UNITED STATES PATENT AND TRADEMARK OFFICE

_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD

_____

APPLE INC.,
Petitioner

_____

U.S. Patent No. 10,721,066

DECLARATION OF NATHANIEL POLISH, PH.D., UNDER 37 C.F.R. § 1.68 IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW

TABLE OF CONTENTS

I. INTRODUCTION 4
II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE 6
III. LEVEL OF ORDINARY SKILL IN THE ART 10
IV. RELEVANT LEGAL STANDARDS 10
V. SUMMARY OF THE '066 PATENT 11
A. Overview of the '066 patent 11
B. Effective Filing Date of the '066 patent 12
VI. CLAIM CONSTRUCTION 12
VII. THE CHALLENGED CLAIMS ARE UNPATENTABLE 13
A. Ground 1: Claims 6-17 are obvious in view of Spatharis and Manjunath 14
1. Spatharis 14
2. Manjunath 15
3. Reasons to combine Spatharis and Manjunath 16
4. Claim 6. 19
5. Claim 7 41
6. Claim 8 41
7. Claim 9 42
8. Claim 10 43
9. Claim 11 46
10. Claim 12 46
11. Claim 13 47
12. Claim 14 49
13. Claim 15 49
14. Claim 16 49
15. Claim 17 49
B. Ground 2: Claims 6-17 are obvious in view of Fuller and Jain 49

1. Fuller 49
2. Jain 50
3. Reasons to combine Fuller and Jain 51
4. Claim 6 51
5. Claim 7 76
6. Claim 8 77
7. Claim 9 78
8. Claim 10 80
9. Claim 11 82
10. Claim 12 82
11. Claim 13 83
12. Claim 14 85
13. Claim 15 85
14. Claim 16 85
15. Claim 17 86

VIII. CONCLUSION 87

I. INTRODUCTION

1. I am making this declaration at the request of Apple Inc. in the matter of the *Inter Partes* Review of U.S. Patent No. 10,721,066 ("the '066 patent").

2. I am being compensated for my work in this matter at my standard hourly rate. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this proceeding. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

3. I have been asked to provide my opinions regarding whether the subject matter of claims 6-17 ("the Challenged Claims") of the '066 patent would have been obvious to a person having ordinary skill in the art ("POSITA") at the time of the alleged invention, in light of the prior art. It is my opinion that the Challenged Claims would have been obvious to a POSITA.

4. In preparing this declaration, I have reviewed the following materials as well as any additional references cited herein:

- Ex.1001: U.S. Patent No. 10,721,066
- Ex.1002: Prosecution History of U.S. Patent No. 10,721,066
- Ex.1005: U.S. Patent No. 6,833,865 ("Fuller")
- Ex.1006: U.S. Patent No. 7,295,752 ("Jain")
- Ex.1007: Graphics File Formats: Second Edition, James D. Murray et al. (1996) ("Murray")

- Ex.1009: U.S. Pub. No. 2006/0227995 ("Spatharis")
- Ex.1010: Introduction to MPEG-7: Multimedia Content Description Interface, 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, *MyPort, Inc. v. Samsung Electronics Co., Ltd. et al.*, 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: The Authoritative Dictionary of IEEE Standards Terms (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")

5. In forming the opinions expressed below, I have considered: the documents listed above; the relevant legal standards, including the standard for obviousness; and my own knowledge and experience based upon my work in the field as described below.

6. Unless otherwise noted, all **emphasis** in any quoted material has been added. Claim terms are *italicized*.

II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

7. My complete qualifications and professional experience are described in my *Curriculum Vitae*, a copy of which can be found in Exhibit 1004. The following is a brief summary of my relevant qualifications and professional experience.

8. I am an independent consultant. All of my opinions stated in this declaration are based on my own personal knowledge and professional judgment. In forming my opinions, I have relied on my knowledge and experience in acoustics, signal processing, design of headsets and sound production equipment, and representations of acoustic sounds.

9. I am over 18 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein. The following provides an overview of some of my experience that is relevant to the matters set forth in this declaration.

10. I expect to testify regarding my background, qualifications, and experience relevant to the issues in this litigation. I have a Ph.D. in Computer Science from Columbia University. I hold the following four degrees from Columbia, spanning the years 1980 to 1993:

- Ph.D. in Computer Science, May 1993, Thesis: Mixed Distance Measures for the Optimization of Concatenative Vocabularies in Speech Synthesis;
- M.Phil. in Computer Science, December 1989;
- M.S. in Computer Science, December 1987; and
- B.A. in Physics, Columbia College, May 1984.

11. For over thirty years, I have run a computer technology development firm that I co-founded, called Daedalus Technology Group. My primary business activity is the development of computer-related products. This activity involves understanding the business objectives of customers, designing products to suit their needs, and supervising the building, testing, and deployment of these products. I develop hardware and software as well as supervise others who do so.

12. Also, from time to time I found other companies in order to pursue particular product opportunities. I develop and ultimately sell these companies. Most of my business activity, however, is as a consulting product developer. From time to time I have also served as an expert witness on computer and software related cases. I am a named inventor on twelve United States patents and am a member of several professional societies, including the IEEE and ACM.

13. I have extensive experience in several areas relevant to this case. My doctoral work was in the area of computer speech. As part of that work I developed signal processing systems and algorithms to filter and manipulate sound, especially speech. That work gave me strong practical insight into such topics as storing, retrieving, and manipulating audio and video information.

14. Much of my professional work over the past thirty years has been as a computer technology product developer. Several of those products involved technologies closely related to the topics in the patent at issue in the current matter, including digital audio systems and associated signal processing. One product in particular was for a company called IVT or burst.com in the later 1990s. That product involved developing a set of systems to record and transmit digital video and audio signals over the internet. Through this work I became very familiar with managing audio and video formats.

15. In addition to this work for IVT, I also worked with the Dun & Bradstreet Corporation in the 1980s to develop a system for interreacting with users of the credit reporting services over the telephone. It was one of the earliest interactive voice response systems that are now commonplace. This work involved the recording and manipulation of audio for use by the computer systems.

16. In 2000, I also supervised the development, design, and deployment of an industrial strength interactive voice response (IVR) system that allows users to access streaming media via their cell phones. This system supported up to 96 streams on a single chassis and was designed for a company called Savos. Under my supervision, the firm developed for this system the web-based content management system and interface for customizable user preferences.

17. Through these and other experiences, I am very familiar with the technologies at issue in this matter.

18. I have reviewed the ’066 patent, and relevant excerpts of the prosecution history of the ’066 patent. Based on my experience and education, and the acceptance of my publications and professional recognition by societies in my field, I believe that I am qualified to offer opinions as to the knowledge and level of skill of one of ordinary skill in the art at the time of the invention of the ’066 patent (which I further describe below).

III. LEVEL OF ORDINARY SKILL IN THE ART

19. 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 '066 patent. That person would have a bachelor's degree in computer science, computer engineering, electrical engineering, or equivalent training, and approximately two years' experience working with audio and visual content capture systems. Lack of work experience can substitute for additional education, and vice versa.

IV. RELEVANT LEGAL STANDARDS

20. I am not an attorney. In preparing and expressing my opinions and considering the subject matter of the '066 patent, I am relying on certain basic legal principles that Apple's counsel has explained to me.

21. I understand that prior art to the '066 patent includes patents and printed publications in the relevant art that predate the priority date of the '066 patent.

22. I have been informed by Apple's counsel that a claimed invention is unpatentable under 35 U.S.C. § 103 if the differences between the claimed invention and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a POSITA. I have also been informed by Apple's counsel that the obviousness analysis considers factual

inquiries, including the level of ordinary skill in the art, the scope and content of the prior art, and the differences between the prior art and the claimed subject matter.

23. I have been further informed by Apple's counsel that there are several recognized rationales for combining references or modifying a reference to show obviousness. These rationales include: (a) combining prior art elements according to known methods to yield predictable results; (b) simple substitution of one known element for another to obtain predictable results; (c) use of a known technique to improve a similar device (method, or product) in the same way; (d) applying a known technique to a known device (method, or product) ready for improvement to yield predictable results; (e) choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success; and (f) some teaching, suggestion, or motivation in the prior art that would have led a POSITA to modify the prior art or to combine prior art teachings to arrive at the claimed invention.

V. SUMMARY OF THE '066 PATENT

A. Overview of the '066 patent

24. The '066 patent describes a "capture device 100" for capturing content such as "[s]till pictures, moving pictures, audio, telemetry or other information." Ex.1001, 4:8-11. The content is processed by a "data converter"

included in the capture device, which the '066 patent describes as "any type of device that will capture the information and place it in some type of digitized format." Ex.1001, 4:10-26. 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, 4:27-35. 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, 5:62-67. The captured content is stored in association with the context tags and the additional metadata. Ex.1001, 6:7-14.

B. Effective Filing Date of the '066 patent

25. I have been instructed to apply January 8, 2007 as the effective filing date of the '066 patent for the purposes of my analysis. In this declaration the phrase "effective filing date" and "time of the invention" refers to January 8, 2007.

VI. CLAIM CONSTRUCTION

26. I understand that claim terms in IPRs are construed according to their "ordinary and customary meaning" to those of skill in the art. I have applied this interpretation throughout my analysis.

27. I also understand that the '066 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**
"*context tag*" (claims 6, 13)	"a searchable element derived from either a data element itself or from the context description element"
"associating the text and image recognition context tags with the digital image" (claims 6, 13) *"storing [. . .] the digital image in association with the text and image recognition context tags"* (claims 6, 13)	Plain and ordinary meaning
"a first digital audio format" (claims 13, 16)	Plain and ordinary meaning
"*the internal storage storing the digital image in association with the text and image recognition context tags*" (claim 6)	Plain and ordinary meaning
"image source" (claims 13, 16)	"an internal or an external image source"

28. I have applied these constructions throughout my analysis.

VII. THE CHALLENGED CLAIMS ARE UNPATENTABLE

29. The discussion in this Declaration provides a detailed analysis of how the asserted prior art references teach each limitation of the Challenged Claims.

30. As part of my analysis, I have considered, and discuss in detail, the scope and content of the prior art and any differences between the alleged invention and the prior art.

31. It is my opinion that the alleged invention recited in the Challenged Claims would have been obvious in view of the teachings of the asserted prior art and the knowledge of a POSITA before the time of the alleged invention of the '066 patent.

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

1. Spatharis

32. Spatharis describes a "system for extracting, processing, and sending metadata associated with audio data and/or video data." Spatharis, 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." Spatharis, [0012]. The camera system includes a "storage device" for "stor[ing] content that is acquired, generated, or processed by any of the subsystems" of the camera system. Spatharis, [0017].

33. Spatharis further describes that the camera system includes a "metadata extraction engine" that "provides analysis of the audio/video signal" including "speech to text conversion" of captured digital audio signals, and "visual

properties face recognition" of captured digital images. Spatharis, [0042]. Spatharis describes that this extracted metadata is associated with and stored with the digital image. Spatharis, [0042].

2. Manjunath

34. 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." Manjunath, 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. Manjunath, 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." Manjunath, 127. 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 '066 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.

3. Reasons to combine Spatharis and Manjunath

35. It is my opinion that 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.

36. As a threshold matter, Spatharis and Manjunath are analogous art because both references are in the same field of endeavor as the '066 patent. The '066 patent relates to capturing and storing audio and image content in association with context tags generated based on the content. Ex.1001, Abstract, 2:35-39, 5:60-6:2. Spatharis and Manjunath are similarly related to capturing and storing audio and image content in association with context tags generated based on the content. Spatharis, [0012], [0013], [0016] (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"); Manjunath, 127-28 (describing the "MPEG-7" format for storing text annotations in association with multimedia content). Thus, Spatharis and Manjunath are analogous art to the '066 patent.

37. Spatharis describes a system "for extracting, processing, and sending metadata associated with audio data and/or video data." Spatharis, 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**." Spatharis, [0044]; *see also id.*, [0026]. Spatharis thus mentions data formats generally, leading a POSITA to look to known formats for metadata storage.

38. Manjunath is a textbook describing known techniques for metadata storage: 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." Manjunath, 127 (italics omitted). In my opinion, 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. Manjunath, 127.

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). 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. In particular, implementing Spatharis with a tried-and-true industry standard such as MPEG-7 provides the benefit of wide-ranging interoperability. Manjunath, 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**.").

40. In my opinion, a POSITA would have had a reasonable expectation of success in making such a combination because Manjunath describes systems operating in the proposed manner. Manjunath, 127-128. Manjunath "offers a practical step-by-step walk through of the components, from systems to schemas to audio-visual descriptors." Manjunath, 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." Manjunath, 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."

