UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Canon Inc., Canon U.S.A., Inc., and Axis Communications AB

Petitioner

v.

Avigilon Fortress Corporation,

Patent Owner

Case: Unassigned

U.S. Patent No. 7,932,923
Issue Date: April 26, 2011

Title: Video Surveillance System Employing Video Primitives

DECLARATION OF JOHN R. GRINDON, D.Sc.

TABLE OF CONTENTS

I, Dr. John R. Grindon, declare as follows:

1. I make this Declaration based upon my own personal knowledge and, if called upon to testify, would testify competently to the matters stated herein.

2. I have been asked by Canon Inc., Canon U.S.A., Inc. and Axis Communications AB (collectively “Petitioner”) to provide technical assistance in connection with the Petition for *inter partes* review of U.S. Patent No. 7,932,923 (“the ’923 patent”). This Declaration, which I understand will be filed with the Petition, is a statement of certain opinions I have formed on issues related to the patentability of claims 1-41 of the ’923 patent.

3. It is my opinion that each of claims 1-41 of the ‘923 Patent is unpatentable in view of the identified prior art. My opinions are stated in detail in this Declaration, with reference to the exhibits attached to the Petition and additional exhibits attached to my Declaration.

4. I am being paid at my customary rate of $500 per hour for my time spent in study, preparation, and testifying in this matter. I am also being reimbursed for reasonable and customary expenses associated with my work. No portion of my compensation is dependent or otherwise contingent upon the specifics of my testimony or the results of this matter. My compensation is not contingent upon the outcome of this *inter partes* review.

I. BACKGROUND AND QUALIFICATIONS

5. My name is John R. Grindon. I reside in Hazelwood, Missouri. I have been retained by Petitioner as an expert witness for this matter. I am competent to testify regarding the opinions set forth in this expert Declaration. I reserve the right to provide further opinions if requested at a later time.

6. My qualifications are summarized here and are presented further in my curriculum vitae, which is attached as Exhibit 1006. I received a Bachelor of Science (B.S.) degree in Electrical Engineering, with First Honors, from the University of Missouri at Rolla, now the Missouri University of Science and Technology. I received a Master of Science (S.M.) degree in Electrical Engineering from the Massachusetts Institute of Technology, and a Doctor of Science (D.Sc.) degree in Electrical Engineering from Washington University in St. Louis.

7. During my university studies, I was awarded the Westinghouse Achievement Scholarship. I was a Hughes Masters Fellow at M.I.T. My doctoral research at Washington University was in the field of signal processing.

8. I have more than 40 years of experience including research, analysis, design and development of electronic systems and software for acquiring, processing, analyzing, extracting information from, and communicating signals

and images. This work includes technology related to that disclosed in the '923 patent. I have experience in both hardware and software for this technology.

9. For more than 20 years, I have provided independent consulting, including in the areas of technology related to the '923 patent. From 1987 until 1990, I served as Executive Vice President and Director of Research at the former Cencit, Inc. Cencit was an engineering research and development company that created systems for sensing and digitizing the shape of three-dimensional ("3D") objects through processing of images. The systems generated 3D data about an object based on data gathered from video image sequences that included the object. The 3D data was extracted from the sequences of video images acquired by video cameras viewing the objects. These systems employed processors, communications and control circuits, memory storage and retrieval capabilities, and software algorithms for identifying image primitives, object extraction, location, and 3D surface mapping. Several U.S. and foreign patents were awarded. For more information about the technology developed at Cencit, see for example U.S. Patent No. 4,846,577, filed in April 1987, issued in July 1989.

10. The systems developed by Cencit were implemented for a variety of uses. For example, one system was used for craniofacial trauma research by the Washington University School of Medicine. Another system was used by the United States Army for research into leak proof fitting of gas masks for biological

warfare. Other applications included the 3D modeling of Space Shuttle astronaut's hands for custom fitting of gloves for use with spacesuits, and scanning of the heads of brain tumor patients for precisely positioning implanted radioactive seeds.

11. Later, I also helped to develop a 3D shape digitization system for the Textile/Clothing Technology Corporation ("TC-Squared") of Cary, North Carolina. This system extracted 3D data of a human body from sequences of images, for the original purpose of fabricating custom-fit clothing. Using a different technology than the Cencit system, but also based upon multiple video cameras, the TC-Squared system employed sinusoidal patterns of light projected onto the subject. A first sequence of coarse sinusoidal patterns was used to generate a sparse set of 3D contours. These 3D contours were unambiguously identified. A second sequence, of fine sinusoidal patterns, was then projected which produces high-resolution 3D contours, but with ambiguities for phase angle multiples of 360°. The 3D contour identifications from the coarse patterns were then used to unambiguously identify, or map, the high-resolution 3D contours that were generated from the fine grating projections. Multiple 3D contours were created using the multiple cameras, each covering a portion of the subject, and these contours were then combined into a uniform composite 3D map of the shape of the object. This system is described in further detail for example in a patent by

myself and several co-inventors, U.S. Patent No. 6,373,963, filed in February 1999, issued in April 2002.

12. My role in the development of the TC-Squared system included among other things development of a design concept, providing technical guidance to the engineering team, analyzing and predicting performance parametrically during the design process, defining system parameters and specifications, development of processing algorithms, and development of camera and system calibration methods and software.

13. During my 25 years at McDonnell Douglas Corporation, now Boeing, from 1962-1987, I started with the title of Engineer and progressed through various positions of increasing responsibility to the position of Branch Chief, Electronics. Among other things, my work there included digital image processing research and development for autonomous cruise missile guidance. This work was based on employing on-board digital video cameras and radar to sense terrain and targets, and processing of the imagery and radar data to derive information, such as shape of the terrain surface and identification of targets from the video images. This work also employed correlating images acquired by an on-board video camera with stored reference images. I also developed algorithms for missile-borne laser radar systems (LIDAR) for processing sequences of depth images to detect objects for

the purpose of enabling low-flying cruise missiles to avoid collisions, for example with power wires in their path.

14. Among other work in image processing at McDonnell Douglas, I led an image processing research and development team, where we developed algorithms for target recognition and classification. This work included extracting primitive features from images and development of classifier algorithms to use the extracted features for identifying and locating objects in the images. I also managed a team to develop algorithms for an automatic target classification and recognition system using on-board infrared ("IR") cameras. I also developed a new class of image processing algorithms for autonomous cruise missile location and guidance using IR cameras to acquire video image sequences of the scenes ahead of the missile, correlating the scene images with stored reference data for vehicle position tracking.

15. I have been qualified as a technical expert for several litigation and patent review matters involving video image acquisition, image processing, image analysis, video surveillance, and information extraction. My expert analysis and testimony has involved patents related to digital video imaging and image processing systems, including those for sensing and measuring the shape and location of objects. My curriculum vitae, attached as Exhibit 1036 to the Petition,

includes a listing of matters of public record for which I have served as an expert and offered testimony at deposition or trial in at least the last five years.

II. MATERIALS CONSIDERED

16. I understand that the Petition and this *inter partes* review proceeding involves the ’923 patent, which is owned by Avigilon.

17. I further understand that the ‘923 patent issued from U.S. Application No. 12/569,116. This is reflected on the face of the ‘923 patent (Ex. 1001).

18. The face of the ‘923 patent also indicates that the earliest priority date it may claim is October 24, 2000, based on the filing date of U.S. Patent No. 6,954,498. I express no opinion as to whether the claims of the ‘923 patent are or are not entitled to this October 24, 2000, priority date, but reserve the right to later do so.

19. I understand that prior art to the ‘923 patent includes all patents and printed publications that were published more than one year before the earliest possible priority date claimed by the ‘923 patent, which as noted above is October 24, 2000. Further, I understand that each of the prior art references discussed in this Declaration was published more than one year before October 24, 2000 and therefore qualifies as “§ 102(b)” prior art.

20. I have been asked to consider whether certain prior art references (namely, *Kellogg*, *Dimitrova* and *Brill*) disclose or suggest the features recited in

the claims of the ‘923 patent. I have also been asked to consider the state of the art in the 2000 timeframe (i.e., before the earliest possible priority date of October 24, 2000), and to compare the claims of the ‘923 patent to the prior art available at that time.

21. I have reviewed and considered the following documents, among others identified herein, in connection with my analysis of the ‘923 patent and this Declaration:

- U.S. Patent No. 7,932,923 to Lipton et al. ("the ’923 patent," Ex. 1001);
- Prosecution history of U.S. Application No. 12/569,116, which led to the issuance of the ‘923 patent (Ex. 1002);
- *Ex Parte* Reexamination Control No. 90/012,876, the *ex parte* reexamination of the ‘923 patent;
- *Inter Partes* Reexamination Control No. 95/001,914, the *inter partes* reexamination of the ‘923 patent;
- U.S. Patent No. 7,868,912 ("the ’912 patent"), which is related to the ‘923 patent;
- *Ex Parte* Reexamination Control No. 90/012,878, the *ex parte* reexamination of the ’912 patent;
- *Inter Partes* Reexamination Control No. 95/001,912, the *inter partes* reexamination of the ’912 patent;

- "Visual Memory" by Christopher James Kellogg (Ex. 1003, "*Kellogg*"), which I understand was published in September 1993 and is prior art to the '923 patent; and
- "Motion Recovery for Video Content Classification" by N. Dimitrova et al. ("*Dimitrova*"), which I understand was published in October 1995 and is prior art to the '923 patent
- "Event Recognition and Reliability Improvements for the Autonomous Video Surveillance System" by Frank Brill et al. (Ex. 1004, "*Brill*"), which I understand was published in December 1998 and is prior art to the '923 patent.

22. My opinions are also based on my experience in video image acquisition, image processing, image analysis, video surveillance, and information extraction.

III. LEGAL STANDARDS FOR PATENTABILITY

23. I have been asked to provide my opinions as to whether the identified prior art (namely, *Kellogg*, *Dimitrova* and *Brill*) teach or render obvious claims 1-41 of the '923 patent from the perspective of a person of ordinary skill in the art ("POSITA") in the 2000 timeframe.

24. I am an electrical engineer and innovator by training and profession. The opinions I express in this Declaration involve the application of my technical

knowledge and experience to the evaluation of certain prior art with respect to the ‘923 patent.

25. My opinions are also formed by my understanding of the relevant law. I am not an attorney. Therefore, for purposes of this Declaration, I have been informed about certain aspects of the law as it relates to my opinions.

26. I understand that for an invention claimed in a patent to be found patentable, it must be (among other things) new and not obvious based on what was known before the invention was made.

27. I understand that the information that is used to evaluate whether an invention was new and not obvious when made is generally referred to as “prior art.” I understand that the prior art includes patents and printed publications that existed before the earliest filing date of the patent (which I have been informed is called the “effective filing date”). I also understand that a patent or published patent application is prior art if it was filed before the effective filing date of the claimed invention and that a printed publication is prior art if it was publicly available before the effective filing date. As noted above, I understand that prior art relative to the ‘923 patent includes all patents or printed publications that were published more than one year before October 24, 2000.

28. I understand that in this *inter partes* review proceeding, the claims must be given their broadest reasonable interpretation consistent with the patent

specification, as understood by a person of ordinary skill in the art. After the claims are construed in this manner, they are then compared to the prior art.

29. I understand that a dependent claim is a patent claim that refers back to another patent claim. A dependent claim, as I understand it, includes all of the limitations of the claim to which it refers.

30. I understand that in this *inter partes* review proceeding, the information that may be evaluated is limited to patents and printed publications. My analysis, which is set out in detail below, compares the claims to printed publications that I understand are prior art to the claims.

A. Anticipation Analysis: 35 U.S.C. § 102

31. I understand that a person cannot obtain a patent on an invention if the prior art included that invention.

32. If an invention is not new, then the invention has been "anticipated" by the prior art.

33. A claim is "anticipated" by the prior art if each and every limitation of the claim is disclosed, expressly or inherently, in a single item of prior art, from which a person of ordinary skill in the art could practice the invention.

34. I have applied the above standards in my evaluation of whether claims 1-41 of the '923 patent are anticipated in light of the prior art.

B. Obviousness Analysis: 35 U.S.C. § 103

35. It is my understanding that a claim is unpatentable if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I also understand that an obviousness analysis takes into account the scope and content of the prior art, the differences between the claimed subject matter and the prior art, and the level of ordinary skill in the art at the time of the invention.

36. I understand that a claimed invention is not patentable if it would have been "obvious" to a person of ordinary skill in the field of the invention at the time the invention was made. This means that even if all the requirements of a claim are not found in a single prior art reference, the claim is not patentable if the differences between the subject matter in the prior art and the subject matter in the claim would have been obvious to a person of ordinary skill in the art at the time the application was filed.

