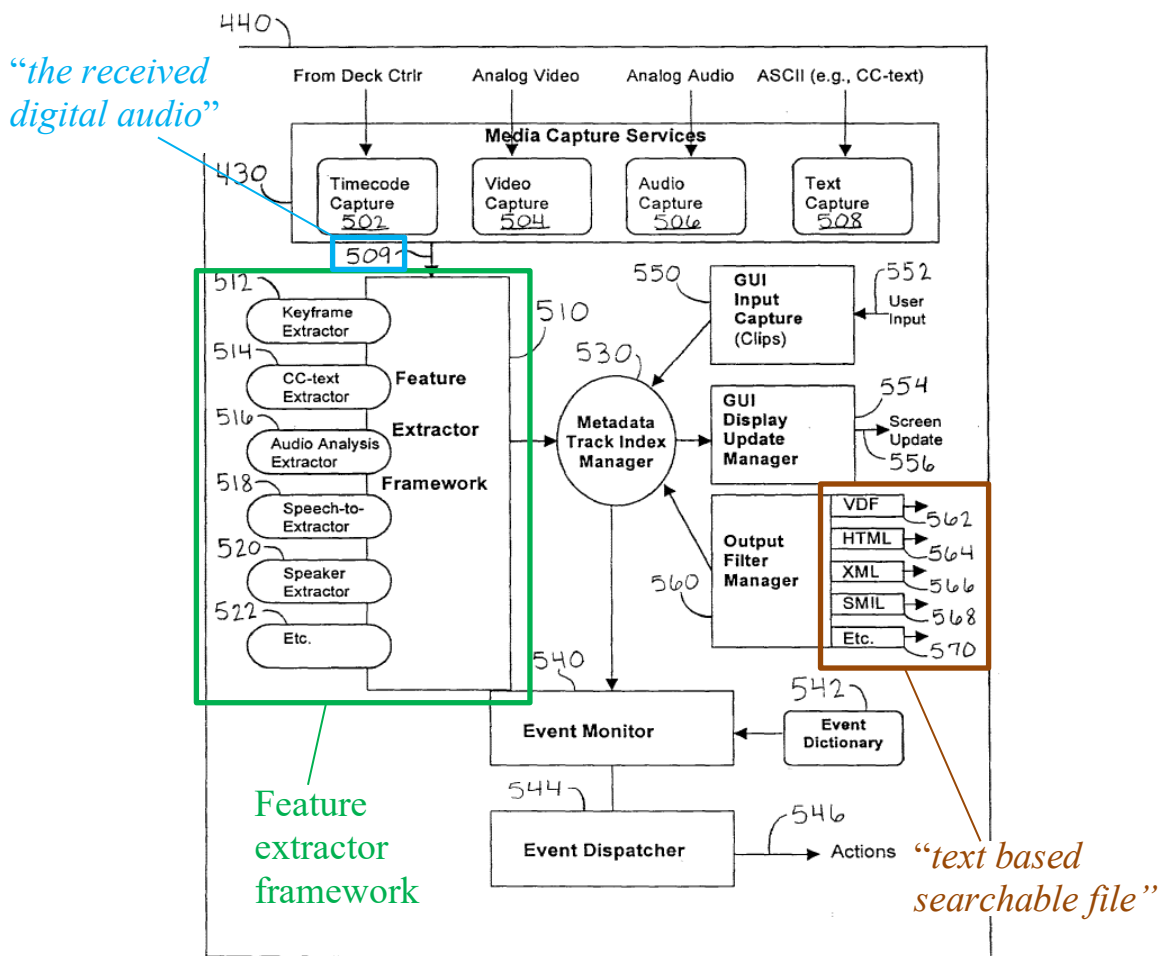
ordinary meaning of this term. *See e.g.*, IEEE Dictionary (Ex.1021), 432

(describing a file as simply a named “collection of data.”). Furthermore, Jain’s

extracted metadata may be full text 628 outputted “in a variety of formats such as

Virage Data Format (VDF) 562, HTML 564, XML 566, SMIL 568 and other 570.”

Ex.1006, 8:41-57, 9:33-10:27; Ex.1003 ¶151.