4. Claim 6.

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

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

42. **First**, Spatharis describes:

> **[A] camera system**, which can also be referred to as an image acquisition and exploitation system, configured to extract, process, and/or send metadata associated with captured sound and/or images (e.g., video) is described. The captured sound and/or images can be referred to as content or captured content. **The camera system not only captures the sound and/or images, but also can extract and/or process metadata from the captured content**. The camera system can make the captured content and/or associated metadata available to a user for exploitation in substantially real-time. The extracting and/or processing

> of metadata from associated captured content can be referred to as enriching or indexing. The processing of captured content can be referred to as manipulation.
> Spatharis, [0012].

43. Spatharis thus teaches a "camera system" that performs a method for "**captur[ing]…content**," such as "**sound and/or images**" ("*method for capturing image and audio information"*), and "extract[ing] and/or process[ing] metadata from the captured content." Spatharis, [0012].

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." Spatharis, [0017].

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 method for capturing image and audio information for storage*."

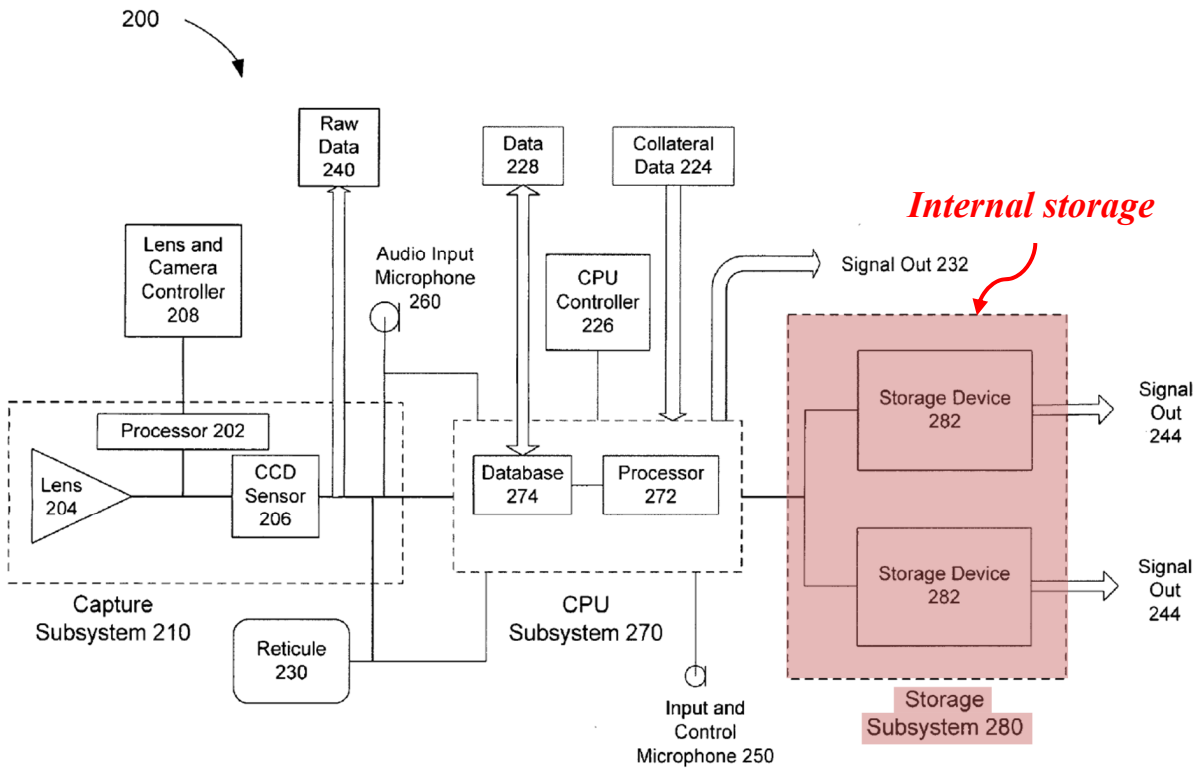
[6.1] ***executing a capture operation with a capture device by the steps of:***

46. 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." Spatharis, [0012]. Spatharis thus renders obvious "*executing a capture operation with a capture device*."

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

47. 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." Spatharis, [0017]. 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)." Spatharis, [0026].



Spatharis, Fig. 2 (annotated)

48. Thus, because Spatharis' camera system includes storage devices, Spatharis in the combination renders obvious "*providing internal storage*."

[6.3] ***interfacing a microphone with an external audio information source that generates external audio information and***

49. Spatharis describes that the camera system includes:

> **[A] capture subsystem 210**, also referred to as a visual and audio capture subsystem, includes a CCD sensor 206, a set of lenses 204, and a processor 202. The CCD sensor 206 captures, through the lenses 204, visual information (e.g., video) that is converted into a digital data. The digital data is made available to a reticule 230 for viewing

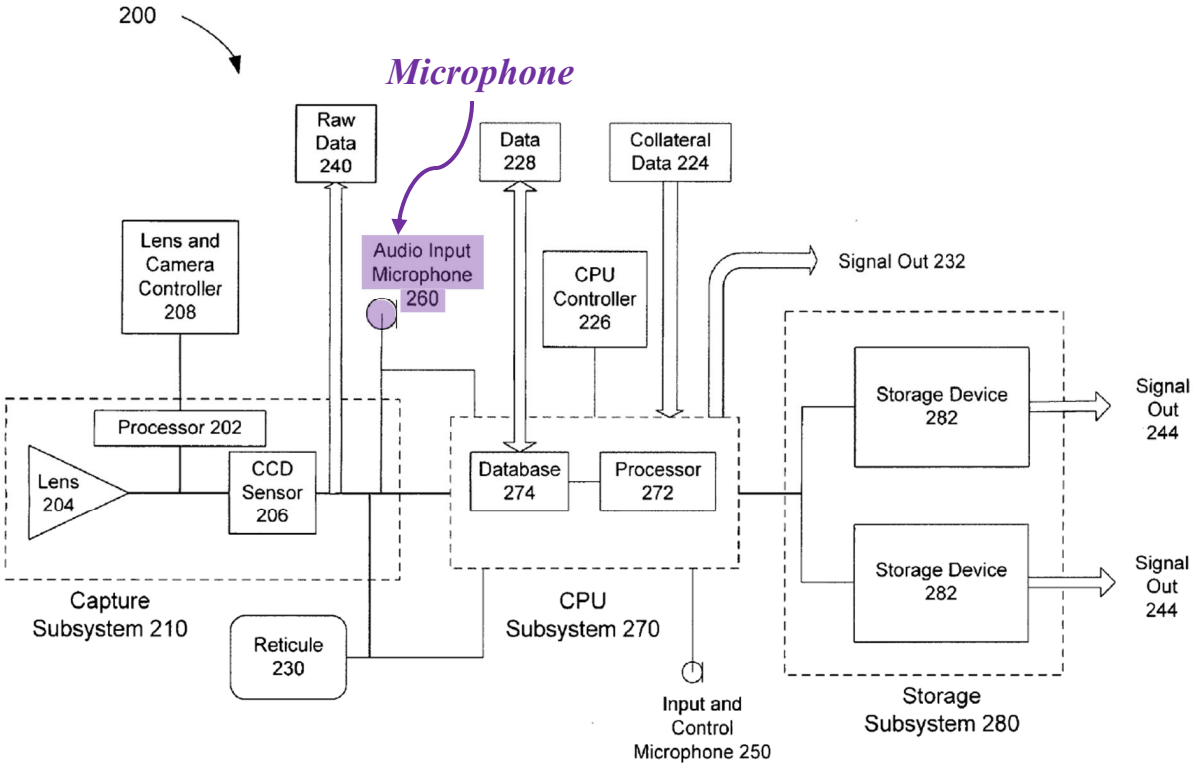
> as a raw (i.e. unenriched) picture. **The capture subsystem 210 also includes a source audio input microphone 260 that digitally captures sound**. The captured digital audio and/or video data, which can be collectively referred to as content, are made available to the CPU subsystem 270 for processing. 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. In some embodiments, the video and/or audio captured by the capture subsystem 210 can be analog signals that are converted into digital signals/data.

Spatharis, [0024].

50. Thus, 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**." Spatharis, [0024]. Spatharis' camera system captures the analog audio signals by "*interfacing"* the microphone 260 with an external environment in which an object ("*an external audio information source"*) is making a sound ("*generating external audio information"*) — consistent with the conventional operating details of microphones which were well-known many years prior to the '066 patent. *See, e.g.*, 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").

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



Spatharis, Fig. 2 (annotated)

52. Thus, because Spatharis' camera system includes a microphone for capturing sound from the environment, Spatharis in the combination renders obvious "*interfacing a microphone with an external audio information source that generates external audio information*."

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]), 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**." Spatharis, [0032]. 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* Spatharis, [0032].

[6.4] ***converting with a first data converter[1] the first external audio information from the microphone***

54. Spatharis describes that the camera system includes:

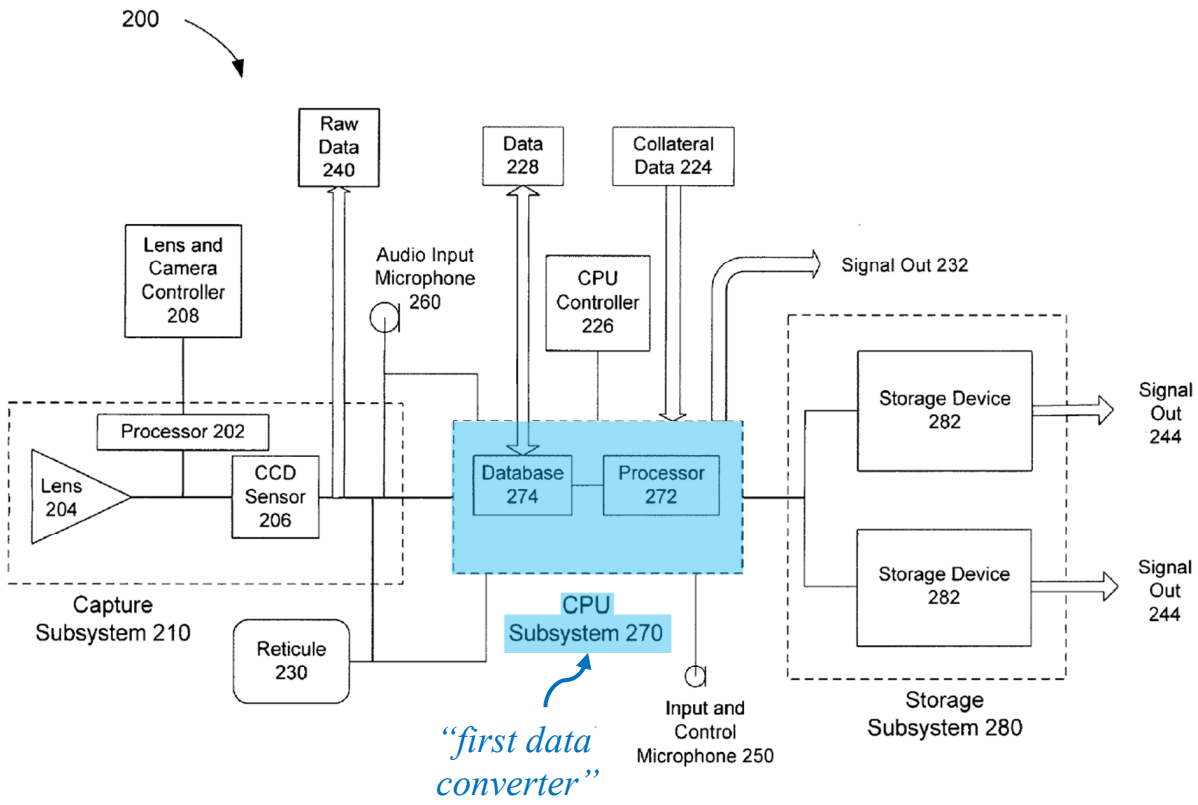
> a CPU subsystem 400 that executes various software applications including an operating system 430 and applications. Such applications can be, for example, third-

[1] 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.

> party applications (i.e., off-the-shelf applications). The applications can include, for example, metadata extraction engines 440 (e.g., Virage's Videologger and/or Audiologger), digital-asset-management 450 (e.g., Artesia Technologies' Digital Asset Management (DAM)) and browsing and editing programs 460 (e.g., Apple's Final Cut Pro).

Spatharis, [0039].