37. In determining the scope and content of the prior art, it is my understanding that a reference is considered relevant prior art to the '923 patent if it falls within the field of the inventor's endeavor. In addition, a reference is prior art if it is reasonably pertinent to the particular problem with which the inventor was involved. A reference is reasonably pertinent if it logically would have commended itself to an inventor's attention in considering his problem. Thus, if a

reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an obviousness analysis.

38. In my opinion, the prior art references identified in the Petition and addressed in this Declaration (including, *Kellogg, Dimitrova* and *Brill*) seek to solve the same types of problems as the ’923 patent and logically would have commended themselves to an inventor’s attention considering the problems of the ’923 patent. In particular, the ’923 patent and the identified prior art concern techniques and systems for detecting objects in a video, determining attributes of the objects, identifying events by applying a user rule or event definition to the attributes, and reporting and/or taking other action in response to inferring the occurrence of such events.

39. I understand that a determination as to whether a claim would have been obvious should be based on four factors (though not necessarily in the following order): (i) the level of ordinary skill in the art at the time the application was filed; (ii) the scope and content of the prior art; (iii) the differences between the claimed invention and the prior art; and (iv) any objective factors indicating obviousness or non-obviousness that may exist in a particular case.

40. It is my understanding that an obviousness analysis cannot be based on hindsight, but must be done using the perspective of a person of ordinary skill in the relevant art as of the effective filing date of the patent claim.

41. I understand the objective factors indicating obviousness or non-obviousness may include: commercial success of products covered by the patent claims; a long-felt but unaddressed need for the invention; failed attempts by others to make the invention; copying of the invention by others in the field; expressions of surprise by experts and those skilled in the art at the making of the invention; and the patentee having proceeded contrary to the accepted wisdom of the prior art. I also understand that any of this evidence must be specifically connected to the invention rather than associated with the prior art or with marketing or other efforts to promote an invention.

42. I understand that the teachings of two or more prior art references may be combined in the manner disclosed in the claim if such a combination would have been obvious to one having ordinary skill in the art at the effective filing date. In determining whether a combination would have been obvious, I understand that the following exemplary rationales may support a conclusion of obviousness:

- Combining prior art elements according to known methods to yield predictable results;
- Simple substitution of one known element for another to obtain predictable results;
- Use of a known technique to improve similar devices (methods, or products) in the same way;

- Applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- "Obvious to try"—choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;
- Known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations are predictable to a person of ordinary skill in the art; and
- Some teaching, suggestion, or motivation in the prior art that would have led a person of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

43. I understand that the obviousness analysis need not seek out precise teachings directed to the specific subject matter of the challenged claim, but instead can take account of the ordinary innovation and experimentation in the relevant field that does no more than yield predictable results.

44. I understand that, in assessing whether there was an apparent reason to modify or combine known elements as claimed, it may be necessary to look to interrelated teachings of multiple patents or prior art references, the effects of commercial demands, and the background knowledge of a person of ordinary skill in the art. I further understand that any motivation that would have applied to a person of ordinary skill in the art, including motivation from common sense or

derived from the problem to be solved, is sufficient to explain why references would have been combined.

45. I understand that modifications and combinations suggested by common sense are important and should be considered. Common sense suggests that familiar items can have obvious uses beyond the particular application being described in a prior art reference, that if something can be done once it would be obvious to do it multiple times, and that in many cases a person of ordinary skill in the art can fit the teachings of multiple patents together in an obvious manner to address a particular problem. The prior art does not need to be directed to solving the same problem that is addressed in the patent.

46. I understand that a person of ordinary skill in the art is also a person of ordinary creativity. In many fields, it may be that there is little discussion of obvious techniques, modifications, and combinations, and it may be the case that market demand, rather than scientific research or literature, will drive a new design. When there is market pressure or design need to solve a particular problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has a good reason to employ the known options. If this leads to the expected success, then it is likely the product of ordinary skill and common sense as opposed to patentable innovation. I understand that if a combination was obvious to try, that may show that it was obvious and therefore unpatentable. That

a particular combination of prior art elements was obvious to try suggests that the combination was obvious even if no one made the combination.

IV. THE ’923 PATENT

A. Overview

47. As part of my analysis, I read and considered the ’923 patent and related prosecution history before the Patent Office. The following overview is not meant to describe my full understanding of the ’923 patent and prosecution history, but rather to highlight the general aspects of the ’923 patent and prosecution history.

48. According to the Abstract, the ’923 patent relates to a “video surveillance” system for extracting video “primitives” (including “attributes” of a detected object) from a video and determining the occurrence of an event based on the primitives. Ex. 1001, Abstract.

49. The ’923 patent states that its detection and storage of primitives is advantageous over conventional systems that search raw video data because it will result in a reduction of the amount of data to be stored or processed as compared to raw video. *Id*., 2:29-33. Accordingly, the disclosed surveillance system extracts “primitives” or “attributes” from the video. *Id*., Abstract. Examples of video primitives include a classification, a size, a shape, a color, a texture, a position, a velocity, a speed, an internal motion, a motion, a salient motion, scene change, etc.

Id., 7:8-12. Events are defined in terms of attributes using the event discriminators. *Id.*, 4:63-5:5. Using these event discriminators, the ’923 patent describes that the “video content can be reanalyzed … in a relatively short time because only the video primitives are reviewed and because the video source is not reprocessed.” *Id.*, 14:63-66.

50. The disclosure of ’923 patent is generally functional and does not describe the technical aspects of the disclosed system in any detail. For example, while the patent states that objects are detected (*id.* 9:30-44, Fig. 5) and attributes of the objects are detected (*id.* 10:49-52, Fig. 5), the ’923 patent does not purport to have invented the technology to perform these functions. To the contrary, the ’923 patent disclosure emphasizes that it is agnostic to the object detection technology used by explaining that “any” relevant algorithm can be used in the disclosed system (*id.* 9:34, 9:39; 10:27-29; 10:42-47 (briefly discussing prior art techniques for classification); 10:49-51 (“video primitives are identified using the information from blocks **51-56** [admitted prior art object detection, tracking, and classification] and additional processing as necessary”).

51. As a further example of the high-level functional nature of the ’923 patent disclosure, the patent merely lists various examples of relevant attributes/primitives without describing how to identify those attributes. *Id.* 7:8-12 (examples of attributes); 7:13-16 (examples of the classification attributes); 7:17-

23 (examples of size attributes); 7:24-28 (examples of color attributes); 7:29-31 (examples of texture attributes); 7:32-36 (examples of internal motion attributes); 7:37-41 (examples of motion attributes).

52. Motion attributes, in particular, include object activities or motion such as entering, exiting, stopping, appearing, disappearing or moving from one place to another. *Id.*, 3:30-33, 7:8-10; 7:37-46, 7:63-67. A trajectory of an object is an example of an attribute. *Id.*, 7:8-10; 7:47-49. An internal motion such as "person having swinging arms and legs" is another example of an attribute. *Id.*, 7:34-36. A scene change attribute includes an object entering a scene and becoming stationary or an object changing position in a scene. *Id.*, 7:66-8:4. Salient motion and scene change attributes can be detected over a period of time. *Id.*, 7:8-10; 7:37-46, 63-67.

53. While example attributes are disclosed, the '923 patent does not purport to have invented the detection of these attributes. And, because all the attributes are described as examples, the '923 patent does not specify that any particular set of attributes are necessary to practice the invention. This lack of disclosure is relevant because a POSITA would understand that a particular system's ability to define events will naturally be limited by the scope of attributes that is detected by the system. The '923 patent, however, does not teach a

POSITA how to select attributes in order to achieve broad flexibility in the events that can be identified.

54. After disclosing the exemplary attributes that might be of interest, the ’923 patent describes the Fig. 3 exemplary process of tasking the system and describes additional exemplary attributes. *See id.* Fig. 2 Block 23 (tasking); 4:30-31 (Fig. 3 discloses an embodiment of tasking); 8:16-19 (object types are identified); 8:20-31 (spatial areas of interest identified); 8:32-36 (temporal attributes identified); 8:37-49 (response is *optionally* identified); 8:50-58 (discriminators describing interactions are identified).

55. The ’923 patent next describes the Fig. 5 exemplary process of detecting attributes and notes that the particular attributes to be identified “are the same as” the attributes discussed with regard to the system *tasking* as described in reference to block 23 of Fig. 2. *Id.* 10:51-52. As shown in Fig. 5, below, blocks 51-57 generally relate to object detection, tracking and classification methods. But they do not contain enough granularity to explain what the universe of attributes a POSITA should detect is and how to achieve that scope. As discussed above, the ’923 patent simply does not purport to have invented methods for detecting these attributes and, thus, it does not provide much disclosure to support them.

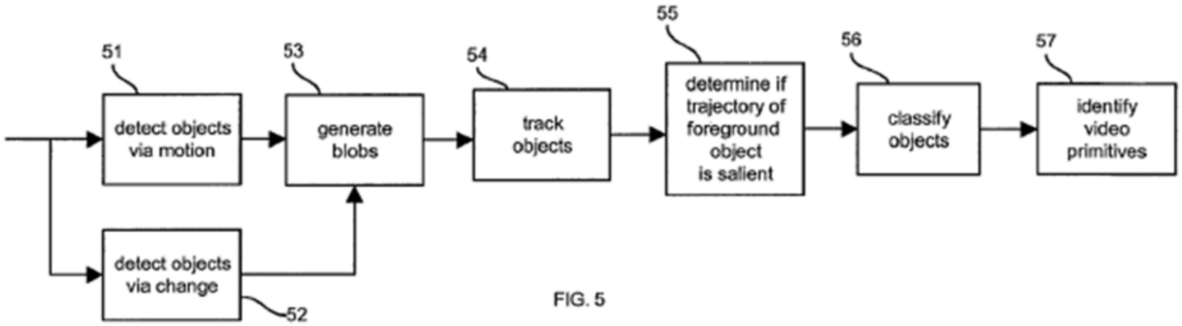


FIG. 5

56. Indeed, the ’923 patent merely describes that “video primitives are identified using the information from blocks 51-56 [of Fig. 5] and additional processing as necessary.” *Id.*, 10:49-51. The patent does not describe what the additional processing is or how a POSITA would determine what is necessary.

57. Figures 4 and 9 both show exemplary flow diagrams depicting the operation of the disclosed surveillance system. *Id.* Fig. 4, Fig. 9, 4:30-31; 40-41. Each of these figures includes archiving the primitives/attributes or accessing an archive as a step in the processes, see blocks 43 and 92 respectively. The ’923 patent, however, provides no detail regarding how this data is stored. Specifically, all the ’923 patent states: “In block **43**, the video primitives from block **42** are archived. The video primitives can be archived in the computer-readable medium **13** or another computer-readable medium. Along with the video primitives, associated frames or video imagery can be archived.” *Id.* 10:58-62. Given this disclosure a POSITA would understand that video primitives or attributes could be stored according to any known techniques in the art, such as in a database.

58. Events of an object in the video are detected by applying a user-defined "event discriminator" to the attributes, without the need to reprocess the original video. *Id*., 4:63-5:1, 5:17-19. Event discriminators can be defined using objects, spatial attributes, and temporal attributes. *Id.*

59. As with the other aspects of the '923 patent disclosure, the patent describes examples of events that can be detected with the system, but it does not describe how the exemplary events are implemented in the context of the disclosed system. For example, the patent describes a "loitering event" as a "person" object in the "automatic teller machine" (ATM) space for "longer than 15 minutes" and between 10:00 p.m. and 6:00 a.m. *Id*., 5:1-5. A system implementing this event would need to translate this natural language definition into a query that can be applied against the attributes detected by the system. For example, this event ***could*** be translated into a search for a combination of five primitives: (1) a "person" object, (2) an "enter" (3) in the "ATM" video space location, (4) with no "exit" for "longer than 15 minutes," and (5) in a time range "between 10:00 p.m. and 6:00 a.m." *Id*. Other collections of attributes could also be used to define this same event. Nevertheless, this collection of attributes would then need to be translated into a query syntax to be applied to the detected and stored attributes. The '923 patent does not disclose details for this process.

60. Once an event is identified, any number of responses or actions may optionally be taken. *Id.* 8:37. For example, a report may be displayed to the user. *Id.*, 11:45-54. Reports may contain information such as the video sequence containing the event, data on the number of event occurrences, identification of the positions in the scene of the event, data on the times when the events occurred, etc. *Id.* According to the '923 patent, the type of response the system will take is determined during the tasking of the system. In particular "responses, if any, are identified for each event discriminator in block 34" of Figure 3. *Id.*, 11:14-16, Fig. 3.