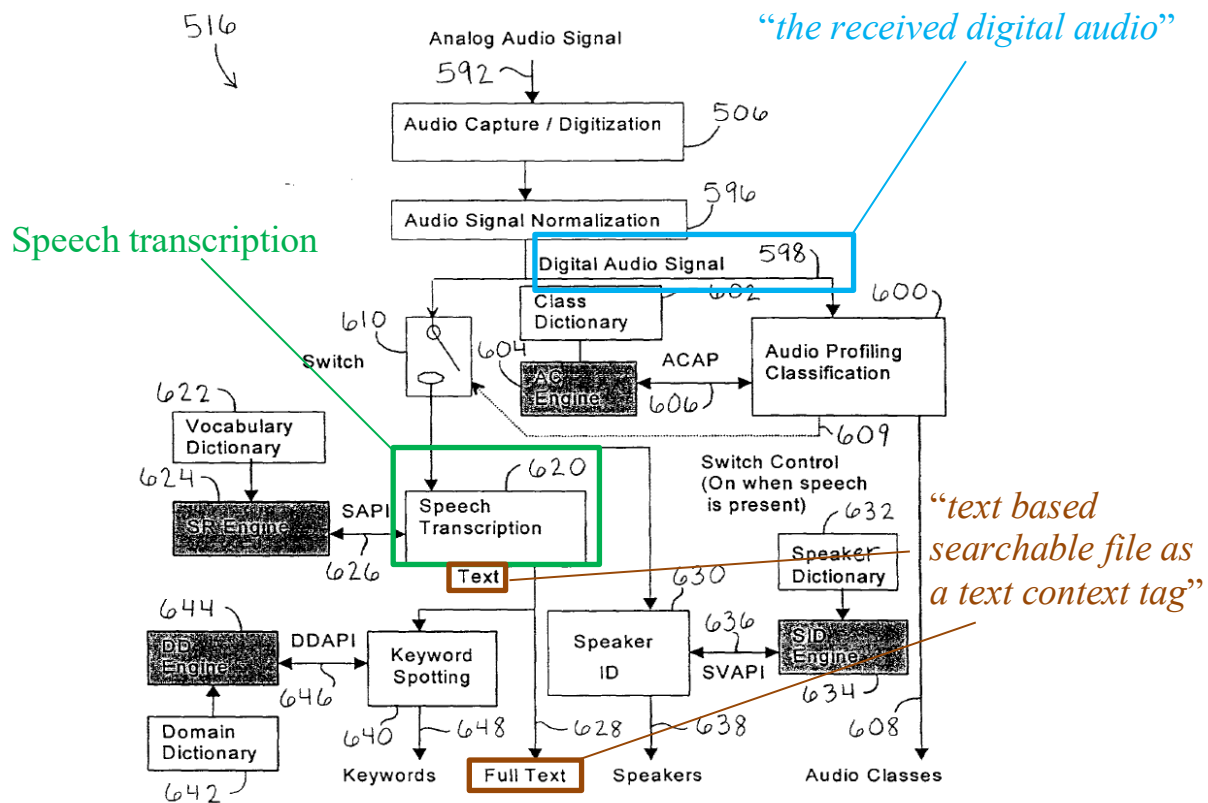


Ex.1006, FIG. 9 (annotated); Ex.1003 ¶151.

Additionally, the full text 628 is “searchable” because the full text 628 can be searched for “keywords of interest” using a “Keyword Spotting Feature

Extractor 640” to “produce a text output 648 limited to the keywords specified by a Domain Dictionary 642.” Ex.1006, 9:33-10:27; Ex.1003 ¶152.

**Second**, the full text 628 (“*text based searchable file*”) and/or the text output 648 are “*a text context tag*” because, as discussed above, they are metadata extracted from the “digital audio signal 598” (“*received digital audio*”) by Fuller’s audio/video analysis engine (“*a media data converter for converting the received set of captured information*”) and thus provide context to the digital audio signal. Ex.1006, 9:33-10:27; Ex.1005, 7:50-67; Ex.1003 ¶153. The full text 628 and/or the text output 648 provide context because they can be searched for “keywords of interest” using a “Keyword Spotting Feature Extractor 640” to “produce a text output 648 limited to the keywords specified by a Domain Dictionary 642.” Ex.1006, 9:33-10:27; Ex.1003 ¶153.