55. Thus, in the combination, Spatharis describes that the "**CPU subsystem 270**" ("*a first data converter*") 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. Spatharis, [0039]. 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**" by the "**CPU subsystem 270**" (the "*first data converter*") included in the camera system 200. Spatharis, [0023]-[0024]. Spatharis' FIG. 2 shows this configuration:



Spatharis, Fig. 2 (annotated)

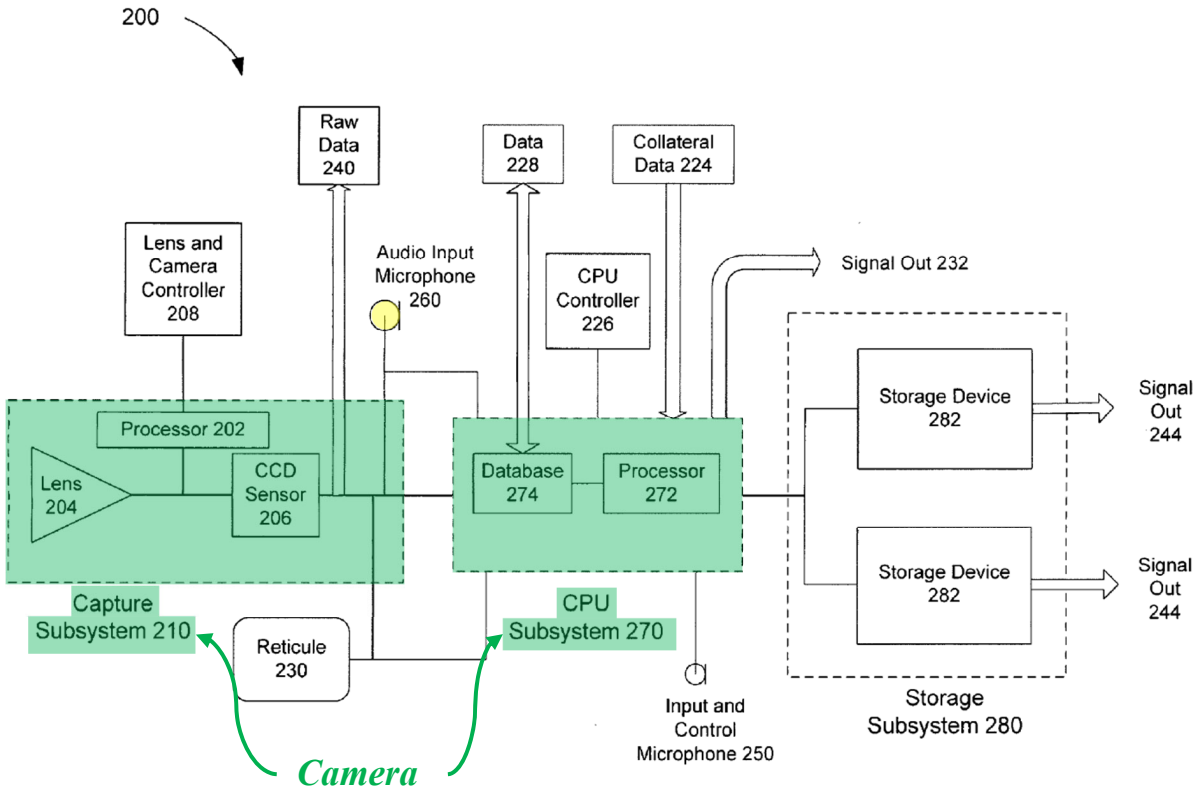
56. Thus, because Spatharis' camera system includes a CPU subsystem that converts the audio signal from the microphone to digital data, Spatharis in the combination renders obvious "*converting with a first data converter the first external audio information from the microphone*."

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

57. **First**, Spatharis teaches a "*camera*" in that the camera system 200 includes a "**capture subsystem 210**." Spatharis, [0023]. 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." Spatharis, [0024]. The capture subsystem 210, including the "still camera," and the CPU subsystem 270, collectively, teach the claimed "*camera*." Spatharis, [0023]-[0024]. Before the effective filing date of the '066 patent (and still today), the term "camera" in the context of digital cameras referred to both the image sensor and associated processor (e.g., CPU subsystem 270 in Spatharis). *See, e.g.,* Hubel (Ex.1025), 2:51-54 (explaining that a digital camera "may broadly be viewed as including an imager and a processor"); Stavely (Ex.1026), [0014] (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.").

58. Spatharis' FIG. 2 shows this configuration:



Spatharis, Fig. 2 (annotated)

59. **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*.") Spatharis, [0023]. The object having its picture taken is an "*external image source*" as claimed.

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

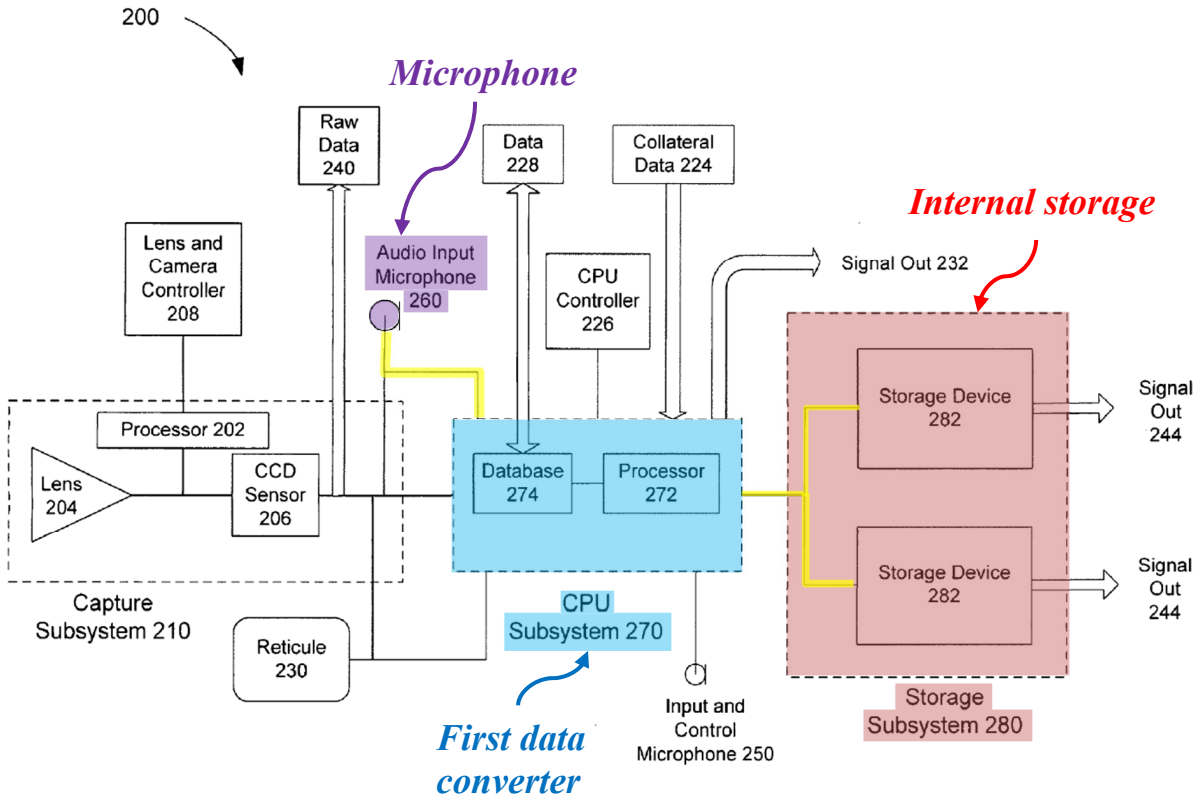
[6.6] ***the first data converter processing the captured external audio information and storing it in a first digital audio format as stored digital audio in internal storage within the capture device***

61. **First**, as previously discussed in [6.4], in the combination, Spatharis' "CPU subsystem 270" teaches "*a first data converter*." In addition, as previously discussed in [6.3] and [6.4], Spatharis teaches that the CPU subsystem 270 "*convert[s] … the first external audio information from the microphone*" into a "digital signals/data." Spatharis, [0024].

62. **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"*). Spatharis, [0023].

63. **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*"). Spatharis, [0026]; *see also* [0044].

64. 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*"). Spatharis, [0023]-[0024], [0026].



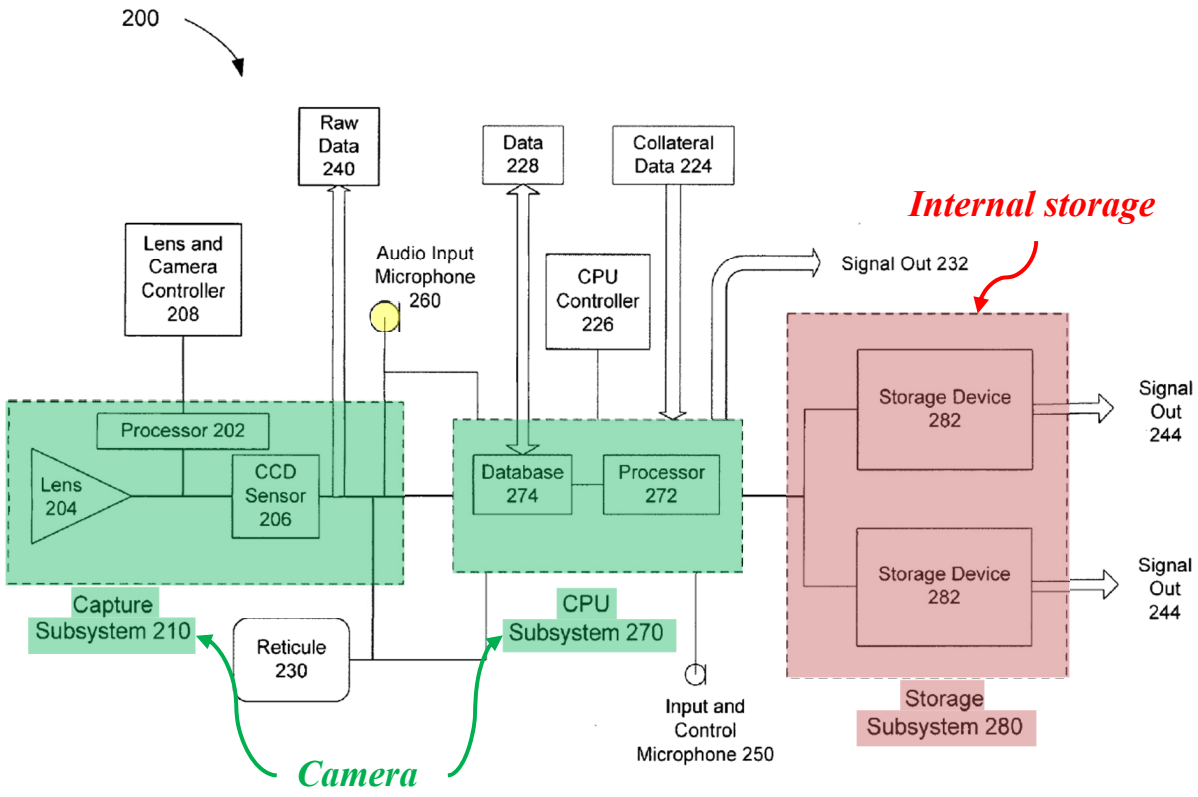
Spatharis, Fig. 2 (annotated)

65. 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*."

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

66. **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:



Spatharis, Fig. 2 (annotated)

67. **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*"). Spatharis, [0023].

68. **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*"). Spatharis, [0023].

69. 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*."

[6.8] ***capturing with a data capture device, as captured data, location information, and time information associated with at least the capture of the image and***

70. **First**, Spatharis describes that the camera system captures "**collateral data**" ("*captured data"*) such as "**date, time**" ("*time information"*) and "**GPS**" ("*location information"*) associated with the capture of the image. Spatharis, [0040]. This collateral data is "generated" ("*captur[ed]"*) by specialized "applications and/or hardware" ("*a data capture device"*). Spatharis, [0040].

71. **Second**, Spatharis' collateral data is "associated with at least the capture of the image" because Spatharis explains that "[c]ollateral data includes… additional marking and/or annotating data that can be **used by the CPU subsystem 270 to process image and/or sound data**." Spatharis, [0033].

72. Thus, because Spatharis' device obtains collateral data such as GPS location and time/date information that is used to mark/annotate image and sound data, Spatharis in the combination renders obvious "*capturing with a data capture*

device, as captured data, location information, and time information associated with at least the capture of the image."

[6.9] ***storing the captured data as stored captured data;***

73. **First**, as discussed at [6.8], Spatharis teaches that the camera system captures "**collateral data**" ("*captured data*") such as "date, time" and "GPS" information associated with the capture of the image. Spatharis, [0040].

74. **Second**, Spatharis further teaches that the "collateral data" is "received by the CPU subsystem," "processed" by one or more "applications" executed by the CPU subsystem, and "output to [the] storage subsystem" ("*internal storage*"). Spatharis, [0040]-[0041].