61. The system of the '923 patent can be implemented with standard, existing computer hardware and video surveillance equipment. *Id.*, 3:47-62, 5:60-6:2, 6:48-50. A computer system is connected to video equipment. *Id.*, 6:46-48. The computer system obtains source video, performs object detection and attribute extraction and stores the attributes in computer-readable medium. *Id.*, 9:23-10:66, Figure 5. Using standard I/O devices, a user can task the computer system with event discriminators. *Id.*, 6:24-26, 6:64-67. Then, the computer system identifies event occurrences based on the extracted attributes. *Id.*, 10:66-11:1.

B. Prosecution History

1. Prosecution

62. During prosecution, claims were found obvious over U.S. Patent Nos. 7,653,635 and 6,721,454. Ex. 1002, 160-163. To distinguish the prior art the patentee amended the claims to recite selecting of a new user rule after detecting the plurality of attributes (or storing the detected attributes). *Id.*, 118-129.

63. After a first examiner interview, the claims were further narrowed to recite that the plurality of attributes include at least one of a physical attribute and a temporal attribute. *Id.*, 98-110. After a second interview, the claims were further narrowed to require video "from a single camera" to distinguish US Patent Publication 2003/0023612. *Id.*, 78-90. Applicant argued that unlike the prior art system, which used two cameras to track players and six additional cameras for ball tracking, the claimed system analyzes multiple attributes in a video from a single camera. *Id.*, 91-93. The claims were then allowed and matured into claims 1-41. *Id.*, 65-71.

2. *Inter partes* reexamination

64. The '923 patent was challenged in an *inter partes* reexamination, Control No. 95/001,914, by Bosch Security Systems, Inc. Ex. 1008, 1. The Patent Office instituted the reexamination, and rejected all claims on the following six separate grounds:

- Claims 1-7, 9-13, and 15-28 are anticipated by *Courtney*
- Claim 14 is obvious over *Courtney*
- Claims 1-7, 9-13 and 15-28 are anticipated by *Shotton*
- Claim 14 is obvious over *Shotton*
- Claims 8 and 29-41 are obvious over *Shotton* and the *Brill '835*[1] patent
- Claims 1-41 are obvious over *Courtney* and the *Brill '835* patent.

Ex. 1009, ¶5; Ex. 1010, ¶¶3-10.

65. The patentee filed an amendment and response on August 27, 2012. Ex. 1011. Before any action by the examiner, the patentee and Bosch settled their dispute, and the Patent Office granted the patentee's petition to terminate the reexamination on February 13, 2013. Ex. 1012, 4-5.

3. *Ex parte* reexamination

66. On May 23, 2013, the '923 patent was anonymously challenged in an *ex parte* reexamination, Control No. 90/012,876. Ex. 1013, 1. The Patent Office instituted the reexamination, and rejected all claims on 10 different grounds:

- Claims 1-9, 13, 15-18, 20-30, 34 and 36-40 are anticipated by *Day-I*
- Claims 1-7, 9-13 and 15-28 are anticipated by *Courtney*
- Claims 1-7, 9-13 and 15-28 are anticipated by *Shotton*
- Claims 14 and 25 are obvious over *Day-I*
- Claims 10, 19, 31, and 41 are obvious over *Day-I* in view of the *Brill '835* patent

---

[1] U.S. Patent No. 6,628,835 ("*Brill* '835 patent"). *Brill* (Ex. 1004), used as a prior art reference in this IPR, states that it is an advancement on prior work and it was published later than *Brill* '835 patent.

- Claims 11, 12, 32 and 33 are obvious over *Day-I* in view of *Day-II*
- Claim 14 is obvious over *Courtney*
- Claim 14 is obvious over *Shotton*
- Claims 8 and 29-41 are obvious over *Shotton* in view of the *Brill '835* patent
- Claims 1-41 are obvious over *Courtney '584* in view of the *Brill '835* patent.

Ex. 1014; Ex. 1015, ¶¶10-21.

67. The patentee filed an amendment replacing claims 1-41 with new claims 42-81. Ex. 1016, 2-21. Specifically, claims 55-58 differed from all other claims in that they included a new limitation requiring that the step of applying the "new user rule" comprised applying the rule to ***only*** the plurality of detected attributes. *Id.*, 8-10.

68. Relying on the above amendment, the patentee distinguished the claims from the prior art reference *Day-I* by arguing:

> New claims 55-58 require application of the user rule "only" to the detected attributes. In contrast, the queries of Day-I are not applied to the attributes stored in the VSDG alone but are applied to object-oriented abstractions. *See* Day-I at p. 405, § 3.1 ("For video data, a user can use combination [sic] of various abstractions to construct his/her view of the video data. The important feature of this hierarchy, and in general for any object-oriented abstractions [sic], is that each terminal node is either a CTO [Conceptual Temporal Object, a CSO [Conceptual Spatial Object], or a PO [Physical Object]. Any complex video query is expressed as a function of these terminal nodes and processing of such query requires execution of some CTO and CSO over specified PO's.") & p. 407, § 3.2.3 ("all these queries generally require processing of various combination [sic] of object hierarchy (shown in Figure 5)").

Id., 78-79. The patentee also made similar arguments to distinguish the claims from the related prior art reference *Day-II*. *Id.*, 79

69. The patentee relied on, among other things, the '923 patent's disclosure that "[t]he video content can be reanalyzed with the additional embodiment in a relatively short time because *only* the video primitives are reviewed and because the video source is not reprocessed." Ex. 1033, ¶ 30 (emphasis in the original); Ex. 1018, 7; *see also* Ex. 1040, ¶148; Ex. 1001, 14:64-67. But in that context the word "only" is used for excluding reprocessing video source as opposed to excluding something other than video primitives, such as abstractions. Neither this citation, nor the rest of patentee's citations in the reexamination (Ex. 1033, ¶ 30; Ex. 1001, 5:19-23, 5:31-32, 10:63-11:1, 14:58-60, 14:66-15:4) supports excluding application of the user rule to abstractions or explains what an abstraction of a primitive is.

70. Although not cited by the patentee, the '923 patent discloses, "[i]n block **31**, one or more objects types of interest are identified in terms of video primitives or *abstractions* thereof." Ex. 1001, 8:16-17 (emphasis added). This statement, which is only related to the object type attributes (*e.g.*, an object, a person, a red object) does not explain what an abstraction is. *Id.*, 8:18-19. There is no further disclosure explaining what an abstraction is or how the distinctions between attributes and abstractions would be embodied in a working system.

71. The Patent Office issued a final rejection on all claims, except claims 55-58. Ex. 1017, ¶19. In response to the final rejection, the patentee narrowed each originally issued independent claim to include the new limitation found in allowable claims 55-58 and canceled claims 42-81. Ex. 1018, 2-6, 9. Claims 1-41 were then found patentable (Ex. 1019, ¶3), and a reexamination certificate issued on May 21, 2014 (Ex. 1020).

4. Prior Proceedings in the Related Patents

72. The ’923 patent is part of a family of patents that ultimately claim priority to U.S. Application No. 09/694,712, filed October 24, 2000. Related U.S. Patent Nos. 7,868,912 (“the ’912 patent”) (Ex. 1034) and 8,564,661 (“the ’661 patent”) (Ex. 1035) are also part of the same family, but they are not in the direct priority chain of the ’923 patent.

73. The ’912 patent was also involved in the Bosch dispute and challenged in an *inter partes* reexamination. Ex. 1024. Like the ’923 patent Bosch reexamination, before any action by the examiner, the patentee settled its dispute with the requestor, and the Patent Office granted the patentee’s petition to terminate the reexamination. Ex. 1025.

74. Subsequently, the ’912 patent was also challenged in an *ex parte* reexamination. Ex. 1026. The Patent Office instituted the reexamination, and rejected all claims on multiple grounds. Ex. 1027; Ex. 1028. After back and forth

arguments with the Patent Office (Ex. 1029; Ex. 1030; Ex. 1031), a reexamination certificate issued on June 25, 2014 (Ex. 1032).

75. The ’661 patent, on the other hand, was not challenged in any reexaminations. But it has been challenged in two IPRs, namely IPR2018-00138 and IPR2018-00140 (hereinafter “the Related IPRs”). The Related IPRs were filed on October 31, 2017, instituted on June 1, 2018, and are currently pending.

C. Level of Ordinary Skill in the Art

76. I have been informed that the factors defining the level of ordinary skill in the art include the types of problems encountered in the art; the prior art solutions to those problems; the rapidity with which innovations are made; the sophistication of the technology; and the educational level of active workers in the field.

77. As indicated above, the earliest possible priority date of the ’923 patent is October 24, 2000. At that time, a person of ordinary skill in the art (“POSITA”) would have (i) a Bachelor of Science degree in electrical engineering, computer engineering, or computer science, with approximately two years of experience or research related to video processing and/or surveillance systems or (ii) equivalent training and work experience in computer engineering and video processing and/or surveillance systems.

78. It is also my opinion that a POSITA would be knowledgeable and familiar with the video processing and information extraction concepts and techniques recited in the claims of the ’923 patent, as they were well known in 2000, as demonstrated above.

79. As of 2000, I was at least a person of ordinary skill in the art.

80. In this Declaration, and for all of my opinions stated herein, I have applied the knowledge of a POSITA as of 2000.

V. CLAIM CONSTRUCTION

81. I have been asked to explain how a person of ordinary skill in the art before the earliest priority date of the ’923 patent would have understood certain claim terms of the ’923 patent.

82. I have been advised that in *inter partes* review proceedings before the U.S. Patent and Trademark Office (“Patent Office”), an unexpired patent claim’s terms receive their broadest reasonable construction in light of the specification, as would be understood by one of skill in the art. One exception to applying this construction is when the patent applicant acts as his or her own lexicographer by defining a claim term in a way that diverges from the broadest reasonable construction.

83. I understand that under current rules the claims herein are interpreted according to their broadest reasonable construction in light of the specification (BRI).

84. I understand that U.S. patent law includes special requirements for claim limitations that are written in "means-plus-function" language. As explained below, it is my opinion that certain limitations of one or more claims of the '923 patent are written in such language because they recite "means for" and do not claim sufficient structure to perform the recited functions. I understand claim limitations with "means-plus-function" language are evaluated using the following framework:

a. The limitation must be evaluated to determine if it is a "means-plus-function" limitation and therefore should be interpreted under what I understand is the version of section 112, paragraph 6 of the U.S. patent statutes. I understand that a claim limitation is a "means-plus-function" limitation if it recites functional language and does not include a sufficient description of structure, material or acts to perform the stated function.

b. If the limitation is determined to be a "means-plus-function" limitation subject to section 112, paragraph 6, then the scope of that limitation includes the corresponding structure, material, or

> acts described in the patent specification for performing its stated function. To construe the claim, it is necessary to review the specification to identify each embodiment clearly linked or associated with the stated function.

85. My opinions regarding the structure disclosed in the ’923 patent specification corresponding to the means-plus-function claim limitations are set forth below.

86. To the extent that additional information becomes available, I reserve the right to continue my investigation and study, which may include a review of documents and information that are presented in this proceeding. I reserve the right to have my opinions provided here to this review only, and I reserve the right to continue my investigation and study such that I am not bound by my opinions in other proceedings.

87. For claim terms other than those reviewed below, I have applied the plain and ordinary meaning of the claim terms under a broadest reasonable construction, as understood to a person of ordinary skill in the art in the October 2000 timeframe. At this time, explicit construction of other claim terms is not necessary. In addition, I understand that the ’923 patent includes a “DEFINITIONS” section in columns 3 and 4 that defines a number of terms, including some of which that appear in the claims. Accordingly, I understand that

the broadest reasonable interpretation of the claim terms defined in the "DEFINITIONS" will be at least as broad as these definitions in the patent. I apply these definitions to these terms where they are used in the claims except as explained below for claim construction purposes.

A. "attributes of the object" (claims 1-7, 9-19, 22-28, 30-41); "attributes of each of the detected first and second objects" (claims 8, 29); "attributes of the detected object" (claims 20, 21)

88. The '923 patent describes "attributes" or "primitives" as "observable" characteristics of an object. Ex. 1001, 7:6-7. The '923 patent explains that object "attributes" or "primitives" can represent an object's physical characteristics, such as "a classification; a size; a shape; a color; a texture; a position." *Id*., 7:8-12. Classifications include a real-world identification of the object, such as identifying it as a person, animal, car, etc. Size, shape, color and texture are self-explanatory. The position of the object is the objects location either in the video image or translated into real-world location.

89. Attributes can also represent an object's temporal characteristics in a video. Examples include "every 15 minutes," "between 9:00 p.m. to 6:30 a.m.," "less than 5 minutes," "longer than 30 seconds," "over the weekend," and "within 20 minutes of." *Id.*, 8:32-36

90. The '923 patent also explains that attributes can represent actions or activities of the object, such as "a speed, an internal motion, a motion…" *Id.*, 7:8-

12. Examples of these activity attributes include: "appearance of an object, disappearance of an object, a vertical movement of an object, a horizontal movement of an object…." *Id.*, 7:37-40. Examples of an internal motion include a "person having swinging arms and leg." *Id.*, 7:32-36. Some attributes are detected over a period of time, *e.g.*, moving from one place to another, trajectory, an object entering a scene and becoming stationary, etc. *Id.*, 7:42-46, 63-67. These typically known as spatio-temporal attributes in the field.