**Ex.1006, FIG. 10 (annotated); Ex.1003 ¶153.**

In this vein, during claim construction in prior litigation, the District Court construed “*context tag*” as “a searchable element derived from either a data element itself or from the context description element.” Ex.1018, 7-9. Regardless of whether this construction is adopted, Fuller and Jain teach that the text based searchable file is “*as a text context tag*” as claimed. As described above, the full text 628 is “a searchable element” and the text output 648 is “limited to the keywords specified by a Domain Dictionary 642” and is *searchable* for those keywords. Ex.1006, 9:33-10:27. Additionally, the full text 628 and text output 648

are both “derived from...a data element...” (e.g., “*the received digital audio*”) because the full text 628 and the text output 648 are both derived from a speech transcription of the digital audio signal 598. *See* Ex.1006, 9:33-10:27; Ex.1003 ¶154. Thus, even using the prior claim construction, the full text 628 and/or the text output 648 teach “*a text context tag.*” Ex.1003 ¶154.

Accordingly, because Fuller’s audio/video analysis engine extracts metadata from the audio as full text, which may be searched by the keyword spotting feature extractor to produce a text output as taught by Jain, Fuller and Jain render obvious “*a media data converter for converting the received set of captured information to convert the received digital audio to a text based searchable file as a text context tag.*” Ex.1003 ¶¶148-155.

**[6.10] *creating an image recognition searchable context tag with image recognition of at least a portion of the digital image and***

**First**, Fuller’s audio/video analysis engine 301 uses “[f]ace identification/recognition” and/or “Optical Character Recognition (OCR)” to analyze “multimedia content such as still images and video” (“*image recognition of at least a portion of the digital image*”). Ex.1005, 2:52-3:9; Ex.1003 ¶156.

**Second**, Fuller’s audio/video analysis engine 301 creates “*an image recognition searchable context tag*” in the form of “metadata descriptions” of the “[f]ace identification/recognition” and/or “OCR” of “multimedia content such as

still images and video.” (“*with image recognition of at least a portion of the digital image*”). Ex.1005, 2:52-3:9. Fuller goes on to explain that these “metadata descriptions...can be effectively used to **index** the content for downstream applications such as **search and browse**.” Ex.1005, 2:52-3:9; Ex.1003 ¶157.

Fuller’s “metadata description” teaches an “*image recognition searchable context tag*” under the District Court’s construction of “*context tag*.” See Ex.1018, 7-9. A “metadata description” is “a searchable element” because the metadata descriptions “can be effectively used to **index** the content for downstream applications such as **search and browse**.” Ex.1005, 2:52-3:9. Additionally, a metadata description is “derived from...a data element...” (e.g., “*the digital image*”) because the metadata descriptions are derived from “multimedia content such as still **images** and video.” Ex.1005, 2:52-3:9; Ex.1003 ¶158.