75. Thus, because Spatharis stores the location and time/stamp information in storage, Spatharis in the combination renders obvious "*storing the captured data as stored captured data*."

[6.10] ***converting with a media data converter the received digital audio to a text based searchable file as a text context tag and***

76. **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*."). Spatharis, [0042]. Spatharis describes that the camera system includes a CPU subsystem that "**executes applications for manipulating content captured by the capture subsystem 210**." Spatharis,

[0029]; *see also id.*, [0015] (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*"). Spatharis, [0042]. 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**.**" Spatharis, [0013]. Spatharis' FIG. 4 shows this configuration:

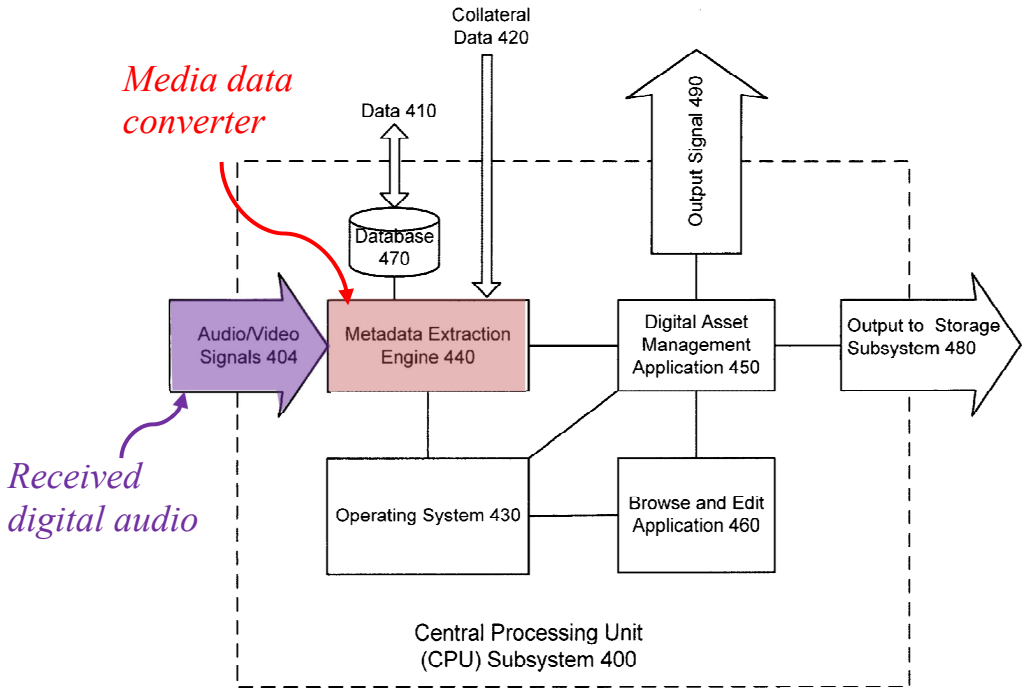


FIG. 4

Spatharis, Fig. 4 (annotated)

77. **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*"). Manjunath, 127. Manjunath explains that "[t]he practice of **using text annotations to search**, catalogue and index multimedia content is both longstanding and widespread." Manjunath, 127. And "MPEG-7 supports this practice with the TextAnnotation data type that allows free text, keyword, structured and dependency structure annotations." Manjunath, 127. Manjunath provides the following examples of an "**XML file**" for text annotations. Manjunath, 104, 127-128:

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

Manjunath, 127 (XML file examples for free text annotations)