91. Accordingly, "attributes" as used in the claims should be construed as "characteristics associated with an object."

B. "new user rule" (claims 1-41)

92. Each independent claim recites a "new user rule." For example, claim 1 recites "identifying an event of the object … by applying the new user rule to the plurality of detected attributes." The term is not defined in the claims, but dependent claims 2 and 23 limit "selecting a new user rule" to "selecting a subset of the plurality of attributes for analysis."

93. The patent specification does not use the term "user rule" or "new user rule" either. But in the claim the event is identified by applying the new user rule. The corresponding feature in the patent that does this is the "event discriminator." The patent states that "an operator is provided with maximum flexibility in configuring the system by using event discriminators." Ex. 1001,

4:63-64. The patent further states that object attributes/primitives are detected and archived, and then "event occurrences are extracted from the video primitives using event discriminators." *Id.*, 10:58-64, Fig. 4; 10:66-11:1 ("[t]he event discriminators are used to filter the video primitives to determine if any event occurrences occurred.").

94. Event discriminators define events by using attributes of objects to specify a particular event. *Id.* at 4:64-5:1, 8:59-9:12. The attributes specified can be one or more spatial or temporal attributes and they can also describe the interaction of multiple objects. *Id.* For example, the '923 patent defines a "loitering event" at an ATM as follows: (1) a "person" object, (2) in the "ATM" video space, (3) for "longer than 15 minutes," and (4) "between 10:00 p.m. and 6:00 a.m." *Id.*, 5:1-5. Thus, with these details specified the system can look for objects that satisfy all these criteria, when that occurs an event is detected. In this way, the event discriminators in the patent perform the same function of allowing the user to identify events as the claimed "new user rule."

95. In view of the above, a POSITA reading the '923 patent would understand that the "new user rule" refers a specified combination of a set of attributes for identifying an event.

96. Accordingly, "new user rule" should be construed to mean "a specified combination of a set of attributes for identifying an event."

C. Independence-based limitations (claims 1-41)

97. During the reexamination of the ’923 patent and in the Related IPRs, Patent Owner has argued that the claimed “independence-based” limitations distinguish its invention over the prior art. These independence-based limitations correspond to the basic concept of the invention, which is that it is advantageous to identify objects in a video and record their attributes because those attributes can later be searched more efficiently than processing and searching the video imagery. Ex. 1001, 15:17-23, 5:8-10. These limitations also attempt to convey that the event does not need to be preset, which would distinguish over prior art that was designed to specifically address certain known events.

98. In order to make these points, the Patent Owner argues that the “independence-based” limitations have the following three requirements (1) identifying an event that refers to one or more objects engaged in an activity by analyzing the detected attributes; (2) the detected attributes are independent of the event identified; and (3) the identified event is not one of the detected attributes. IPR2018-00138, Paper No. 11, 25 (September 4, 2018); Ex. 1016, 37-39.

99. I address each of these concepts below using Patent Owner’s numbering scheme. However, point (2) is addressed last for ease of analysis.

1. Independence Argument (1): requires identifying an event that refers to one or more objects engaged in an activity by analyzing the detected attributes

100. The claim limitations corresponding to Argument (1) are:

- Claim 1: "identifying an event of the object …by applying the new user rule to the plurality of detected attributes"
- Claim 8: "identifying an event …by applying the new user rule to the plurality of detected attributes"
- Claim 9: "means for identifying an event of the object …by applying a selected new user rule to the plurality of attributes"
- Claim 20: "the new user rule providing an analysis of a combination of the attributes to detect an event"
- Claim 22: "identifying an event of the object …by applying the new user rule to the plurality of detected attributes"
- Claim 29: "identifying an event …by applying the new user rule to the plurality of detected attributes"
- Claim 30: "means for identifying an event …by applying a selected new user rule to the plurality of attributes stored in memory"

101. Patent Owner also asserts that these limitations require identifying an event by analyzing the detected attributes. IPR2018-00138, Paper No. 11, 16-17 (September 4, 2018); Ex. 1016, 37-38. This claim language, however, is broader and quite generic in that it merely states that the event is identified "by ***applying*** the new user rule to the plurality of detected attributes." Ex. 1001, Reexamination Certificate 1:45-47 (emphasis added). This raises an issue regarding whether or not this claim language limits the process used to determine correspondence

between the detected attributes and the attributes specifying the event in the new user rule.

102. The ’923 patent specification does not use the terms apply/applying to describe identifying an event, nor does it disclose *analyzing* attributes. *See, e.g., id.*, 4:64-5:1, 6:63-64, 7:2-6, 10:63-64, 10:66-7:1. Thus, a POSITA would have no reason to limit the ways one might apply the user rule to the detected attributes. As such, a POSITA would understand that the claimed “applying” could encompass any mechanism for analyzing the detected attributes to determine if they satisfy the user rule criteria. This could be accomplished by querying a database.

103. In my opinion, this claim language should cover any way of determining correspondence between a user rule and a set of detected attributes.

2. Independence Argument (3): the identified event is not one of the detected attributes

104. Argument (3) corresponds to claim language in each of the independent claims and in the ’923 patent this concept is stated as part of the same phrase as the Argument (1) claim language.

105. The claim limitations corresponding to Argument (3) are highlighted:

- Claim 1: “identifying an event of the object ***that is not one of the detected attributes of the object*** by applying the new user rule to the plurality of detected attributes”
- Claim 8: “identifying an event ***that is not one of the detected attributes of the first and second objects*** by applying the new user rule to the plurality of detected attributes”

- Claim 9: "means for identifying an event of the object ***that is not one of the detected attributes of the object*** by applying a selected new user rule to the plurality of attributes"
- Claim 20: "the new user rule providing an analysis of a combination of the attributes to detect an event ***that is not one of the detected attributes***"
- Claim 22: "identifying an event of the object ***that is not one of the detected attributes of the object*** by applying the new user rule to the plurality of detected attributes"
- Claim 29: "identifying an event ***that is not one of the detected attributes of the first and second objects*** by applying the new user rule to the plurality of detected attributes"
- Claim 30: "means for identifying an event …by applying a selected new user rule to the plurality of attributes stored in memory…***the event not being one of the detected attributes***"

106. This argument addresses the Patent Owner's interest in avoiding prior art that simply identifies one detected attribute. Essentially, if a system only identifies events that are merely single detected attributes, it would be indistinguishable from a system that is preset to only detect certain predetermined events. This would be contrary to the patent's stated goal of allowing a user to later define a new event based on the detected attributes. *See, e.g.,* Ex. 1001, 4:63-5:1, 7:2-6.

107. For example, a single activity attribute such as "appear," "enter," or "exit" would meet the patent's basic definition of an event, *i.e.*, "an object engaged on an activity." For example, the user rule "identify that an object appears" needs

to search for only one attribute "appear" and it then identifies any object that meets the "appear" criteria. This single attribute test would meet the '923 patent's definition of an "event" because it would identify objects engaged in the "appear" activity. Thus, this argument redefines "events" in the context of the '923 patent claims to clarify that in the context of the claims an event must be defined in terms of more than a single attribute. Accordingly, a claimed event is not merely "one or more objects engaged in an activity" as recited in the specification's definition section. Ex. 1001, 3:44-45.

108. The patentee specifically addressed this issue in the reexamination of the '923 patent:

> the specification of the '923 patent discloses ***some identified events that are the same as a detected attribute***. *See* `707 application at ¶ 98 ("an object appears").

Ex. 1016, 38 (emphasis added). As discussed above, the example the patentee provided, "an object appears," identifies ***any*** object that "appears." Thus, it identifies every occurrence of the "appear" attribute.

109. Relying on the Argument (3) claim language, the patentee argued that the ***claimed*** event is more than a single attribute.

> the specification of the '923 patent also discloses events that are not detected attributes. *See, e.g., id* at ¶ 98 ("a person appears; a red object moves faster than 10 m/s"); & ¶ 99 ("two objects come together; a person exits a vehicle; a red object moves next to a blue object"). The claims of the '923 patent require

> identification of an event that is not a detected attribute and are silent regarding identification of an event that is a detected attribute. *See* Zeger Dec., ¶ 56.

Ex. 1016 at 39. Here, the patentee identified "a person appears" as an event that is within the scope of the claim because it is not merely a single event attribute. Instead, this event requires two attributes, the "appear" activity attribute plus a "person" object classification attribute. Thus, this illustrates the concept that single activity attributes are not events within the scope of the claim—although they would be events in the context of the patent disclosure—and two attribute events are within the scope of the claim.

110. Given the above admission that the patent disclosure includes both single attribute events and multiple attribute events, as well as the open-ended comprising format of the claims, a POSITA would understand that a prior art reference is not excluded from being relevant merely because it is capable of identifying single attribute events. Instead, a prior art reference is disclaimed if it can ***only*** identify single attribute events.

111. Similarly, the fact that a prior art system records activity attributes does ***not*** provide an adequate basis to distinguish that prior art from this claim limitation. This is relevant to Patent Owner's arguments that a prior art reference merely records pre-defined events. IPR2018-00138, Paper No. 11, 7-10, 52-53

(September 4, 2018); IPR2018-00140, Paper No. 11, 7-11, 40-41, 53-54 (September 4, 2018). Of course, there is nothing wrong with a prior art system recording activity attributes, such as enter, exit, appear or disappear. Indeed, the '923 patent describes systems that record exactly those attributes. *See*, *e.g.*, Ex. 1001, 3:30-33, 7:37-39, 8:60, 8:63, 8:67. The pertinent question for this limitation and the claims is whether a system can combine these attributes with other attributes to create a new event that is not merely one of the detected attributes.

112. Further illustrating this point is the fact that the '923 patent claims recite this limitation as "***identifying an event*** of the object that is not one of the detected attributes of the object." Thus, this limitation specifically limits "the identified event," which is the event specified by the user rule with a "***plurality*** of detected attributes." Ex. 1001, Reexamination Certificate 1:44-47 (emphasis added). Accordingly, this limitation clarifies that ***the claimed event specified by applying the user rule*** cannot be a single activity attribute.

113. Based on the above, in my opinion the construction of Argument (3) limitation requires that the ***claimed*** user defined "event" comprises a minimum of two attributes.

3. Independence Argument (2): the detected attributes are "independent" of the event identified or detected

114. Argument (2) asserts that the detected attributes are "independent" of the event that is identified or detected. As an initial matter, it is important to understand that the "event identified" is the event specified by the user rule and therefore the requirements of Argument (3) above must apply to this event and it cannot not merely be one of the detected attributes. Again, for the reasons stated above, this means that the when evaluating the prior art one must determine whether the attributes that are detected by the prior art are independent of events that are specified by user rules.

115. The '923 patent claim language corresponding to the Argument (3) limitations requires:

- "the plurality of attributes that are detected are independent of which event is identified" (claims 1, 8, 22, 29)
- "the attributes to be detected are independent of the event to be detected" (claim 20)
- "for identifying the event independent of when the attributes are stored in memory" (claim 9, 30)

116. As the claims require, the ***claimed*** independence is between the detected attributes and the event that is defined by the user rule and identified. This is shown in the context of Claim 1 as follows:

> **[1.3]** selecting a ***new user rule*** after detecting the plurality of attributes; and

> **[1.4]** after detecting the plurality of attributes and after selecting the new user rule, ***identifying an event of the object*** that is not one of the detected attributes of the object ***by applying the user rule*** to the plurality of detected attributes, wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes;
>
> **[1.5]** wherein the plurality of attributes that are detected are ***independent of which event is identified***,

117. As stated in [1.5], the claimed attributes are required to be independent of "which event is identified." That identified event has antecedent basis in [1.4] where it is specified as the event defined by the claimed new user rule. This claim language should, therefore, be understood to require that the detection of attributes is independent from, *i.e.*, not affected by, the user rule. In the context of the claims, the new user rule tasks the system when it is selected and used to identify a particular identified event. *See*, Ex. 1001, 6:64-67 ("Without tasking, the video surveillance system operates by detecting and archiving video primitives and associated video imagery without taking any action").

118. In my opinion, a POSITA would understand that this limitation is met if a user rule can define an event that applies to an arbitrary mix of detected attributes and the definition of the event by the user rule is not used to alter the selection of attributes that are collected. The same attributes are detected for every user rule regardless of which user rule is selected. Accordingly, the proper

construction of this limitation merely requires that ***the event detection process does not alter the attribute detection process***.