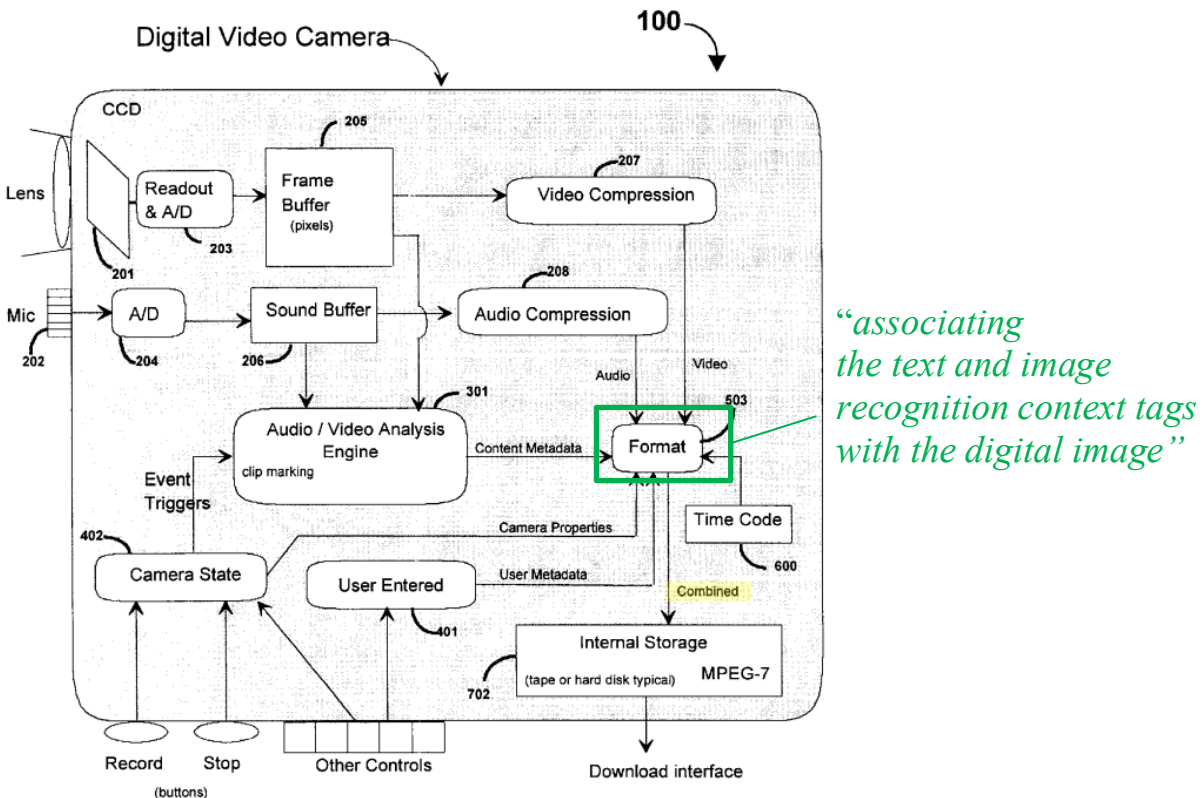
Thus, because Fuller’s audio/video analysis engine analyzes multimedia content, such as still images, to generate a metadata description, which can be indexed and searched, Fuller in the combination renders obvious “*creating an image recognition searchable context tag with image recognition of at least a portion of the digital image*.” Ex.1003 ¶¶156-159.

**[6.11] associating the text and image recognition context tags with the digital image, and**

**First**, as described above at [6.9] and [6.10], the audio/video analysis engine

301 may perform “metadata extraction” to obtain metadata (“*text and image recognition context tags*”) from the digital audio and video. Ex.1005, 7:50-67.

**Second**, Fuller’s device includes a formatting unit 503 that synchronizes (“*associate[es]*”) the “[c]ontent-based metadata from the analysis engine 301” (“*the text and image recognition context tags*”) “with the video content” (“*the digital image*”). Ex.1005, 7:50-8:15. The formatting unit 503 outputs the synchronized content and metadata “in the form of an MPEG-7 stream” (“*associating the text and image recognition context tags with the digital image*”). Ex.1005, 8:1-15; Ex.1003 ¶161. Fuller also discusses that the “combined metadata and image file” may also be stored as “a FlashPix formatted file” in the “storage unit 700.” Ex.1005, 6:65-7:10. Fig. 4 (annotated below) illustrates the “[c]ombined” MPEG-7 stream flowing out of the formatting unit 503 to the storage 702:



Ex.1005, FIG. 4 (annotated); Ex.1003 ¶161.

Thus, because the metadata extracted from the digital audio and video is sent to the formatting unit and synchronized with the video content, Fuller in the combination renders obvious “*associating the text and image recognition context tags with the digital image.*” Ex.1003 ¶¶160-162.

**[6.12] the internal storage storing the digital image in association with the text and image recognition context tags.**

First, as discussed above at [6.2], Fuller’s digital video camera 100 includes a frame buffer 205, sound buffer 206, and internal storage 700, 702, 703, 704—all of which are “*internal storage.*”

**Second**, as shown above with respect to [6.11], the formatting unit 503 synchronizes the content-based metadata with the video content “in the form of an MPEG-7 stream.” Ex.1005, 8:1-15.

**Third**, Fuller explains that the “MPEG-7 stream then passes to an internal storage unit 702, which may be a digital tape, a hard disk, or other storage media” (*“the internal storage storing the digital image in association with the text and image recognition context tags”*). Ex.1005, 8:1-15. Fuller also discusses that the “combined metadata and image file” may also be stored as “a FlashPix formatted file” in the “storage unit 700.” Ex.1005, 6:65-7:10; Ex.1003 ¶165.

Thus, because Fuller’s device stores the MPEG-7 stream (which includes the video content, content metadata, and collateral metadata) as a file in at least internal storage 702, Fuller in the combination renders obvious “*the internal storage storing the digital image in association with the text and image recognition context tags.*” Ex.1003 ¶¶163-166.