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

Manjunath, 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>
```

Manjunath, 128 (XML file examples for structured annotations)

78. As previously discussed in §VII.A.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.

79. 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 "*converting with a media data converter the received digital audio to a text based searchable file as a text context tag*."

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

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

81. **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. Spatharis, [0042]. 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**.**" Spatharis, [0013]. 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*." Spatharis, [0013].

82. **Third**, Spatharis' system creates an image recognition "*searchable*" context tag: "[m]etadata can be used to aid a user and/or a computer system to, for example, **index/organize/classify**, retrieve and/or process captured content." Spatharis, [0013]. 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." Manjunath, 127. As previously discussed in §§VII.A.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.

83. 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*."

[6.12] ***associating the text and image recognition context tags with the digital image and the captured data, and***

84. **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. Spatharis, [0013], [0023], [0029], [0042].

85. **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*"). Spatharis, [0044].

86. 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*."

[6.13] ***storing in the internal storage the digital image in association with the text and image recognition context tags in addition to the stored captured data.***

87. 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." Spatharis, [0026]. 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]). Spatharis, [0026]; *see also id.*, [0012] ("The captured sound and/or images can be referred to as content").

5. Claim 7

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

88. **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. Spatharis, [0023]-[0024].

89. **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."** Spatharis, [0015].

90. 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*." Spatharis, [0015].

6. Claim 8

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

91. 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*"). Spatharis, [0023], [0024].

Spatharis' "still camera" captures a still image representing the state of the external image source at a moment in time. Spatharis, [0023].

92. Accordingly, because Spatharis' system captures an image at an instant in time using a digital still/video camera, Spatharis renders obvious "*the step of capturing captures the image with the camera from the external image source at an instant in time*."

7. Claim 9

[9.1] ***The method 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.***

93. 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*"). Spatharis, [0023]-[0024]. 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*"). Spatharis, [0012], [0015], [0022], claim 7. It is also generally

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

8. Claim 10

[10.1] ***The method of claim 6, and further including: transmitting with a transmitter associated with the capture device the composite data set[2] to the location on the network[3];***

94. 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)." Spatharis, [0034]. "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**." Spatharis, [0034]. Transmitting stored content ("*the composite data set*") to a library system makes use of "I/O ports" ("*a transmitter*"). Spatharis, [0019], [0034]. "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 180 can also be used to extract signals, content,

[2] I have been instructed to interpret the term "*the composite data set*" as "*a composite data set*."

[3] I have been instructed to interpret the term "*the location on the network*" as "*a location on a network*."

and/or metadata from the camera system 100 to outside systems." Spatharis, [0019]; *see also id.*, [0034]. 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. Spatharis, [0022].

95. 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 "*transmitting with a transmitter associated with the capture device the composite data set to the location on the network*."

[10.2] ***wherein a system disposed at the location on a network operates to receive with a receiver the transmitted composite data set from the transmitter associated with the capture device as a received set of captured information, provide a database, and store in the database the received composite data set.***

96. **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 operates to receive … the transmitted composite data set from the transmitter*"). Spatharis, [0034]. 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. *See, e.g.*, Cox (Ex.1017), 9:5-12 (describing a "a personal

computer…capable of receiving" data from a network through a "cable interface"); Hershey (Ex.1016), 6:8-18 (describing the use of a "communication interface" for sending and receiving data from a network).

97. **Second**, in my opinion, Spatharis teaches storing the received data in a "*database*" because the "asset-management," "archiving," and "library" systems described in Spatharis store the received data, as such systems were well-known to do. Spatharis, [0034]; *see, e.g.*, Chadwick (Ex.1014), Abstract, [0023] (describing a system that "archives" content and extracted metadata in a database); Hsiao (Ex.1015), Abstract (describing an "archival" system that stores information in a database).

98. 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 operates to receive with a receiver the transmitted composite data set from the transmitter associated with the capture device as a received set of captured information, provide a database, and store in the database the received composite data set*."

9. Claim 11

[11.1] ***The method of claim 10, wherein the step of capturing with 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.***

99. *See* [9.1].

10. Claim 12

[12.1] ***The method of claim 11, wherein the step of transmitting transmits the received composite data set to the location on the network after at least the stop event associated with the processing of the captured external audio information.***

100. **First**, as described above at [9.1], Spatharis' camera system stores content after a "*stop event*."

101. **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)." Spatharis, [0034]. "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**." Spatharis, [0034]. The "stored content" 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*"). Spatharis, [0034].

102. Accordingly, because Spatharis describes transmitting the stored content to an archive or library system, Spatharis in the combination renders obvious "*wherein the step of transmitting transmits the received composite data set to the location on the network after at least the stop event associated with the processing of the captured external audio information*."

11. Claim 13

[13.0] ***A method for capturing image and audio information for storage, comprising the steps of:***

103. *See* [6.0].

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

104. *See* [6.2].

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

105. *See* [6.3].

[13.3] ***converting with a first data converter the external audio information from the microphone;***

106. *See* [6.4].

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

107. *See* [6.5].

[13.5] ***capturing with a capture device, as captured data, location information and time information associated with at least the capture of the image and***

108. *See* [6.8].