119. In the Related IPRs, Patent Owner has argued that this limitation should be construed to mean "the plurality of detected attributes are detected ***without regard to or knowledge*** of ***a predefined/predetermined list of events of interest***" amongst which at least one event is identified. IPR2018-00138, Paper No. 11, 29 (September 4, 2018); IPR2018-00140, Paper No. 11, 30 (September 4, 2018) (emphasis added). I disagree with this proposed construction because it is not supported by the claim language, specification or prosecution history of the '923 patent. I do not see any discussion in the intrinsic record of a "predefined/determined list of events." Thus, I cannot conclude what constitutes a predefined/determined list of events. I have reviewed Dr. Bovik's declaration in support of Patent Owner's Response to the '661 patent IPR. I note that Dr. Bovik identified no support in the intrinsic record for this list concept. Ex. 1038, ¶¶ 50, 56, 60. This language can be used to sow confusion to the extent that Patent Owner confuses activity attributes and events, as has been done. *Id.*, ¶ 80 (equating an index of objects and object attributes to a list of events); *id.*, ¶ 92 (equating a schema that describes logical relationships between attributes with an event list). The claim only requires independence from the event identified by the

user rule, not a list, and the claim construction should not add this unstated concept.

120. Similarly, the requirement of Patent Owner's construction that detection occurs "without regard to or knowledge of" the list is also unsupported by the specification, claims and intrinsic evidence as a whole. Putting aside the lack of a list, as discussed above, the intrinsic evidence provides no basis to understand what constitutes "regard to" or "knowledge of" an identified event. This limitation certainly should not require getting into the mind of the engineer that designed the system to understand what hypothetical events she might have imagined when considering relevant attributes.

121. The intrinsic evidence does not explain who or what's knowledge or regard is relevant to the disclosed system; as a result, this proposed construction is hopelessly vague. The Patent Owner cites the following text as support for this limitation:

> In block 23 of FIG. 2, the video surveillance system is tasked. Tasking occurs after calibration in block 22 and is optional. Tasking the video surveillance system involves specifying one or more event discriminators. Without tasking, the video surveillance system operates by detecting and archiving video primitives and associated video imagery without taking any action, as in block 45 in FIG. 4.

See Ex. 1001, 6:61-67; IPR2018-00138, Paper No. 7, 14-15 (March 2, 2018); IPR2018-00140, Paper No. 7, 14-15 (March 2, 2018).

122. This passage does not disclose what it means to have "regard for" the events and who or what is prohibited from having "knowledge" of the events. It simply states that in the disclosed embodiment tasking is optional. In the claimed system, however, tasking is required because the claim specifically requires selecting a new user rule. These limitations proposed by the Patent Owner are not found in or required by the claim language.

123. Indeed, contrary to Patent Owner's arguments, some embodiments of the claimed system have knowledge of the tasked events when the attributes are detected. For example, the '923 patent discloses and claims the real-time detection of events, which involves the system being tasked with a user rule and subsequently detecting attributes to determine whether the event defined by the user rule occurred. Ex. 1001, 2:48-50, 9:14-19, 9:25-26, 17:5-9, 19:31-35. Obviously, once a new user rule is defined, the system operating in a real-time mode has knowledge of that ultimate event to be determined, and all of this occurs ***before*** the attributes are even detected.

124. Another point of confusion created by the Patent Owner's construction is that it is misused to attempt to distinguish certain prior art, like *Courtney*, by arguing those references merely detect ***pre-defined*** "events."

IPR2018-00140, Paper No. 11, 11-13 (September 4, 2018). The chief problem with the Patent Owner's argument is that it does not apply the definition of event specified in the claims according to the Argument (3) claim language, *i.e.*, that the claimed event is not a single attribute.

125. To overcome rejections over *Courtney* in the reexamination, the patentee argued that *Courtney* was distinguishable because when it analyzed video it merely detected "events." Ex. 1016 at 49. The patentee's argument was largely based on the mere fact that *Courtney* uses the word "event" to refer to ***single activity attributes***, such as appear, disappear, enter, and exit. Ex. 1021, 10:52-61. While it is true that *Courtney* uses the word "event," the '923 patent detects these exact same things and calls them attributes. Ex. 1001, 3:30-33.

126. However, as discussed above, the patentee admitted that the '923 patent uses these terms in such a way that a single activity, like appear, could be considered an event or an activity attribute according to the '923 patent specification. Ex. 1016, 38-39. Indeed, the patentee clarified that only a plurality of detected attributes is considered for claim interpretation purposes. *Id.* Thus, using this clarification, the mere fact that *Courtney* happened to label certain detected activities like "appear" as an "event" (Ex. 1021, 10:52-61) rather than an "attribute" is nothing more than a semantic difference. In my opinion, this is not a patentably significant technical distinction compared to the '923 patent claims.

127. The patentee tried to further distinguish *Courtney* by arguing that referencing "events" by location or time does not meet the "identifying events" requirement in the claim. Ex. 1016, 49. The patentee argued that this distinguished *Courtney* because *Courtney* merely discloses a user rule that queries its single activity attribute "events" and adds a time and location attribute, *e.g.*, an object appears at a certain time and location. *Id.*

128. The patentee's support for this limitation is the definition of the "event" in the '923 patent, which states that events "may be referenced with respect to time or location." Ex. 1001, 3:44-46. The first problem with this argument is that it ignores the patentee's admission that the '923 patent refers to single activity attributes as events. And it ignores the fact that the '923 patent defines a number of events that include location (position) attributes and time attributes. *See, e.g.*, Ex. 1001, 9:2 ("an object ***appears*** at ***10:00 p.m.***") (emphasis added); 5:1-5. The '923 patent disclosure does not suggest that these attributes should not be considered part of the event description.

129. Moreover, there is ***no claim*** limitation that expressly prohibits the event from including location and time attributes. The mere use of the word "independent" certainly does not dictate this concept. Indeed, to the contrary, the claims specifically require the event definition that includes temporal attributes. Ex. 1001, Reexamination Certificate 1:37-39; *see also id.* 16:22-25. Thus, it seems

incorrect to prohibit location and time attributes from being considered attributes as part of an event definition as Patent Owner appears to argue.

130. To the extent Patent Owner maintains its argument that it somehow disclaimed considering the time or location attributes of events, this argument should at most only exclude the claim from covering a system that can only identify an "event" as a single pre-defined activity attribute, plus a time attribute, and/or a location attribute. This, however, would not distinguish over the prior art presented in this petition which is capable of much more sophisticated *ad hoc* event definitions.

131. In view of the above, the proper construction of this limitation merely requires that the event detection process does not alter the attribute detection process. Should the Patent Owner further argue that indexing an "event" by time or location was somehow disclaimed, that should only prohibit the claim from covering events defined as only a single predefined activity attribute plus a time, and/or a location attribute.

D. "wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes" (claims 1-19, 22-29); "wherein the analysis of the combination of the attributes to detect the event comprises analyzing only the combination of the attributes" (claims 20-21); "wherein the applying the selected new user rule to the plurality of attributes Stored in memory comprises

applying the selected new user rule to only the plurality of attributes stored in memory" (claims 30-41)

132. As discussed above, in the '923 reexamination, the patentee added the limitation "wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes" in order to distinguish the *Day-I* and *Day-II* references. Ex. 1018, 3-6, 9. The patentee argued that this language distinguished the *Day* references because "the queries of *Day-I* are not applied to the attributes stored in the VSDG alone but are applied to object-oriented abstractions." Ex. 1016, 78-79; Ex. 1031, ¶133. Patentee further argued that § 3.2.3 of *Day-I* stated that "all these queries generally require processing of various combination [sic] of object hierarchy (shown in Figure 5)." Ex. 1016, 78-79; *see* Ex. 1022, Fig. 5 below. Patentee made similar arguments as to *Day-II*. Ex. 1016, 79.

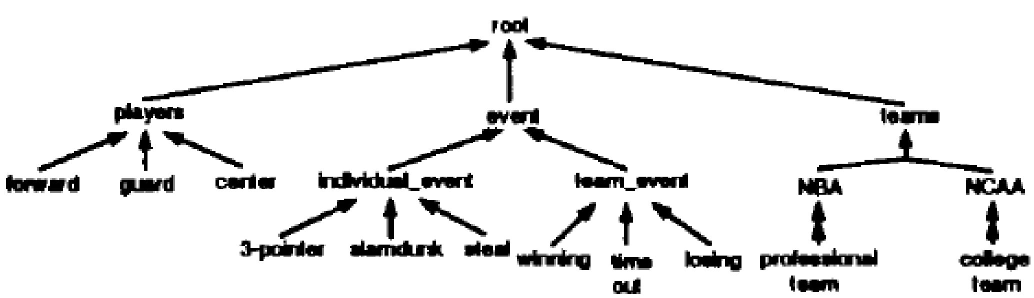


Figure 5: Fan's view

133. The patentee's reexamination arguments to overcome the *Day* prior art are not well founded. The problem is that the '923 patent's limited disclosure

does not comport to this claim language and the fine distinctions over the *Day* references as argued by the patentee. The alleged specification "support" that the patentee cites for this limitation is minimal. *See* Section IV(B)(3). As discussed, the patentee cited to the '923 patent's disclosure that "[t]he video content can be reanalyzed with the additional embodiment in a relatively short time because only the video primitives are reviewed and because the video source is not reprocessed." Ex. 1033, ¶ 30 (emphasis in the original); Ex. 1018, 7; *see also* Ex. 1001, 14:64-67. This does not support the patentee's argument that the patent describes only using primitives and not abstractions because the word "only" functions to exclude reprocessing the source video. Indeed, the patent expressly discloses that the system can process abstractions. Ex. 1001, 8:16-17. There is simply no disclosure explaining what using primitives versus abstractions means or how one would embody a working system that did not process abstractions, as the patentee argued to overcome the *Day* references.

134. What does it mean for a system to only identify objects by primitives or to only identify objects by abstractions of primitives? The '923 patent provides no explanation.[2] Thus, the patent does not support detailed distinctions over the

[2] Note, for example, the '923 patent considers "classification" as an example of an attribute. Ex. 1001, 7:6-9. The patent further gives example classifications of

prior art on this basis. Moreover, while this claim language and the distinction over the *Day* references are directed to "applying" the new user rule to the detected attributes, the corresponding patent disclosure relates to defining the event, not applying the event to the attributes. Ex. 1001, 8:16-19.

135. Nevertheless, the patentee essentially argued that the *Day* references were distinguishable because they ***always*** require the processing of abstractions of the attributes and not the attributes themselves when applying a user rule. *Id.* This is because *Day* always references an object hierarchy structure, *e.g.*, a tree structure, in performing a search. *See* Ex. 1022, Fig. 5.

136. Patentee argued that a search of the *Day* database, *i.e.*, an application, requires traversing the tree through the higher-level abstractions. It is hard to

---

vehicle and police car. *Id.* 7:13-15. Thus, these are clearly both attributes of an object according to the patent. However, given the fact that a police car is a sub-type of vehicles, one could in some circumstances consider vehicle to be an abstraction of police car. The '923 patent provides no disclosure that would permits a POSITA to distinguish between attributes and abstractions in this disclosure.

understand how the ’923 patent supports this distinction because it has ***no*** disclosure of how the attribute data is stored, nor does it disclose how the attribute data is searched during the application process. Nevertheless, this issue is not particularly relevant to the present issues because the prior art at issue here does not require that its attribute data be stored in a tree structure like Fig. 5 of *Day-I*.

137. One further important thing to note about this limitation is that the applying step “***comprises***” applying the new user rule to only the detected attributes. The use of the open-ended term “comprises” encompasses systems that employ searches of object-oriented abstractions, so long as they can also employ searches of only the attributes themselves. Thus, applying the patentee’s argument in view of the *Day* references would only exclude prior art systems that ***always*** require event searches to process abstractions rather than just the attributes themselves, as occurs when data is stored in the tree structure shown in Fig. 5 of *Day-I*. Ex. 1022, Fig. 5.

138. In sum, this limitation should at most only limit claims as excluding coverage of systems that always reference an object hierarchy structure such as a tree structure that requires traversal of abstractions to apply the user rule.

E. Means-plus-function elements (claims 9-19, 30-41)

139. It is my understanding that the following elements of independent claims 9 and 30 should be construed under pre-AIA 35 U.S.C. § 112(6) because

they recite "means for" without reciting sufficient structure to perform the recited functions.

1. "means for detecting an object in a video from a single camera" [9.1]; "means for detecting first and second objects in a video from a single camera" [30.1]

140. A POSITA would understand the '923 patent specification recites two corresponding structures for these limitations, both of which are conventional motion and/or change detection algorithms that are utilized on a computer system or equivalent video processing system to detect objects: 1) Ex.1001, 5:61-64, 9:33-35 (Fig. 5 (51)); and 2) 5:61-64, 9:39-41 (Fig. 5 (52)). .