## 5. Claim 7

**[7.1] *The system of claim 6, wherein the first data converter captures the first external audio information from the microphone during generation thereof.***

**First**, as discussed above at [6.3] and [6.4], Fuller’s digital video camera 100 includes an A/D converter 204 that converts “*the first external audio information from the microphone*” to a digital form. Ex.1005, 7:30-51; *see also id.*,

5:29-42; Ex.1003 ¶166. For example, Fuller specifies that “audio is captured by a microphone 202” and that “[this] signal[ ] is digitized by...an A/D converter 204.” Ex.1005, 7:30-51.

**Second**, Fuller’s process converts captured content “*during generation thereof*” because it is configured to “automatically extract metadata in real-time from the digital content simultaneously with the capture of the content.” Ex.1005, 4:30-32; Ex.1003 ¶168.

Accordingly, because Fuller’s system includes an A/D converter that captures audio in the form of an analog audio signal from a microphone simultaneously with capturing the audio, Fuller renders obvious “*the first data converter captures the first external audio information from the microphone during generation thereof.*” Ex.1003 ¶¶167-169.

## **6. Claim 8**

**[8.1] *The system of claim 6, wherein the camera captures the image from the external image source at an instant in time.***

As discussed above at [6.5], Fuller’s digital video camera 100 includes a visual sensor 201 such as a CCD chip and an A/D unit 203 (“*camera*”) that captures an image (“*image from the external image source*”) at an instant in time, which would be necessary in order for the image to be a **still** image (“*the camera captures the image from the external image source at an instant in time*”).

Ex.1005, 5:29-42, 6:36-50, 7:10-51. Furthermore, Fuller notes that a video includes “video frames [which] are captured sequentially by a CCD sensor 201.”

Ex.1005, 7:32-34.

Accordingly, because Fuller’s system captures an image at an instant in time using a digital still/video camera which may include a visual sensor such as a CCD chip, Fuller in the combination renders obvious “*the camera captures the image from the external image source at an instant in time.*” Ex.1003 ¶¶170-171.

## 7. Claim 9

**[9.1] *The system of claim 6, wherein the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.***

Fuller’s digital video camera 100 includes an analog-to-digital (A/D) converter 204 (“*a first data converter*”) that processes audio (“*captured external audio information*”) from “the start...of [a] recording segment[ ]” (“*start event representing the time that capture of the audio is initiated*”) to the “stop of [a] recording segment[ ]” (“*a stop event representing the time that capture of the external audio information is complete*”). Ex.1005, 4:1-16; *see also id.*, 5:43-57, 7:50-68, 8:43-9:6, claims 4 and 17; Ex.1003 ¶172.