[13.6] ***storing the captured data as stored captured data;***

109. *See* [6.9].

[13.7] ***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,***

110. *See* [6.6].

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

111. *See* [6.7].

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

112. *See* [6.10]. Spatharis' metadata extraction engine teaches both a "*media data converter*" (recited in [6.10]) and a "*second data converter*" (recited in [13.9]). Spatharis, [0013], [0042].

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

113. *See* [6.11].

[13.11] ***associating the text and image recognition context tags with the digital image and with the stored captured data; and***

114. *See* [6.12].

[13.12] ***storing in the internal storage the digital image in association with the text and image recognition context tags in addition to the stored captured data.***

115. *See* [6.13].

12. Claim 14

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

116. *See* [6.5].

13. Claim 15

[15.0] ***The method of claim 13, wherein the step of converting with the first data converter captures the first external audio information from the microphone during generation thereof.***

117. *See* [7.1].

14. Claim 16

[16.0] ***The method of claim 13, wherein the step of capturing captures with the camera the image from the image source at an instant in time.***

118. *See* [8.1].

15. Claim 17

[17.0] ***The method of claim 13, wherein the step of converting with 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.***

119. *See* [9.1].

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

1. Fuller

120. 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." Fuller, 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. Fuller, 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. Fuller, 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. Fuller, 2:52-3:9. The captured content is stored in association with the metadata descriptions and the additional metadata. Fuller, 8:1-15.

2. Jain

121. Fuller incorporates Jain (U.S. patent applications Ser. No. 09/134,500) by reference. Fuller, 7:50-68. Jain describes a "speech transcription module" which is an example implementation of the audio/video analysis engine of Fuller. Fuller, 7:50-68; Jain, 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. Jain, 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. Fuller, 8:1-15.

3. Reasons to combine Fuller and Jain

122. It is my opinion that 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. As noted above, Fuller incorporates Jain by reference. Fuller, 7:50-68.

123. 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." Fuller, 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.

124. 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.

4. Claim 6

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

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

126. **First**, Fuller describes "a conventional **method**" by which "[d]igital video frames are captured sequentially by a CCD sensor 201, while audio is captured by a microphone 202" ("[*a*] *method for capturing image and audio information*"). Fuller, 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. Fuller, 7:30-51.

127. **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.*" Fuller, 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." Fuller, 8:1-15.

128. Thus, Fuller's method of 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*] *method for capturing image and audio information for storage*" as claimed.

[6.1] ***executing a capture operation with a capture device by the steps of:***

129. 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 ("*executing a capture operation*"). *See, e.g.,* Fuller, 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**." Fuller, 1:26-32; *see also id.*, 3:9-18. Fuller's capture device is shown in Fig. 4 below.

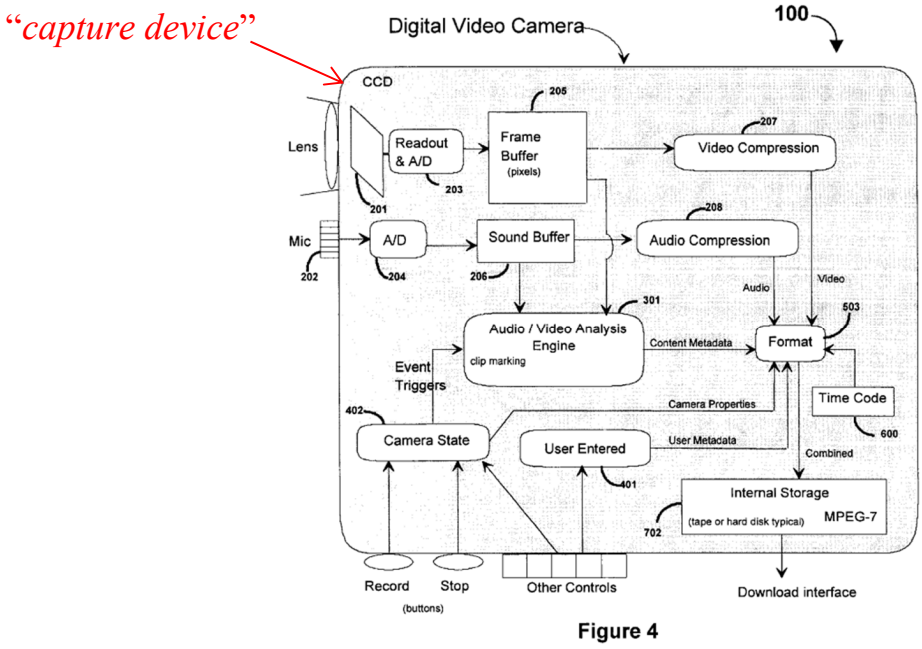


Figure 4

Fuller, FIG. 4 (annotated)

130. Fuller's digital video camera "*execut[es] a capture operation*" when "[d]igital video frames are captured sequentially by a CCD sensor 201" and "audio is captured by a microphone 202." Fuller, 7:30-51.

131. Thus, because Fuller describes capturing video and audio with a digital video camera, Fuller in the combination renders obvious "*executing a capture operation with a capture device.*"

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

132. Fuller's digital video camera "*provid[es] internal storage*" because it includes a frame buffer 205, sound buffer 206, and internal storage 700, 702, 703, 704. Fuller, 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. Fuller, 7:30-51.

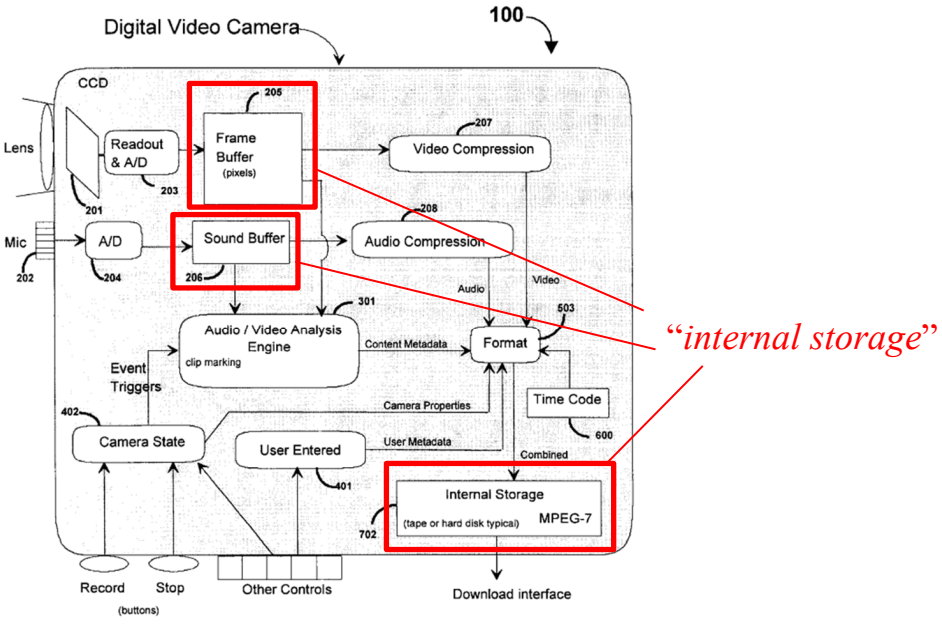


Figure 4

Fuller, FIG. 4 (annotated)

133. Thus, Fuller's device includes a frame buffer, the sound buffer, and internal storage, which in the combination renders obvious "*providing internal storage.*"

[6.3] ***interfacing a microphone with an external audio information source that generates external audio information and***

134. Fuller's microphone 202 ("*microphone*") captures audio information from the environment and thus "*interface[es] … with an external audio information source that generates external audio information*." Fuller, 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"). Fuller's microphone 202 is shown, for example, in Fig. 4 below.

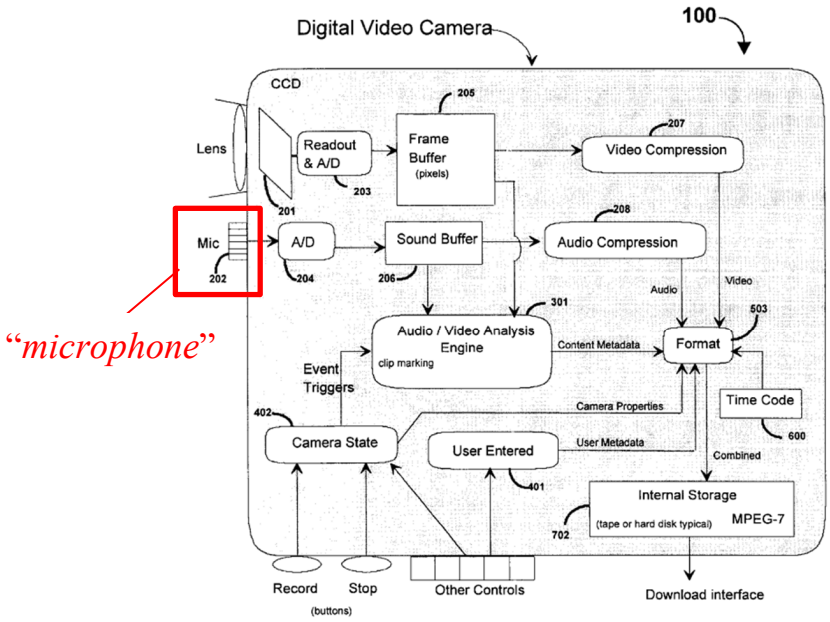


Figure 4

Fuller, FIG. 4 (annotated)

See also, e.g., Fuller, FIGS. 1, 5.

135. Thus, because Fuller's microphone captures audio information from the environment, Fuller in the combination renders obvious "*interfacing a microphone with an external audio information source that generates external audio information.*"

[6.4] ***converting with a first data converter[4] the first external audio information***

[4] 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

from the microphone,

136. 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 ("*converting… the first external audio information from the microphone*"). Fuller, 7:30-51, 5:30-36 (output of microphone is "converted to digital form").

137. 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)." Fuller, 7:30-51. 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. Fuller, 7:30-51, 5:31-35. Fuller's A/D converter 204 is shown in Fig. 4 below.

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.

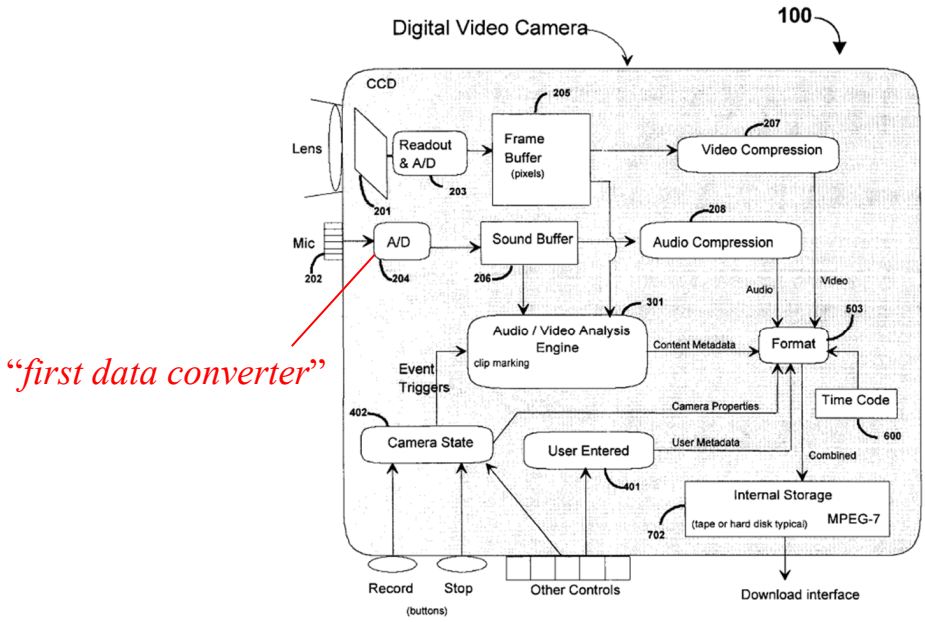


Figure 4

Fuller, FIG. 4 (annotated)

See also, e.g., Fuller, FIGS. 1, 5.

138. 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 "*converting with a first data converter the first external audio information from the microphone*."

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

139. **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." Fuller, 5:29-42, 7:30-51.

140. **Second**, Fuller's camera takes pictures and records video ("*interface[es] … with an external image source to capture an image therefrom*.").

Fuller, 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. Fuller, 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.

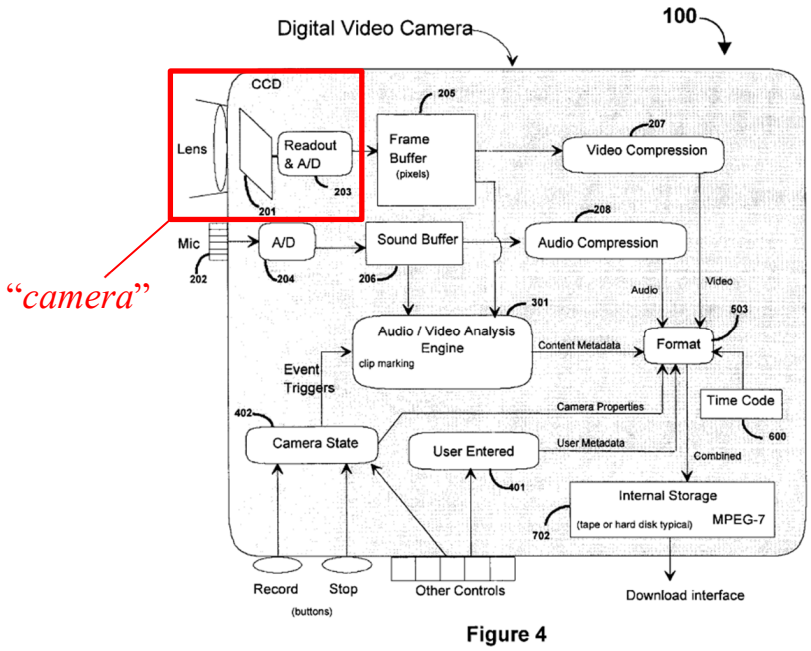


Figure 4

Fuller, FIG. 4 (annotated)

See also, e.g., Fuller, FIGS. 1-3, 5.

141. 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 "*interfacing a camera with an external image source to capture an image therefrom*."

[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,

142. **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." Fuller, 5:30-36, 7:30-51. 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*"). Fuller, 5:30-36, 7:30-51.

143. **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"). Fuller, 5:30-39 (the "digital content is then stored in a memory unit 205, 206"), 7:30-51.

144. 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.*, IPR2023-00025, Paper 21 at 43 (Feb. 16, 2024) ("We understand Fuller's disclosure of 'digital form' to be equivalent to a digital format.").