2. "means for detecting a plurality of attributes of the object by analyzing the video from said single camera, the plurality of attributes including at least a physical attribute and a temporal attribute, each attribute representing a characteristic of the detected object" [9.2]; "means for detecting a plurality of attributes of the object by analyzing the video from said single camera, each attribute representing a characteristic of the respective detected object" [30.2]

141. A POSITA would understand that the '923 patent specification recites a corresponding structure for these limitations, which is a conventional computer-vision algorithm that is utilized on a computer system or equivalent video processing system for detecting attributes. Ex.1001, 5:61-64, 10:49-51 (Fig. 5 (57)). Examples of a physical attribute of an object include size, shape, color and texture, etc. Ex. 1001, 7:8-9. Examples of a temporal attribute of an object include

"every 15 minutes," "between 9:00 p.m. to 6:30 a.m.," "less than 5 minutes," etc. Ex. 1001, 8:32-36.

3. "means for selecting a new user rule after the plurality of detected attributes are stored in memory" [9.4]

142. A POSITA would understand the '923 patent specification recites a corresponding structure for this limitation, which is a conventional user interface software and "I/O devices" such as "a keyboard; a mouse; a stylus; a monitor," utilized on a computer system or equivalent video processing system: Ex.1001, 5:61-64; 6:23-28, 6:61-64 (Fig. 2 (23)), 15:11 (Fig. 9 (91))..

4. "means for identifying an event…" [9.5]-[9.8]; "means for identifying an event of the first object interacting with the second object…" [30.4]-[30.6]

143. A POSITA would understand that the '923 patent specification recites a corresponding structure for these limitations, which is a conventional query mechanism utilized on a computer system or equivalent video processing system to detect an event: Ex.1001, 5:61-64, 10:63-64 (Fig. 1(44)), 10:66-11:1, 9:14-17, 14:57-60, 15:7-10, 14:63-66, 8:65-67. Any functional limitations containing the "new user rule" limitation or the independence-based limitations should be construed according to the claim constructions set forth in Sections V(B) and V(C), respectively. A POSITA would readily understand that the other functional

limitations are performed by the query mechanism and no further construction is necessary.

VI. OVERVIEW OF PRIOR ART

A. State of the Prior Art and Applicant-Admitted Prior Art

144. In my opinion, and based on my personal experience, the state of the relevant art was well developed prior to October 2000. Techniques to detect objects in a video and identify attributes of those objects were known by persons of ordinary skill in the art, as well as methods for searching and identifying events based on those attributes. The hardware, software, and algorithms for implementing such features in a video surveillance system were also known. All of this is acknowledged and verified by the ’923 patent and the prior art made of record during the prosecution of the ’923 patent.

145. As admitted in the ’923 patent, “the video surveillance system of the invention draws on well-known computer vision techniques from the public domain.” Ex. 1001, 5:6-10.

146. As admitted in the ’923 patent, detecting objects in video was known in the art. *See*, *e.g.*, *id.*, 1:27-2:25. Also, as acknowledged in the patent, the particular object detection and tracking scheme disclosed in the patent “can be replaced with any detection and tracking scheme, as is known to those of ordinary skill.” *Id.*, 10:27-29.

147. As admitted in the ’923 patent, detecting various types of video attributes was known in the art. *See, e.g.*, *id.*, 10:37-41 (citations to prior art for determination of trajectories and salient motion); 10:44-47 (object “[c]lassification can be performed by a number of techniques”); 10:54-56 (indicating that size can be obtained from known calibration techniques).

148. The ’923 patent discloses conventional hardware for implementing a video surveillance system. *Id.*, 3:47-62, 5:60-6:2, Figs. 1. In my opinion, a POSITA reading the ’923 patent would be familiar with the identified hardware and conclude that none of the hardware components identified in the ’923 patent is novel.

149. The use of databases for storing video metadata was known in the art since at least the 1990s. *See*, *e.g.*, Ex. 1003, 11-12 and the video databases referenced therein.

150. Creating and applying queries to data in databases were well-known in the art. For example, U.S. Patent No. 5,918,225, which was filed in 1997 and issued in 1999, discloses a SQL-based database system. In describing the state of the art (*see* 1:51-54), the ’225 patent acknowledges that “[t]he general construction and operation of a database management system is known in the art. *See*, *e.g.*, Date, C., *An Introduction to Database Systems*, Volume I and II, Addison Wesley, 1990.” The ’225 patent goes on to discuss (*see* 2:23-24) traditional systems for

"finding one or a few records which meet a given query condition," before disclosing its improvements on the known systems.

B. Kellogg

151. *Kellogg* discloses a visual memory system that tracks objects in a video, detects and stores information about the objects, and responds to user queries specifying events concerning those objects. Ex. 1003, Abstract, 69. *Kellogg*'s system can handle multiple objects in a scene, such as multiple people or a person and a vehicle. *Id.*, 77, 79.

152. Figure 4-1 (reproduced below) provides a diagram of the system, including components for "image processing" to detect objects and data about objects in a video (*i.e.*, attributes), a "visual memory" where attributes are stored, and a "graphical query interface" to enable a user to define query specifications (*i.e.*, user rules or event definitions) for identifying events based on the attributes. *Id.*

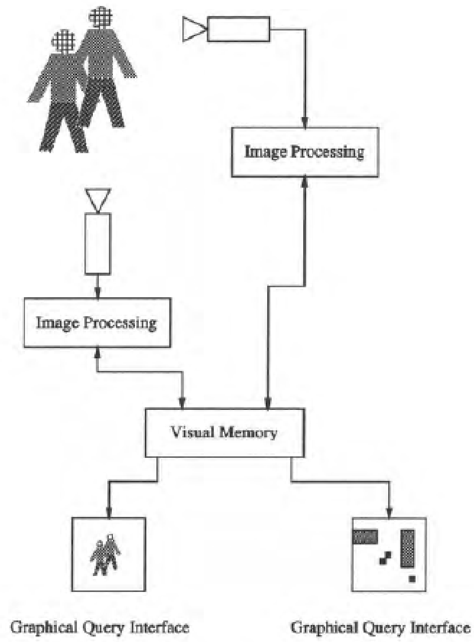


Figure 4-1: Scene monitoring prototype

153. "The object-oriented database on which the visual memory [system] builds provides basic support for object storage and retrieval." *Id.*, 50. *Kellogg*'s visual memory system "provides a powerful and expressive [querying] mechanism for retrieving information" that is "designed to meet a wide variety of retrieval needs, providing flexibility in specifying objects of interest." *Id.*, 53.

154. *Kellogg's* system detects and stores a variety of object attributes. The system may detect an object's area, duration, trajectory (*id.*, 22); the object's class, centroid, orientation, bounding box (*id.*, 24); relative special attributes ("west" or "near") (*id.*, 30); time stamps or intervals for the valid times when the object existed (*id.*, 36-37, 52); relative temporal attributes ("before" or "after") (*id.*, 40-41); and attributes specific to certain object types (height of a person) (*id.*, 71).

155. *Kellogg* uses database queries to detect events. *Id.*, 53 (using SQL-based object query language). *Kellogg* explains that rules "are implemented as part of the query language to allow the query language to optimize object retrieval." *Id.*, 54. *Kellogg* describes that its query mechanisms provide "great flexibility in spatial and temporal query specification," allowing queries to "include[] spatial or temporal keywords," "spatial or temporal object[s]," or even "the result[s] of another query." *Id.*

156. *Kellogg* discloses querying events that are not merely a detected attribute. For example, *Kellogg* discloses using three classes of attributes: spatial, temporal, or spatiotemporal. *Id.*, 54-64. Users can define events of their choosing, such as determining when a person (an object classification) does or does not intersect an arbitrary user-drawn rectangle ("enter area activity") by comparing the persons' position. *Id.*, 54-58, *see also,* 58 (finding when objects are in the scene during the same time). More complex spatiotemporal searches are also disclosed, which allow, *e.g.*, identifying an "approach" event that finds "all objects that came within 3 units of a given object on its trajectory [*i.e.*, a given moving object] during a certain set of valid times." *Id.*, 63. Other examples of *ad hoc* user rules disclosed by *Kellogg* include, "[w]atch for anything that comes within 3 feet of that button" (*id.*, 68), or determine if "anybody [came] into the room between 12:00 and 1:00" (*id.*, 80).

157. Indeed, *Kellogg* discloses the same "loiter" event as the '923 patent. *Kellogg* discloses that the user could "specify that an alarm should fire only if an object remains in a region for a suspicious amount of time." *Id.*, 80; *see* Ex. 1001, 5:1-5. *Kellogg*'s storing of basic attributes allows it to detect a myriad of other events, such as a "theft" query defined as a piece of "furniture" leaving an area, or a "bad dog" event defined as an "animal" entering a prohibited area.

C. Brill

158. *Brill* discloses an Autonomous Video System ("AVS") system for moving object detection and event recognition. Ex. 1004, 4. *Brill*'s AVS system is shown in Figure 1:

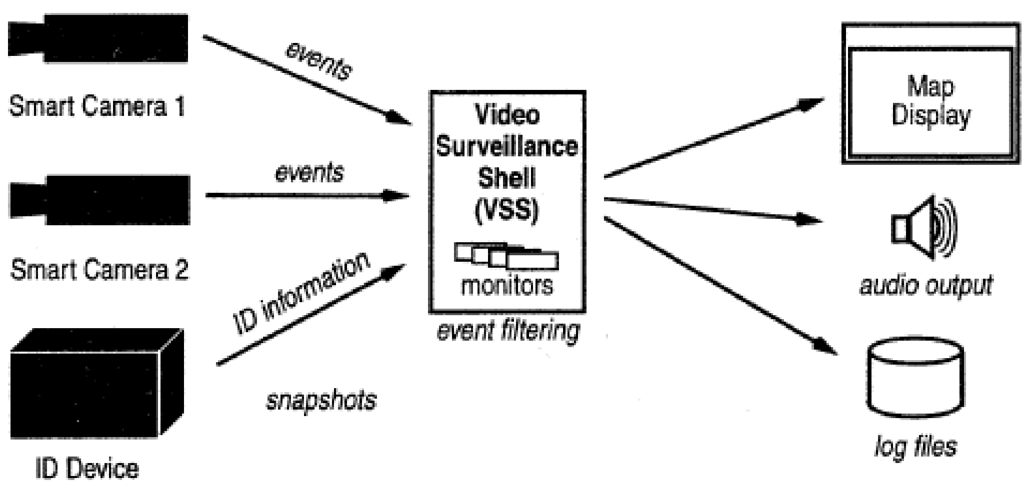


Figure 1: AVS system diagram

159. As illustrated in Figure 1, smart cameras process and analyze video to send activity attributes and other attributes (*e.g.*, object ID (Ex. 1004, 11), object type, location, and time stamp (*id.*, 13)) to the Video Surveillance Shell ("VSS").

At the VSS, user-entered "monitors" are applied to detect events according to the '923 patent. *Id.*, 13.

160. *Brill* teaches a graphical user interface that allows a user to develop and submit queries to identify events. Ex. 1004, 13. The interface allows a user to define events. *Id.*, Figure 11.

Figure 11: Selecting a type of simple event

161. In the example shown in Figure 11, the user defines a rule for a new "Loiter by the door" event, which involves a "person" object engaging in a "loiter" activity. It further requires an "outside the door" region and a duration of "5.0" second. For this user rule, the "days of week" and "time of day" attributes are not analyzed or queried, but a user can use those attributes to further narrow down search. *Brill's* system can be used to infer other new events. For example, an "arrival at work" event could be defined using the Figure 11 interface to select

"person" and "briefcase" objects "outside the door" between "8:50 am and 9:10 am" on "Monday" morning. *Id.* The responses that could be taken when the event is detected include issuing a visual or audible alert (e.g., "flash" or "popup" visual alerts or the "beep" and "voice" audible alerts) or recording data about the event (e.g., the "log" and "plot" action options).

162. *Brill* discloses object detection and tracking methods using background models to distinguish objects of interest from the video background. *Id.*, 6-9. *Brill* also discloses an advanced detection and tracking method based on the probability of an object being at given pixel that allows reliable tracking even after more than one objects overlap in a given scene. *Id.*, 14-18.

D. Dimitrova

163. *Dimitrova* discloses video classification and retrieval systems that enable users to formulate queries and identify events. Ex. 1006, 25-27. In doing so, *Dimitrova*'s system detects and tracks objects in a video scene, and then performs "motion analysis" on video sequences to extract basic activity attributes, *e.g.*, waving, running, walking, strolling or hurrying. *Id.*, 7, 17, 19. Some of the basic activity attributes are not detected in a vacuum, but instead inferred by using certain spatial and/or temporal information. *Id.*, 17. For example, strolling, walking and hurrying are distinguished by using temporal information. *Id.*

164. The system also extracts other attributes that represent the characteristics of the detected objects, *e.g.*, category (*e.g.*, "person" or "pet") (*id.*, 19); relative size (*e.g.*, "big") (*id.*), color (e.g., "brown") (*id.*); parts representing a human figure, *e.g.*, head, torso, arms and legs (*id.*).