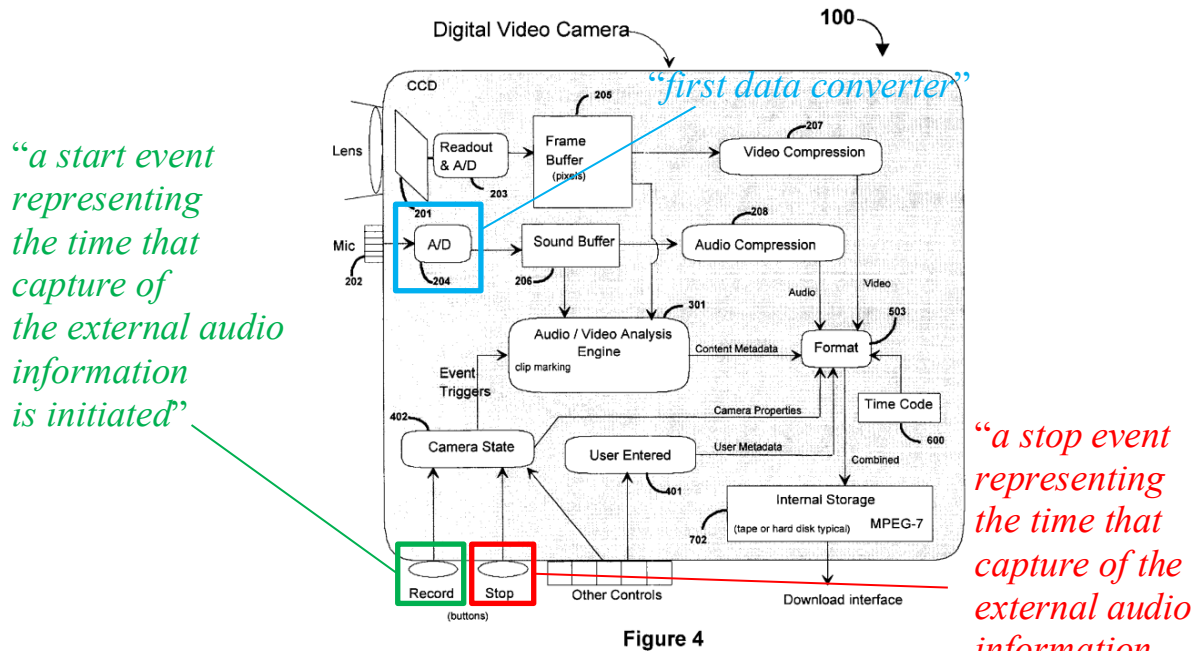


Figure 4

Ex.1005, FIG. 4 (annotated); Ex.1003 ¶172.

See also, Ex.1005, FIG. 5.

Thus, because Fuller’s digital video camera includes an A/D converter that processes audio from the start of a recording segment to the stop of a recording segment, Fuller renders obvious “the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.” Ex.1003 ¶¶172-73.

## 8. Claim 10

**[10.1] *The system of claim 6, and further including: a transmitter associated with the capture device for transmitting the associated stored digital image in association with the text and image recognition context tags to the location on the network***<sup>9</sup>;

Fuller's digital video camera 100 includes a transmitter associated with the capture device for transmitting the "MPEG-7 stream" which, as discussed above at [6.11], includes the metadata extracted from the digital audio and video and the video content ("*the associated stored digital image in association with the text and image recognition context tags*"), to "a host computer or other media processing device" ("*a transmitter associated with the capture device for transmitting the associated stored digital image in association with the text and image recognition context tags to the location on the network*"). Ex.1005, 8:1-15; Ex.1003 ¶174.

Fuller describes "a host computer or other media processing device" which is "*the location on the network*" because the "host computer or other media processing device" can perform "media asset management" using, for example, "WebWare." Ex.1005, 1:45-65, 8:1-15; Ex.1003 ¶174.

---

<sup>9</sup> There is no antecedent basis for the term "*the location on the network.*" For purposes of this proceeding, Petitioner interprets this term as "*a location on a network.*" Ex.1003 ¶174.

Thus, because Fuller’s digital video camera transmits the composite data set to a host computer or other media processing device over a network, Fuller in the combination renders obvious “*a transmitter associated with the capture device for transmitting the associated stored digital image in association with the text and image recognition context tags to the location on the network.*” Ex.1003 ¶¶174-175.

**[10.2] *a system disposed at the location on a network and including: a receiver for receiving the transmitted associated stored digital image in association with the text and image recognition context tags from the transmitter associated with the capture device as a received set of captured information, a database, and the database storing the received associated stored digital image in association with the text and image recognition context tags.***

**First**, Fuller describes that the “host computer or other media processing device” can include a “[d]igital media asset management system[ ] (DMMS[ ])” (“*a system disposed at the location on a network*”) which operates to receive with a receiver the “MPEG-7 stream” which, as discussed above at [6.11], includes metadata extracted from the digital audio and video and the video content (“*transmitted associated stored digital image in association with the text and image recognition context tags from the transmitter associated with the capture device*”), as a received set of captured information (“*a system disposed at the location on a network and including: a receiver for receiving the transmitted associated stored digital image in association with the text and image recognition*

*context tags from the transmitter associated with the capture device as a received set of captured information*"). Ex.1005, 1:45-65, 8:1-15; Ex.1003 ¶176.

**Second**, Fuller states that “[t]he use of various forms of metadata (data about the digital content) has emerged as a way to organize the digital content in **databases** and other storage means such that a specific piece of content may be easily found and used.” Ex.1005, 1:33-65. Fuller describes an example of this concept of a searchable database: “the digital content is entered into the DMMS” which can “exploit metadata to allow constrained searches for specific digital content” (“*a database, and the database storing the received associated stored digital image in association with the text and image recognition context tags*”) Ex.1005, 1:45-65; Ex.1003 ¶177.