145. 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.,* Hindersson (Ex.1019), [0042] ("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."), [0032] (referring to "PCM format."); Cox-2 (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.").

146. Thus, in my opinion, the "digital form" referred to by Fuller is a "*digital audio format*" such as a PCM format. Fuller, 5:22-42, 7:30-51. Accordingly, Fuller renders obvious a "*first digital audio format*" as claimed.

147. **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*."

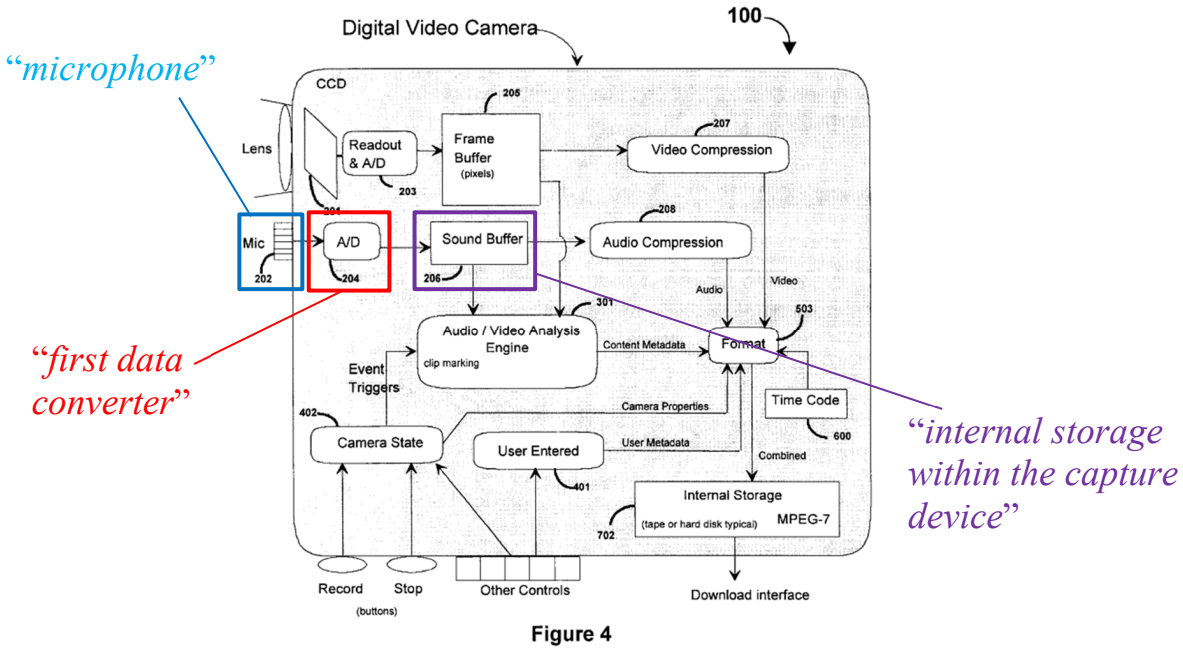


Figure 4

Fuller, FIG. 4 (*annotated*)

See also, e.g., Fuller, FIGS. 1, 5.

148. 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*."

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

149. **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").

150. 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"). Fuller, 7:30-51. 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."

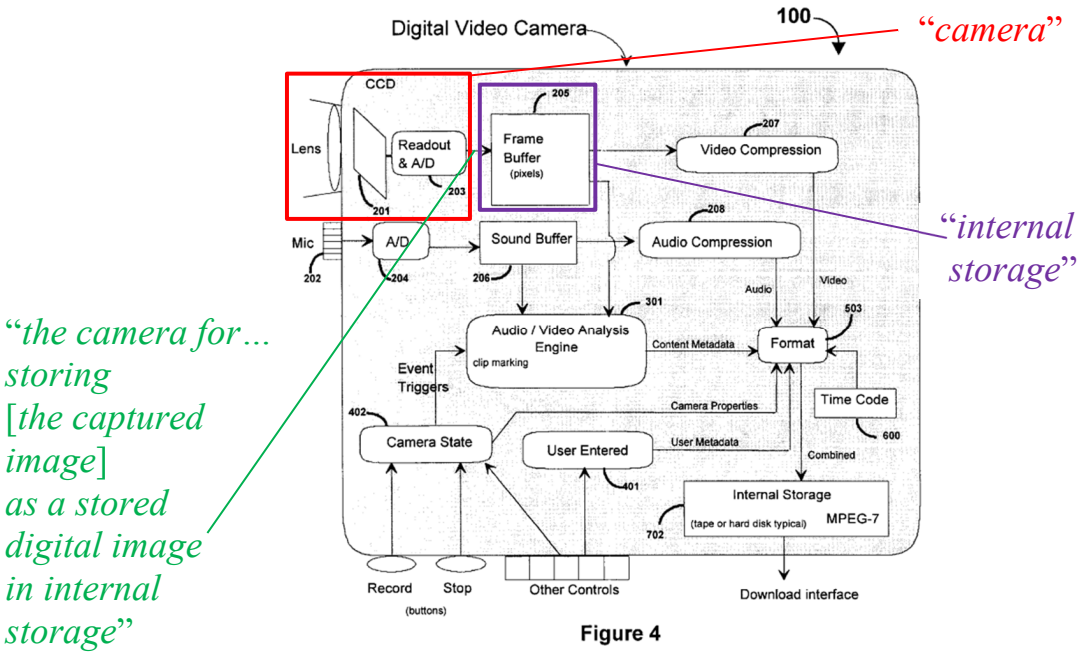


Figure 4

Fuller, FIG. 4 (annotated)

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

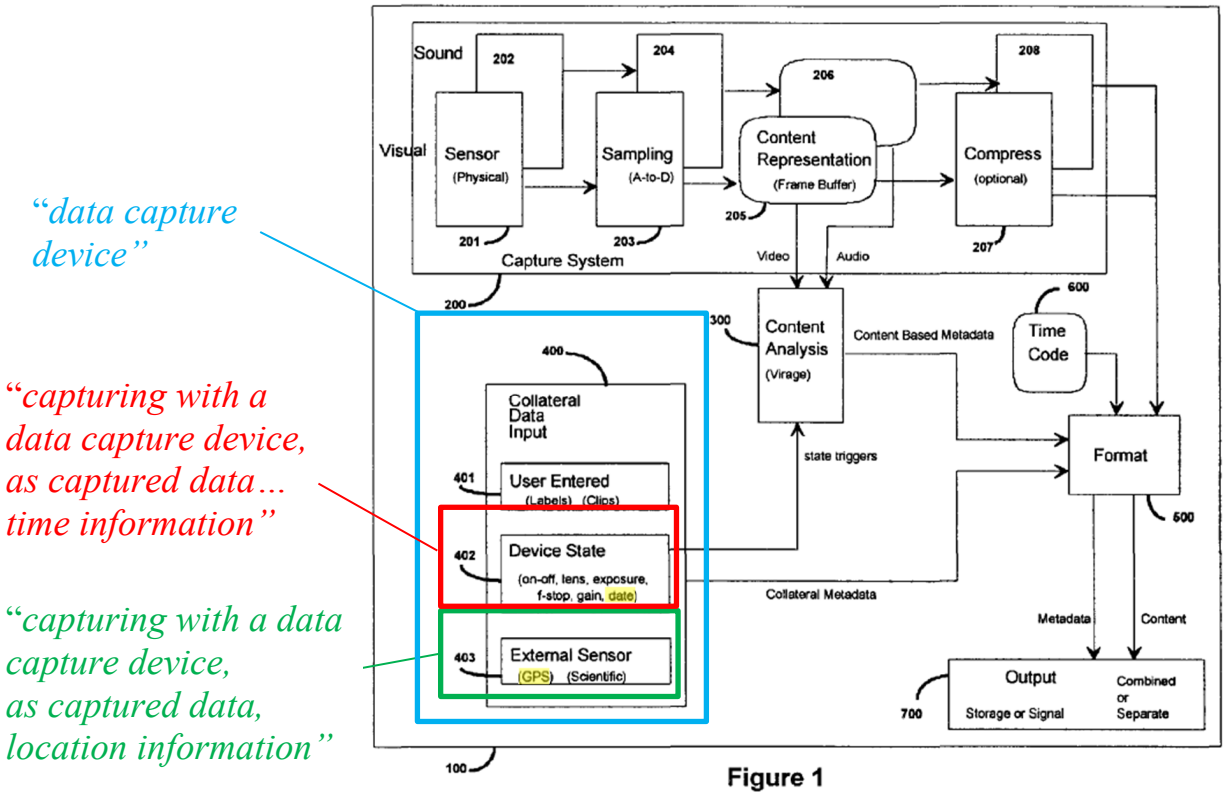
151. Thus, because the visual sensor passes the captured information to the A/D unit for processing and storage as a digital image in the frame buffer, Fuller renders obvious "*the camera for processing the captured image and storing it as a stored digital image in internal storage*."

[6.8] ***capturing with a data capture device, as captured data, location information, and time information associated with at least the capture of the image***

152. **First**, Fuller's digital video camera 100 includes a collateral data input block 400 ("*data capture device*") that captures time and location information. Fuller, 5:22-65. For example, the block 400 may include an external sensor 403 (e.g., GPS unit) to capture location information and a camera state 402 to capture "time/date" information. Fuller, 5:22-65; *see also id.*, 2:24-40, 2:52-3:8, 3:45-4:23, 6:50-65, 9:33-40.

153. **Second**, Fuller's external sensor 403 of block 400 ("*data capture device*") captures "position derived from GPS satellites." ("*capturing with a data capture device, as captured data, location information*"). Fuller, 5:22-65. Because the location information for a particular image capture is taken "at the point of capture," it is "*associated with at least the capture of the image*." Fuller, 2:19-40, 3:45-4:23, 5:41-57, 6:12-36, 6:50-65, claims 1, 5, 6, 14, 18, and 19.

154. **Third**, Fuller's "collateral metadata, such as date/time" is obtained "at the point of content capture" ("*associated with at least the capture of the image*"). Fuller, 2:19-40.



Fuller, FIG. 1 (annotated)

155. Thus, because Fuller's device records GPS satellite position information and times and dates of shots at the point of content capture, Fuller in the combination renders obvious "*capturing with a data capture device, as captured data, location information, and time information associated with at least the capture of the image*."

[6.9] ***and storing the captured data as stored captured data;***

156. **First**, as described above at [6.8], Fuller's digital video camera 100 captures the GPS satellite position information ("*location information*") and the "date/time" ("*time information*").

157. **Second**, Fuller's digital video camera 100 stores the "collateral metadata," which includes the GPS satellite position information and the "date/time," as a file, such as a "FlashPix formatted file" or "MPEG-7 stream" in the internal storage 700, 702 ("*storing the captured data as stored captured data*"). Fuller, 5:56-7:10; *see also id.*, 8:1-15.

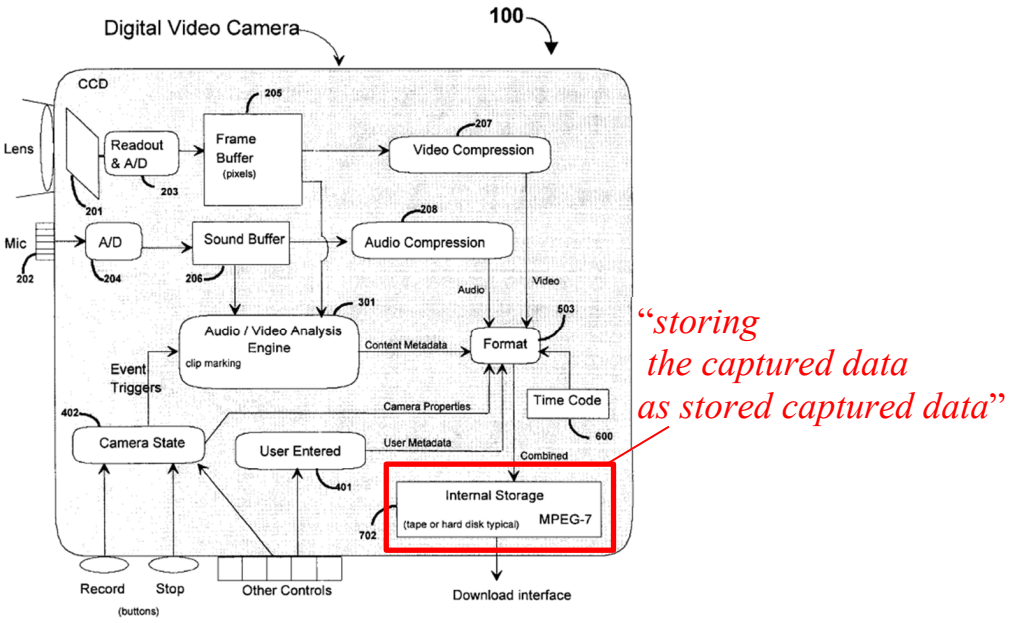


Figure 4

Fuller, FIG. 4 (annotated)

158. Thus, Fuller's device stores the "collateral metadata" as a file, which in the combination renders obvious "*storing the captured data as stored captured data*." Fuller, 5:56-7:10; *see also id.*, 8:1-15.

[6.10] ***converting with a media data converter the received digital audio to a text based searchable file as a text context tag***

159. **First**, Fuller's digital video camera 100 includes an audio/video analysis engine 301 ("*media data converter*") that converts media, such as audio and video signals, into extracted metadata. Fuller, 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. Fuller, 7:50-67.

160. 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.

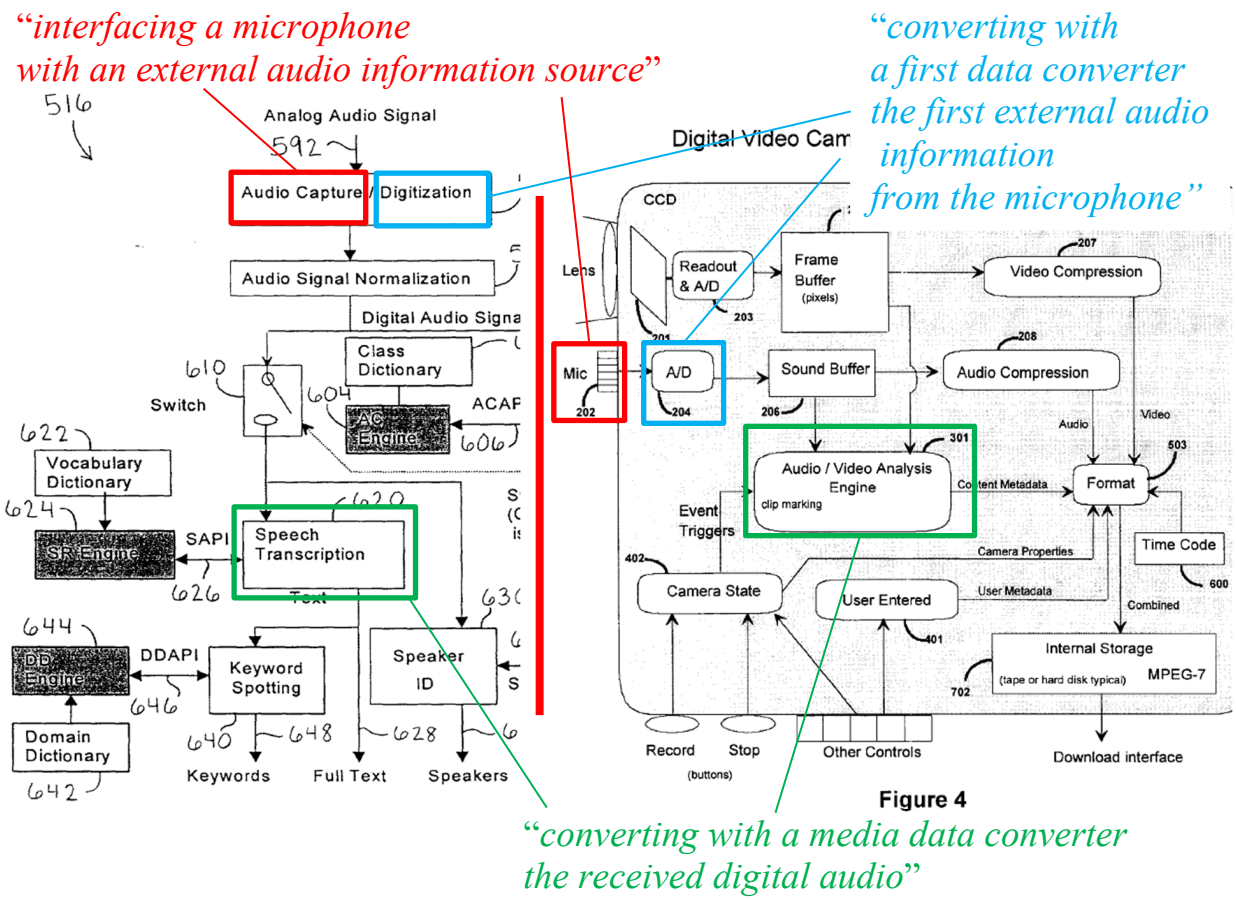
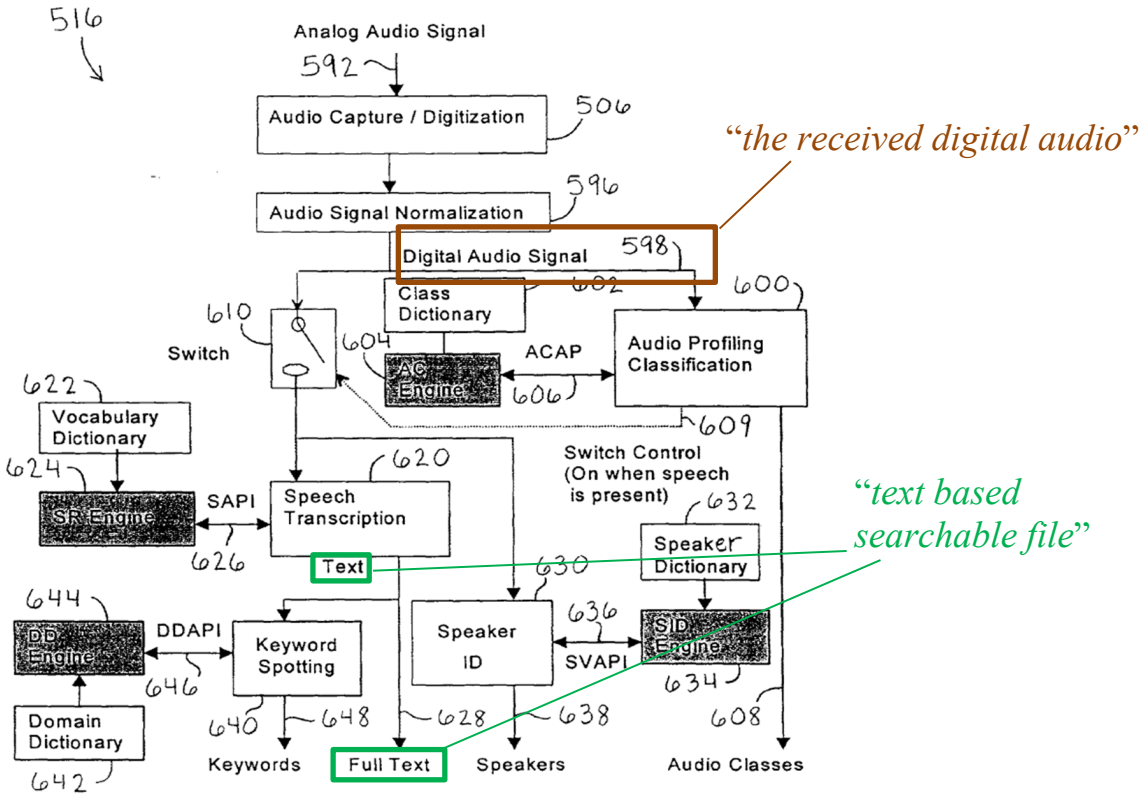


Figure 4

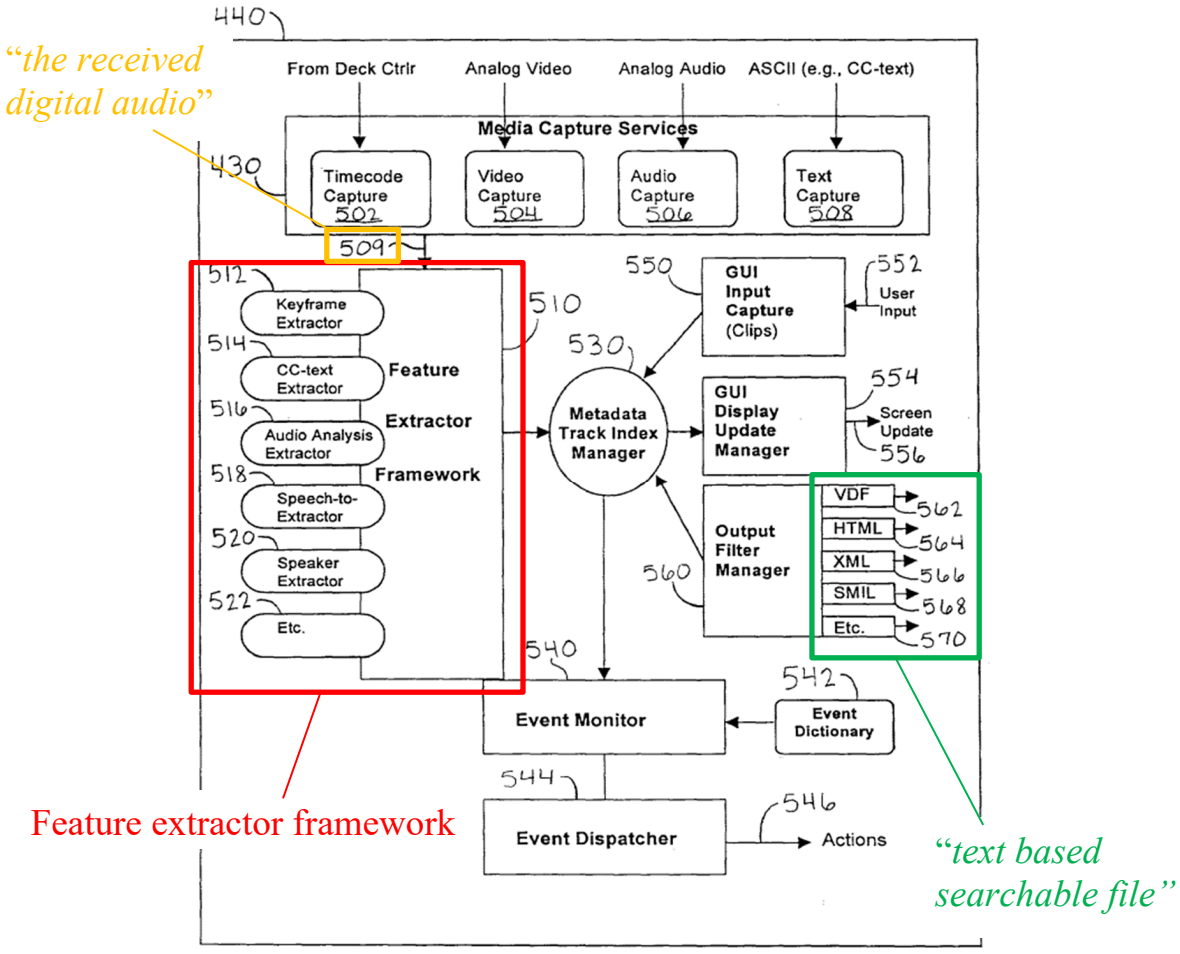
Jain, FIG. 10 (partial, annotated). **Fuller, FIG. 4 (partial, annotated)**

161. 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 ("*converting with a media data converter the received digital audio to a text based searchable file"*). Jain, 9:33-10:27.



Jain, FIG. 10 (annotated)

162. 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*"). Jain, 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." Jain, 8:41-57, 9:33-10:27.

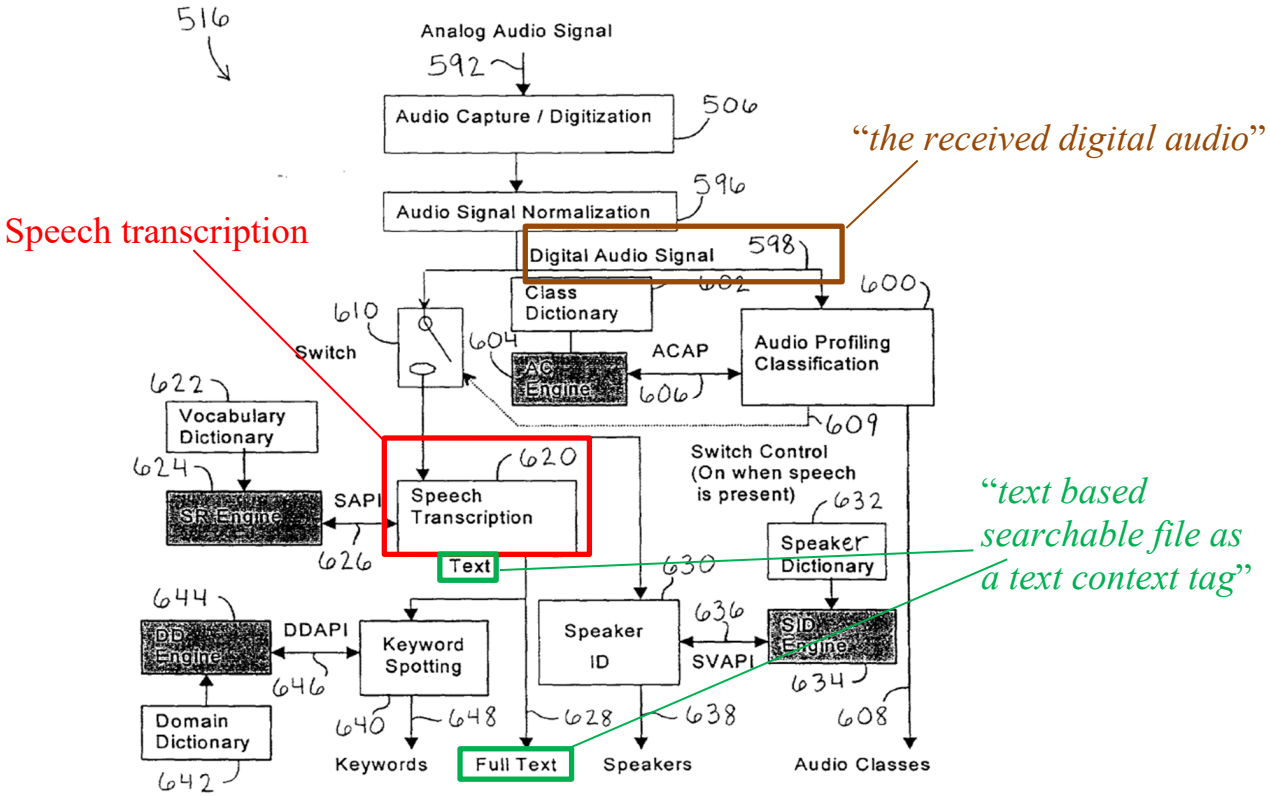


Jain, FIG. 9 (annotated)

163. 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." Jain, 9:33-10:27.

164. **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 ("*media data converter*") and thus provide context to the digital audio signal. Jain, 9:33-10:27; Fuller, 7:50-67. 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." Jain, 9:33-10:27.



Jain, FIG. 10 (annotated)

165. In this vein, I understand that, 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. Jain, 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* Jain, 9:33-10:27. Thus, even using the prior claim construction, the full text 628 and/or the text output 648 teach "*a text context tag*."

166. 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 "*converting with a media data converter the received digital audio to a text based searchable file as a text context tag*."

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

167. **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*"). Fuller, 2:52-3:9.

168. 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*"). Fuller, 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**." Fuller, 2:52-3:9.

169. 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**." Fuller, 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." Fuller, 2:52-3:9.

170. 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*

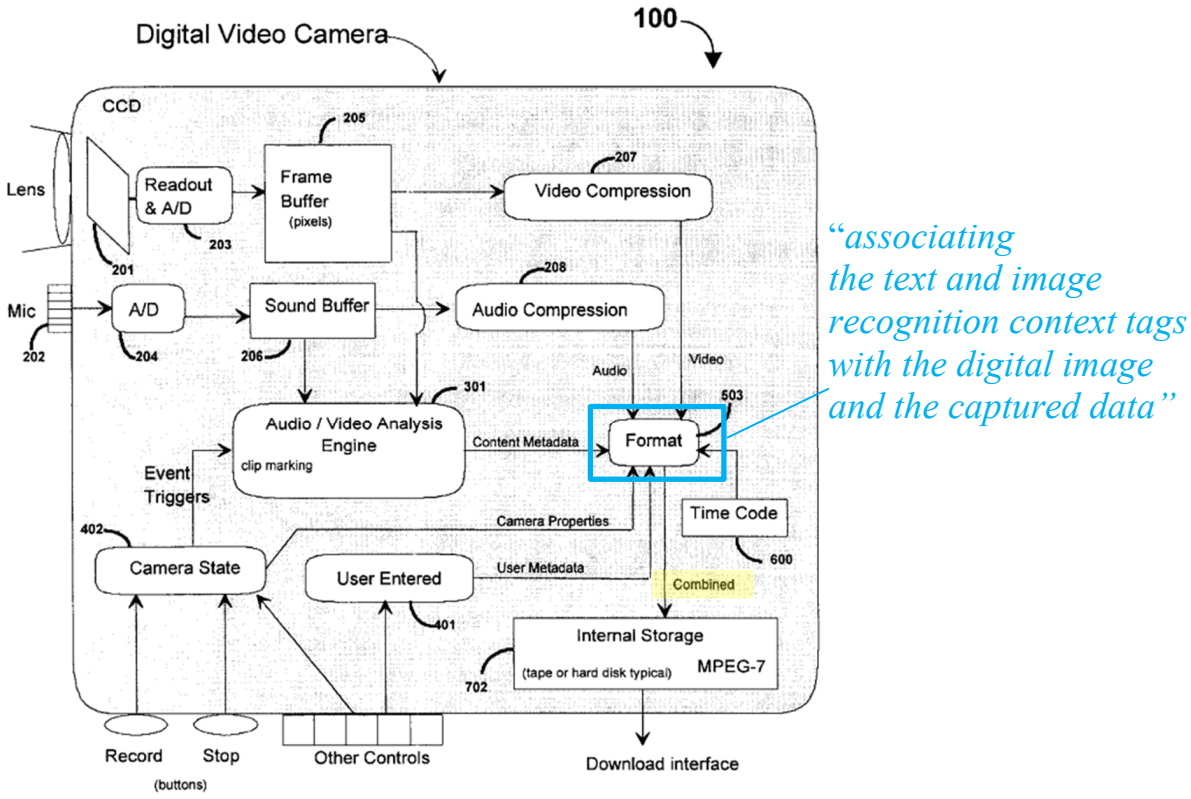
an image recognition searchable context tag with image recognition of at least a portion of the digital image."

[6.12] ***and associating the text and image recognition context tags with the digital image and the captured data,***

171. **First**, as described above at [6.10] and [6.11], 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. Fuller, 7:50-67.

172. **Second**, as described above at [6.8] and [6.9], Fuller teaches obtaining location and time data ("*captured data*").

173. **Third**, Fuller's device includes a formatting unit 503 that synchronizes ("*associate[es]*") the "[c]ontent-based metadata from the analysis engine 301" and "collateral metadata from block[ ] 402" ("*the captured data*") "with the video content" ("*the digital image*"). Fuller, 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 and the captured data*"). Fuller, 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." Fuller, 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:



Fuller, FIG. 4 (annotated)

174. Thus, because both the metadata extracted from the digital audio and video and the collateral metadata are sent to the formatting unit and synchronized, Fuller in the combination renders obvious "*associating the text and image recognition context tags with the digital image and the captured data*."

[6.13] ***and storing in the internal storage the digital image in association with the text and image recognition context tags in addition to the stored captured data.***