165. The information of video sequences is then stored using object attributes ("O"), object motion attributes ("M"), and video attributes ("V"), or an "OMV" triplet. *Id.*, 19-20. The extracted activity attributes—*e.g.*, walking—become part of the object motion attributes (M) that also include other motion attributes such as trajectory, velocity, torsion, etc. *Id.*, 20. Object attributes (O) include convex hull, object skeleton, centroid, texture (*id.*) as well as other attributes that represent the characteristics of the detected objects as discussed above. Video attributes (V) contain video sequence identity and frame information. Ex. 1006, 20.

166. *Dimitrova*'s queries use these collected attributes to identify events that are as simple as "retrieve all the video sequences in which a pet walks and makes a trajectory" to events that are more complex such as "retrieve all the sequences in which a tall person is waving while the president walks." Ex. 1006, 20, 25. In the former query, motion attributes such as an activity attribute "walking" and "trajectory" are combined with an object classification attribute "pet" to identify user defined *ad hoc* events that have a certain trajectory arbitrarily

set by a user in the query. In the latter query, a user creates a rule with multiple activity attributes and object attributes for identifying that which could be called a "parade" event. *Id.* *Dimitrova* identifies a "parade" event by searching for all "tall" "persons" associated with a "waving" activity "while" the "president" is associated with a "walking" activity. *Id.*

E. Motivation to Combine Kellogg and Brill

167. A POSITA would have found it obvious to combine *Kellogg's* visual memory system with the features of *Brill*'s system. As demonstrated by the hundreds of references cited on the face of the '923 patent, the state of the art was quite crowded prior to October 2000. Ex. 1001, References Cited. A POSITA would have been aware of object detection methods, attribute detection methods, and querying mechanisms, like those disclosed in *Kellogg* and *Brill*. A POSITA would have combined elements of *Kellogg* and *Brill* to provide enhancements or achieve particular design objectives, while yielding predictable results.

168. *Kellogg* teaches detecting one or more objects in a video from a single camera. Ex. 1003, 30-31, 77, 79, Figure 3-5. Those objects could be multiple people or a person and vehicle in a single field of view. *Id.*, 56-57, 79. *Kellogg* also contemplates detecting interactions between those objects in general. *Id.*, 65-67; *see also* Section VIII(A)(5)(f).

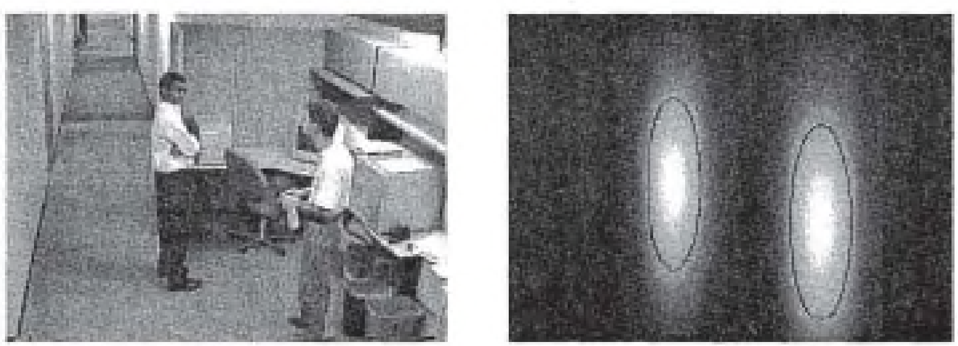
169. *Brill* teaches, among other things, an enhanced event detection platform that reliably handles recognizing interactions of multiple objects, especially two objects where interactions of human-vehicle and human-human are involved. Ex. 1004, 6.

170. *Brill* teaches improvements to object tracking so that tracking is not lost when a person's image overlaps another object, such as a car. Ex. 1004, 6-9. In some systems this situation would cause the objects to appear to merge resulting in the loss of tracking until the person walks away from the other object or the other object moves. *Id.* This problem would have been within the knowledge of a POSITA who employed *Kellogg*'s system to monitor a human and a vehicle, and any interaction thereof. Ex. 1003, 79 (the alarm region monitors both human and vehicle). *Brill* teaches that its "new approach involve[es] additional image differencing…[which] allows objects to be detected and tracked even when their images overlap the image of the car." Ex. 1004, 6. *Brill* specifically teaches a background-model based technique. *Id.*, 6-9. And a POSITA would have been highly motivated to combine the teachings of *Brill* with *Kellogg*'s monitoring system to solve the loss of tracking issue.

171. A POSITA using *Kellogg*'s system to monitor multiple humans and any of their interaction also faced a separate issue. As described in *Brill,* a POSITA knew that it is difficult to monitor movements of one or more people in a

scene because they move unpredictably, may move close to one another, and may occlude each other. Ex. 1004, 14. Indeed, *Kellogg* describes to some extent methods that address these types of uncertainties. Ex. 1003, 31-35. But as *Brill* explains, when two people are in a single scene, it was difficult to maintain the separate tracks of the two people once they merge into a single large region. Ex. 1004, 14. A POSITA employing *Kellogg*'s system would have faced this same issue.

172. *Brill* introduces a new method which maintains an estimate of the size and location of the objects, and that creates a separate image which approximates the probability that the object intersects that pixel location. Ex. 1004, 14-15, Figure 15.



(a) (b)

Figure 15: Probability image for the locations of the people in the scene

173. *Brill* teaches that even after an occlusion, the objects are reliably detected and tracked by relying on non-overlapping areas. *Id.*, 15. A POSITA would have been highly motivated to combine the teachings of *Brill* with *Kellogg*'s

monitoring system to solve the occlusion issue and enhance detection and tracking of multiple moving objects.

174. The motivation to combine *Kellogg* and *Brill* is further evidenced by the fact that they were both related to the AVS systems developed by Texas Instruments. Indeed, at least one prior art publication confirms that the visual memory database of *Kellogg* was combined with the AVS system disclosed in *Brill*. Ex. 1007, ¶¶40-44.

175. In sum, both references provide ample express teaching, suggestion, and motivation to combine *Brill*'s teachings with *Kellogg*'s system. Moreover, as both being related to the AVS systems developed by Texas Instruments, a POSITA would be able to predictably combine the teachings of the two references, without requiring extensive modification to the overall system. *Brill* would be an obvious source because, like *Kellogg*, *Brill* is directed to video surveillance systems using object detection methods, attribute detection methods, and querying mechanisms. *See, e.g.*, Ex. 1003, 24-25, 50, 54, 62-63, 68-69, 71, 77; Ex. 1004, 6-9, 12-14. Similar to *Kellogg's*, *Brill's* system can detect multiple objects and their interactions in a single video scene. *See, e.g.*, Ex. 1003, 62-63; Ex. 1004, 12-13. Given the similarities of their subject matter and teachings, a POSITA would have immediately recognized *Brill*'s advanced multi-object detection and tracking

method would have readily worked in *Kellogg*'s system with a reasonable expectation of success.

F. Motivation to Combine Dimitrova and Brill

176. A POSITA would have found it obvious to combine *Dimitrova's* video classification and retrieval systems with the features of *Brill*'s system. As demonstrated by the hundreds of references cited on the face of the '923 patent, the state of the art was highly advanced prior to October 2000. Ex. 1001, References Cited. A POSITA would have been aware of object detection methods, attribute detection methods, and querying mechanisms, like those disclosed in *Dimitrova* and *Brill*. A POSITA would have combined elements of *Dimitrova* and *Brill* to provide enhancements or achieve particular design objectives, while yielding predictable results.

177. *Dimitrova* teaches detecting and tracking one or more object in a single video scene. Ex. 1006, 3, 11, 25, Figs. 4, 5. This process occurs before motion analysis that leads to extracting and storing activity attributes. *Id.*, 7, 13. The detected objects can consist of moving objects and stationary objects like a moving toy and two stationary cups in Fig. 5, or multiple moving objects like a tall person waving and the president walking. Ex. 1006, 25. *Dimitrova* also contemplates monitoring various types of interactions between those objects. *Id.*, 25, Fig. 5; *see also* Section VIII(C)(5)(g).

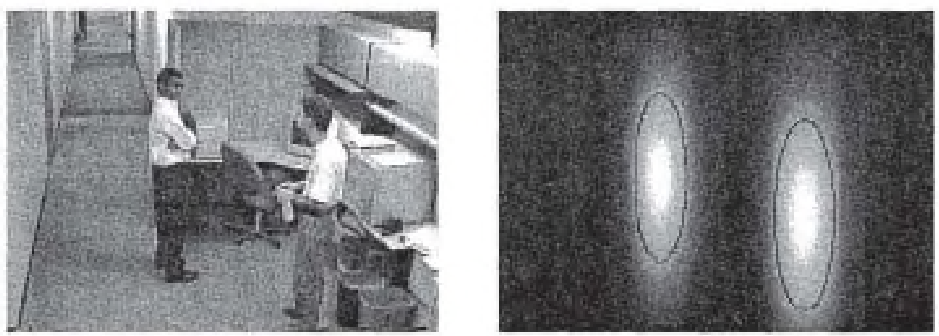
178. As described above under the section on motivation to combine *Kellogg* and *Brill*, *Brill* teaches, among other things, an enhanced event detection platform that reliably handles recognizing interactions of multiple objects. Ex. 1004, 6-9. *Brill* teaches improvements to object tracking so that tracking is not lost when a person's image overlaps that of another object, such as a car. In some systems this situation would cause the object to appear to merge resulting in the loss of tracking until the person walks away from the other object. *Id.* This problem would have been within the knowledge of a POSITA who employed *Dimitrova*'s system to monitor multiple objects. Take the scene including a moving toy and two cups, for example, in *Dimitrova*. Ex. 1006, Fig. 5.

179. *Brill* teaches that its "new approach involv[es] additional image differencing…[which] allows objects to be detected and tracked even when their images overlap." *Id.* *Brill* specifically teaches a background-model based technique. Ex. 1004, 6-9. A POSITA would have been highly motivated to combine the teachings of *Brill* with *Dimitrova*'s monitoring system to solve the tracking loss issue.

180. A POSITA using *Dimitrova*'s system to monitor multiple moving humans or human-human interaction also faced a separate issue. As described in *Brill*, a POSITA knew that it is difficult to monitor movements of one or more people in a scene because they move unpredictably, may move close to one

another, and may occlude each other. Ex. 1004, 14. As *Brill* explains, when two people are in a single scene, it was difficult to maintain the separate tracks of the two people once they merge into a single large region. *Id.*

181. *Dimitrova* specifically teaches monitoring a basketball game that involves detecting and tracking multiple moving humans and various human-human interactions in a given scene. Ex. 1006, Fig. 11. A POSITA employing *Dimitrova*'s system would have faced the issue taught in *Brill*. *Brill* introduces a new method which maintains an estimate of the size and location of the objects, and that creates a separate image which approximates the probability that the object intersects that pixel location. Ex. 1004, 14-15, Figure 15.



(a) (b)

Figure 15: Probability image for the locations of the people in the scene

182. *Brill* teaches that even after an occlusion, the objects are reliably detected and tracked by relying on non-overlapping areas. *Id.*, 15. A POSITA would have been highly motivated to combine the teachings of *Brill* with

Dimitrova's monitoring system to solve the occlusion issue and enhance detection and tracking of multiple moving objects.

183. Indeed, *Dimitrova* also signals to a POSITA that "[m]ore-sophisticated" detection algorithms would be beneficial. Ex. 1006, 32. Moreover, as both references teach a computer-based system employing video cameras, a POSITA would be able to predictably combine the teachings of the two references, without requiring extensive modification to the overall system. *Brill* would be an obvious source because, like *Dimitrova*, *Brill* is directed to video surveillance systems using object detection methods, attribute detection methods, and querying mechanisms. *See, e.g.*, Ex. 1006, 8-20, 25; Ex. 1004, 6-9, 12-14. Similar to *Dimitrova*'s system, *Brill's* system can detect multiple objects and their interactions in a single video scene. *See, e.g.*, Ex. 1006, 25; Ex. 1004, 6-9, 12-14. Given the similarities of their subject matter and teachings, a POSITA would have immediately recognized *Brill*'s advanced multi-object detection and tracking method would have readily worked in *Dimitrova*'s system with a reasonable expectation of success.