Thus, because Fuller’s host computer receives the video content in association with the content-based metadata from a network and stores it in the DMMS (which Fuller explains may be implemented as a “database”), Fuller in the combination renders obvious “*a system disposed at the location on a network and including: a receiver for receiving the transmitted associated stored digital image in association with the text and image recognition context tags from the transmitter associated with the capture device as a received set of captured information, a database, and the database storing the received associated stored digital image in association with the text and image recognition context tags.*” Ex.1003 ¶¶176-178.

## 9. Claim 11

**[11.1] *The system of claim 10, wherein the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.***

As discussed above at [9.1], because Fuller’s digital video camera 100 includes an A/D converter that processes audio from the “start” of a recording segment to the “stop” of a recording segment, Fuller renders obvious “*the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.*” Ex.1003 ¶179.

## 10. Claim 12

**[12.1] *The system of claim 11, wherein the transmitter transmits the associated stored digital image in association with the text and image recognition context tags to the location on the network after at least the stop event associated with the processing of the captured external audio information.***

**First**, as discussed above at [10.1], Fuller’s “host computer or other media processing device” is “*the location on the network.*”

**Second**, as discussed above at [9.1], Fuller describes that audio and video may be combined to form “audiovisual content.” Ex.1005, 1:46-2:3. Because Fuller’s digital video camera 100 is configured such that “time codes from the time code generator 600 are applied to synchronize the metadata with the video content,

” the device then “[o]utput[s]” “the compressed audio/video stream with the metadata” “in the form of an MPEG-7 stream” and, in turn, the “[t]he storage unit 702 may then download the MPEG-7 data to a host computer or other media processing device.” Ex.1005, 8:1-15; Ex.1003 ¶181. Fuller thus renders obvious “*the transmitter transmits the associated stored digital image in association with the text and image recognition context tags to the location on the network after at least the stop event associated with the processing of the captured external audio information.*” Ex.1003 ¶¶180-181.

## 11. Claim 13

**[13.0] *A system for capturing image and audio information for storage, comprising:***

*See* [6.0]. Ex.1003 ¶182.

**[13.1] *internal storage;***

*See* [6.2]. Ex.1003 ¶183.

**[13.2] *a microphone interfacable [sic] with an external audio information source that generates external audio information and***

*See* [6.3]. Ex.1003 ¶184.

**[13.3] *a first data converter for capturing the first external audio information from the microphone;***

*See* [6.4]. Ex.1003 ¶185.

**[13.4] *a camera interfacing with an image source to capture an image therefrom;***

See [6.5]. Ex.1003 ¶186.

**[13.5] *the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio within the capture device,***

See [6.6]. Ex.1003 ¶187.

**[13.6] *the camera for processing the captured image and storing it as a stored digital image;***

See [6.7]. Ex.1003 ¶188.

**[13.7] *a second data converter for converting the received digital audio to a text based searchable file as a text context tag and***

See [6.9]. Fuller's audio/video analysis engine 301 teaches both a "media data converter" (recited in [6.10]) and a "second data converter" (recited in [13.9]). Ex.1005, 7:31-8:15; Ex.1003 ¶189.

**[13.8] *creating an image recognition searchable context tag with image recognition of at least a portion of the digital image and***

See [6.10]. Ex.1003 ¶190.

**[13.9] *associating the text and image recognition context tags with the digital image; and***

See [6.11]. Ex.1003 ¶191.

**[13.10] *the internal storage storing the digital image in association with the text and image recognition context tags.***

See [6.12]. Ex.1003 ¶192.

**12. Claim 14**

**[14.1] *The system of claim 13, wherein the image source is an external image source.***

*See* [6.5]. Ex.1003 ¶193.

**13. Claim 15**

**[15.1] *The system of claim 13, wherein the first data converter captures the first external audio information from the microphone during generation thereof.***

*See* [7.1]. Ex.1003 ¶194.

**14. Claim 16**

**[16.1] *The system of claim 13, wherein the camera captures the image from the image source at an instant in time.***

*See* [8.1]. Ex.1003 ¶195.

**15. Claim 17**

**[17.1] *The system of claim 13, wherein the first data converter processes the captured external audio information from a start event representing the time that capture of the external audio information is initiated to a stop event representing the time that capture of the external audio information is complete.***

*See* [9.1]. Ex.1003 ¶196.