175. **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*."

176. **Second**, as shown above with respect to [6.12], the formatting unit 503 synchronizes the content-based metadata and collateral metadata with the video content "in the form of an MPEG-7 stream." Fuller, 8:1-15.

177. **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" ("*storing in the internal storage the digital image in association with the text and image recognition context tags in addition to the stored captured data*"). Fuller, 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." Fuller, 6:65-7:10.

178. 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 "*storing in the internal storage the digital image in association with the text and image recognition context tags in addition to the stored captured data*."

5. Claim 7

[7.1] ***The method of claim 6, wherein the first data converter captures the first***

external audio information from the microphone during generation thereof.

179. 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. Fuller, 7:30-51; *see also id.*, 5:29-42. For example, Fuller specifies that "audio is captured by a microphone 202" and that "[this] signal[ ] is digitized by… an A/D converter 204." Fuller, 7:30-51.

180. 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." Fuller, 4:30-32.

181. 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*."

6. Claim 8

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

182. 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 ("*a 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 step of capturing captures the image with the camera from the external image source at an instant in time*"). Fuller, 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." Fuller, 7:32-34.

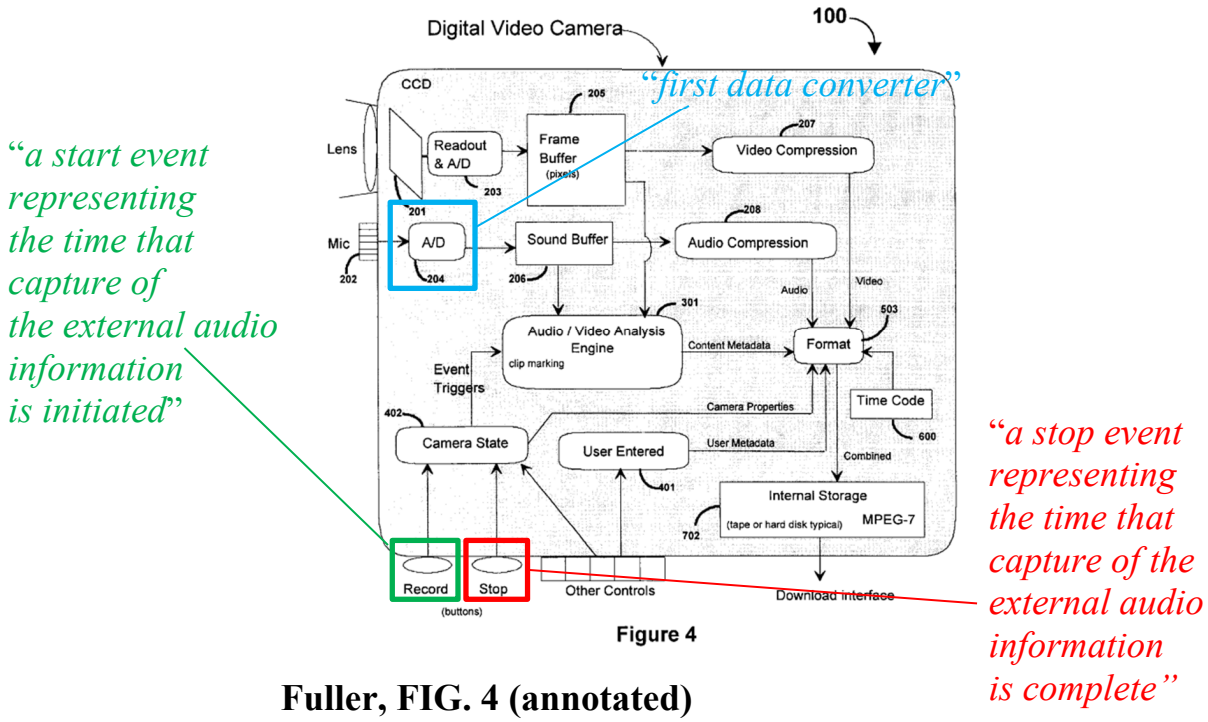
183. Accordingly, because Fuller's system captures an image at an instant in time using a digital still/video camera which may include a visual sensor such as a CCD chip, Fuller in the combination renders obvious "*the step of capturing captures the image with the camera from the external image source at an instant in time*."

7. Claim 9

[9.1] ***The method 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.***

184. 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*"). Fuller, 4:1-16; *see also id.*, 5:43-57, 7:50-68, 8:43-9:6, claims 4 and 17.



Fuller, FIG. 4 (annotated)

See also, Fuller, FIG. 5.

185. 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."

8. Claim 10

[10.1] ***The method of claim 6, and further including: transmitting with a transmitter associated with the capture device the composite data set to the location on the network;***

186. Fuller's digital video camera 100 transmits with a transmitter associated with the capture device the metadata extracted from the digital audio and video, the collateral metadata, and the video content ("*composite data set*") to "a host computer or other media processing device" ("*transmitting with a transmitter associated with the capture device the composite data set to the location on the network*"). Fuller, 8:1-15. 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." Fuller, 1:45-65, 8:1-15.

187. 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 "*transmitting with a transmitter associated with the capture device the composite data set to the location on the network."*

[10.2] ***wherein a system disposed at the location on a network operates to receive with a receiver the transmitted composite data set from the transmitter associated with the capture device as a received set of captured information, provide a***

database, and store in the database the received composite data set

188. 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 metadata extracted from the digital audio and video, the collateral metadata, and the video content ("transmitted composite data set from the transmitter associated with the capture device") as a received set of captured information ("a system disposed at the location on a network operates to receive with a receiver the transmitted composite data set from the transmitter associated with the capture device as a received set of captured information"). Fuller, 1:45-65, 8:1-15.

189. 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." Fuller, 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" ("*provide a database, and store in the database the received composite data set*") Fuller, 1:45-65.

190. Thus, because Fuller's host computer receives a set of captured information 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 operates to receive with a receiver the transmitted composite data set from the transmitter associated with the capture device as a received set of captured information, provide a database, and store in the database the received composite data set.*"

9. Claim 11

[11.1] ***The method of claim 10, wherein the step of capturing with 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.***

191. 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 step of capturing with 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*."

10. Claim 12

[12.1] ***The method of claim 11, wherein the step of transmitting transmits the***

received composite data set to the location on the network after at least the stop event associated with the processing of the captured external audio information.

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

193. **Second**, as discussed above at [9.1], Fuller describes that audio and video may be combined to form "audiovisual content." Fuller, 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." Fuller, 8:1-15. Fuller thus renders obvious "*the step of transmitting transmits the received composite data set to the location on the network after at least the stop event associated with the processing of the captured external audio information*."

11. Claim 13

[13.0] ***A method for capturing image and audio information for storage, comprising the steps of:***

194. *See* [6.0].

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

195. *See* [6.2].

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

196. *See* [6.3].

[13.3] ***converting with a first data converter the external audio information from the microphone;***

197. *See* [6.4].

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

198. *See* [6.5].

[13.5] ***capturing with a capture device, as captured data, location information and time information associated with at least the capture of the image and***

199. *See* [6.8].

[13.6] ***storing the captured data as stored captured data;***

200. *See* [6.9].

[13.7] ***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,***

201. *See* [6.6].

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

202. *See* [6.7].

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

203. *See* [6.10]. 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]). Fuller, 7:31-8:15.

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

204. *See* [6.11].

[13.11] ***associating the text and image recognition context tags with the digital image and with the stored captured data; and***

205. *See* [6.12].

[13.12] ***storing in the internal storage the digital image in association with the text and image recognition context tags in addition to the stored captured data.***

206. *See* [6.13].

12. Claim 14

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

207. *See* [6.5].

13. Claim 15

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

208. *See* [7.1].

14. Claim 16

[16.1] ***The method of claim 13, wherein the step of capturing captures with the***

camera the image from the image source at an instant in time.

209. *See* [8.1].

15. Claim 17

[17.1] ***The method of claim 13, wherein the step of converting with 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.***

210. *See* [9.1].

VIII. CONCLUSION

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

Dated: 8/26/25

Nathaniel Polish, Ph.D.