**IX. CONCLUSION**

For at least the reasons described in the present Petition, the Challenged Claims are unpatentable.

Respectfully submitted,

Dated: September 9, 2025  
HAYNES AND BOONE, LLP  
2801 N. Harwood Ave, Ste 2300  
Dallas, Texas 75201  
Customer No. 27683

/Scott T. Jarratt/  
Scott T. Jarratt  
Lead Counsel for Petitioner  
Registration No. 70,297

**X. MANDATORY NOTICES**

**A. Real Party-in-Interest**

Pursuant to 37 C.F.R. §42.8(b)(1), Petitioner certifies that the real party-in-interest is Apple Inc.

**B. Related Matters**

Pursuant to 37 C.F.R. § 42.8(b)(2), to the best knowledge of the Petitioner, the '017 patent is involved in the following cases:

<b>Case Heading</b>	<b>Number</b>	<b>Court</b>	<b>Date</b>
<i>MyPort Technologies, Inc. v. Apple, Inc.</i>	1-24-cv-01337	DDE	Dec 6, 2024
<i>Samsung Electronics Co., Ltd. v. MyPort, Inc.</i>	IPR2023-00023	PTAB	Oct 7, 2022
<i>MyPort, Inc. v. Samsung Electronics Co., Ltd. et al</i>	2-22-cv-00114	EDTX	Apr 15, 2022

**C. Lead and Back-up Counsel and Service Information**

Lead Counsel

Scott T. Jarratt  
HAYNES AND BOONE, LLP  
2801 N. Harwood Ave, Ste 2300  
Dallas, TX 75201

Phone: (972) 739-8663  
Fax: (214) 200-0853  
scott.jarratt.ipr@haynesboone.com  
USPTO Reg. No. 70,297

Back-up Counsel

Calmann J. Clements  
HAYNES AND BOONE, LLP  
2801 N. Harwood St., Suite 2300  
Dallas, TX 75201

Phone: (972) 739-8638  
Fax: (214) 200-0853  
calmann.clements.ipr@haynesboone.com  
USPTO Reg. No. 66,910

Dan Smith  
HAYNES AND BOONE, LLP  
2801 N. Harwood Ave, Ste 2300  
Dallas, TX 75201

Phone: (972) 739-8634  
Fax: (214) 200-0853  
dan.smith.ipr@haynesboone.com  
USPTO Reg. No. 71,278

Dirk Bernhardt  
HAYNES AND BOONE, LLP  
2801 N. Harwood Ave, Ste 2300  
Dallas, TX 75201

Phone: (972) 739-8659  
Fax: (214) 200-0853  
dirk.bernhardt.ipr@haynesboone.com  
USPTO Reg. No. 82,329

Please address all correspondence to lead and back-up counsel. Petitioner consents to service in this proceeding by email at the addresses above.

**CERTIFICATE OF WORD COUNT**

Pursuant to 37 C.F.R. § 42.24(d), Petitioner hereby certifies, in accordance with and reliance on the word count provided by the word-processing system used to prepare this Petition, that the number of words in this paper is 12,866. Pursuant to 37 C.F.R. § 42.24(d), this word count excludes the table of contents, table of authorities, mandatory notices under § 42.8, certificate of service, certificate of word count, appendix of exhibits, and any claim listing.

Dated: September 9, 2025

/Scott T. Jarratt/  
Scott T. Jarratt  
Lead Counsel for Petitioner  
Registration No. 70,297

**CERTIFICATE OF SERVICE**

The undersigned certifies that, in accordance with 37 C.F.R. § 42.6(e) and 37 C.F.R. § 42.105, service was made on Patent Owner as detailed below.

*Date of service* September 9, 2025

*Manner of service* Priority Mail Express

*Documents served* Petition for *Inter Partes* Review Under 35 U.S.C. § 312 and 37 C.F.R. § 42.104 of U.S. 9,832,017;  
Petitioner's Exhibit List;  
Exhibits Ex.1001-1007 and Ex.1009-1026;  
Petitioner's Power of Attorney.

*Persons served* Law Office of Bill Naifeh  
P.O. Box 803423  
Dallas, TX 75380

/ Scott T. Jarratt/  
Scott T. Jarratt  
Lead Counsel for Petitioner  
Registration No. 70,297