VII. THE PETITION PRESENTS NEW ISSUES OF PATENTABILITY AND THE BOARD SHOULD NOT EXERCISE ITS DISCRETION TO REJECT THE PETITION

184. None of the prior art references used in this Petition was cited in a ground for rejection during prosecution, in the *inter partes* reexamination, or in the

ex parte reexamination of the ’923 patent. The Board has already found that the prior art presented herein raises new patentability issues and is not substantially the same as the prior art references previously considered by the PTO against this patent family. IPR2018-00140, Paper No. 8, 6-7 (June 1, 2018).

1. Dispute Regarding the Independence-Based Claim Elements in the Prior Proceedings

185. In each *ex parte* reexamination of the ’923 patent and the ‘912 patent, the patentee argued that the primary references failed to disclose the “independence-based claim elements.” Ex. 1016, 46-52, 54-61; Ex. 1029, 40-46, 51-54. These arguments were directed to the independence claim construction issued discussed above in Section V(C), including attempting to distinguish the ’923 patent from *Courtney*.

186. In the Related IPRs against the ’661 patent, the crux of Patent Owner’s Preliminary Response is that none of the cited references—namely *Kellogg*, *Brill* and *Dimitrova*—disclose the “independence-based claim elements” because they are no different than *Courtney*. IPR2018-00138, Paper 7, 19-28 (March 2, 2018); *see also* IPR2018-00140, Paper 7, 19-22 (March 2, 2018). The Board rejected that argument. IPR2018-00140, Paper No. 8, 6-7 (June 1, 2018). After the ID, Patent Owner further argued that *Dimitrova* is like *Courtney* because

it merely indexes predefined events. IPR2018-00140, Paper No. 11, 37-38 (September 4, 2018).

2. Kellogg Alone or Kellogg and Brill

187. Unlike *Courtney*, *Kellogg* does <u>not</u> apply its queries to a predefined list of events. *Id.* *Kellogg* stores only very basic attributes in its database, such as object's area, duration, and trajectory (Ex. 1003, 22), object's class, centroid, orientation, and bounding box (*id.*, 24), relative spatial attributes (*id.*, 30), time stamps or intervals for the valid times when the object existed (*id.*, 36-37, 52), and volume or height (*id.*, 25, 71). With these basic attributes *Kellogg* allows a user to later invent any arbitrary collection of attributes to define a new event that can be identified based on the collected attributes. For example, after video attributes are stored, a user might create a completely new event to uncover, such as whether an object classified as a person intersected with a new, arbitrarily defined rectangular region. Ex. 1003, 55 ("[s]elect p from Person where p intersects %rectangle."). Notably, the *Kellogg* system does not and could not pre-define and index this intersection activity because the user had not come up with it yet at least because the rectangle was defined in the user's query. Instead of being predefined, this intersection is actually computed based on more fundamental movement attributes. *Id.*

188. Moreover, unlike *Courtney*'s query, *Kellogg's* query does not limit the search to a single activity attribute, but also searches for an object classification attribute, *e.g.*, person. Ex. 1003, 55 ("[s]elect p from Person where p intersects %rectangle."). This is exactly the sort of additional attribute that Patent Owner identified as characterizing its invention. *See* Ex. 1016 at 38-39 ("the specification of the '923 patent also discloses events that are not detected attributes … [*e.g.*,] 'a ***person*** appears'") (emphasis added).

189. In contrast, *Courtney*'s query cannot identify an *ad hoc* after-the-fact intersection event like this because it is not one of *Courtney*'s predefined events (appearance, disappearance, entrance, exit, deposit, removal, motion and rest). Ex. 1021, 10:50-61. Thus, as the Board has already found, *Kellogg* teaches "independence" and *Kellogg* is a substantially different art than what was before the PTO during the reexamination.

190. As explained fully below in Section VIII(A)(1)(e), *Kellogg* also discloses embodiments that apply a new user rule only to the detected attributes themselves without having to use any higher-level abstractions. Indeed, *Kellogg* discloses a specific embodiment where the attribute data is stored in a "bucket index," which is merely a collection of object attributes without any higher-level organizational structure. *See* Section V(D). This is fundamentally different than the *Day* references where the attribute data was *stored* in a hierarchical tree

structure, such that the lower level attributes could only be reached by traversing the abstractions.

3. Dimitrova and Brill

191. Unlike *Courtney*, *Dimitrova* does not apply its queries to a predefined list of events. *Dimitrova* stores only very basic attributes including activity attributes and other object or motion related attributes in its database. Ex. 1006, 17, 20. With these basic attributes *Dimitrova* allows a user to later invent any arbitrary collection of attributes to define a new event that can be identified based on the collected attributes. Ex. 1006, 25, 29.

192. For example, after video attributes are stored, a user might create a completely new rule to uncover a "parade" event where a tall person is waving while the president walks. Ex. 1006, 25 ("retrieve all the sequences in which a tall person is waving while the president walks."). While basic activity attributes such as "waving" and "walking" are stored, *Dimitrova*'s system does not and could not pre-define and index the "parade" event that the user is ultimately interested in because the user only creates this event by later mixing the basic activity attributes of waving and walking with other attributes like "person" classifications and "tall" size characteristics.

193. This is also true for the simpler query "retrieve all the video sequences in which a pet walks and makes a certain trajectory". *Id.*, 20. The user here is not

just interested in the "walking" activity *per se*, but more so in the trajectory that is later arbitrarily defined. Moreover, unlike *Courtney*'s query, *Dimitrova's* query does not limit the search for a single activity attribute "walking," but also searches for an object classification attribute, *e.g.*, pet. *Id.*

194. In contrast, *Courtney*'s query cannot identify an *ad hoc* after-the-fact trajectory event like this because it is not one of *Courtney* 's predefined events (appearance, disappearance, entrance, exit, deposit, removal, motion and rest). Ex. 1021, 10:50-61. Thus, as the Board has already found, *Dimitrova* teaches "independence" and this petition presents substantially different art than what was before the PTO during the reexamination.

195. As explained fully below in Section VIII(A)(1)(e), *Dimitrova* also discloses embodiments that apply a new user rule only to the detected attributes themselves without using any higher-level abstractions. For example, the pet walking query above searches only collected attributes without any higher-level abstraction, such as the object classification attribute "pet," the activity attribute "walking" and the trajectory attribute.

VIII. CLAIMS 1-41 OF THE ’923 PATENT ARE UNPATENTABLE OVER THE PRIOR ART

A. *Kellogg* Anticipates Claims 1-41

1. Independent Claim 1

a. “detecting an object in a video from a single camera” [1.1]

196. As shown in Fig. 5, the ’923 patent describes that object detection (51, 52) occurs before attribute detection (57). Ex. 1001, 9:30-10:57. The ’923 patent provides two embodiments used in detecting objects (*e.g.*, detecting objects via motion (51) and detecting objects via change (52)), both of which take into consideration that objects in a scene move at certain time points. *Id.*, 9:33, 9:39, Fig. 5. But, ’923 the patent does not purport to have invented object detection. In fact, the ’923 patent admits that many object detection methods were known. *Id.*, 9:35-38, 9:44-48. The ’923 patent does not limit the object detection to a specific algorithm and allows a person skilled in the art to use “any” such algorithm available. *Id.*, 9:34-35, 9:39-41. Moreover, the claim limitation does not even require using any algorithm. The limitation merely requires that an object is detected and nothing more.

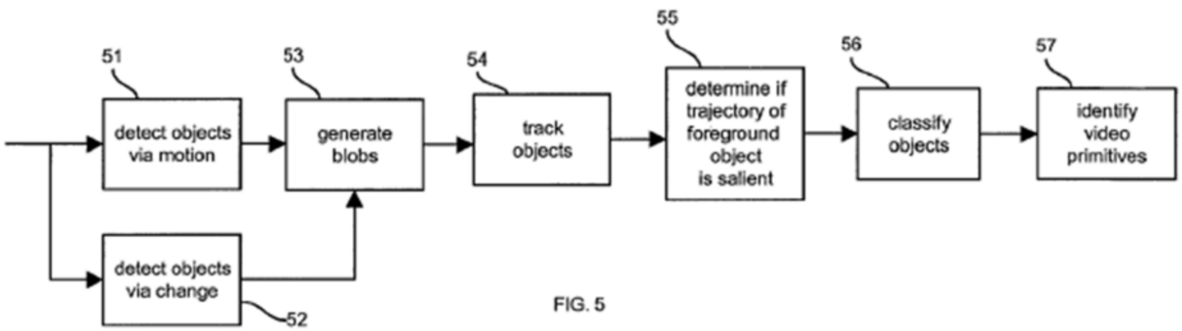


FIG. 5

197. *Kellogg* discloses detecting objects, tracking them, and storing information about them in the visual memory. Ex. 1003, 68-69. For example, *Kellogg* discloses "real-time processing of CCD camera images." *Id.*, 77. *Kellogg* discloses that when a new object is found, *i.e.*, detected in a video, an object identifier ("OID") is assigned to that object. *Id.*, 50. *Kellogg* also discloses that "image processing using video cameras ***tracks*** objects and stores information about them." *Id.*, 68. *Kellogg* further explains that "a security system could track a person walking down a hallway and store a new version describing that person's location every tenth of a second." *Id.* Thus, the *Kellogg* system first detects an object and then tracks it, which is consistent with not only the '923 patent's detection and tracking process but also the detection and tracking process admitted as prior art in the '923 patent. Ex. 1001, Fig. 5, 10:29-30.

198. *Kellogg* discloses "real-time processing of CCD camera images" and software that "tracks ***people*** walking in its ***field of view***." Ex. 1003, 77 (emphasis added). The plural word "people" emphasizes that more than one object is

detected in a video from single camera. *Kellogg* also contemplates detecting at least two different objects in a video from a single camera. *Id.*, 79 (monitoring both people and vehicles at an outdoor alarm region); Figure 4-9 (monitoring three objects—each represented as small squares—from a single camera (red)).

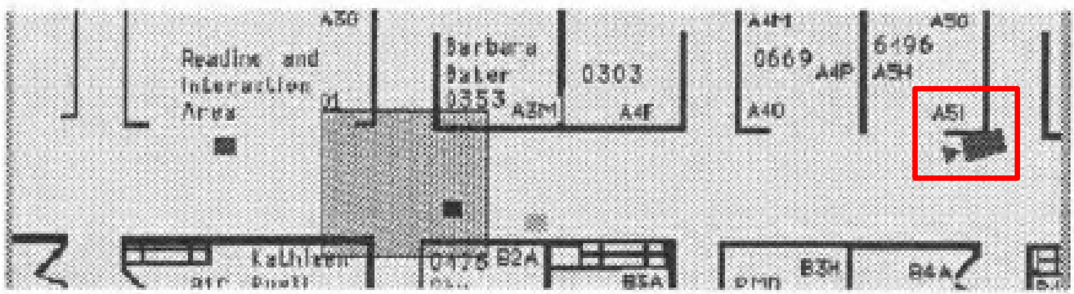


Figure 4-9: Graphical query results

199. As shown in Figure 3-5, *Kellogg* detects the spatial relationship among different objects, *i.e.*, first and second objects, and keeps track of them in a single scene. *Id.*, 30-31; *see also* 21 (index tracks "centroids of moving objects").

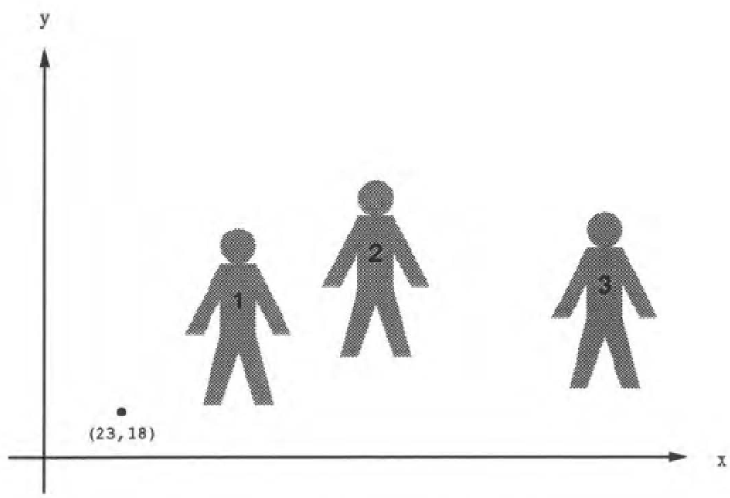


Figure 3-5: Relative spatial objects

200. In view of *Kellogg*'s disclosure, a POSITA would understand that *Kellogg* discloses this element.

201. To the extent the Patent Owner argues that the limitation somehow requires detecting an object from a single camera, that would be an incorrect reading because the limitation only requires that an object is detected in a ***video***. Also, such a reading would be inconsistent with the specification where the only disclosed embodiment is that the "computer system 11 obtains source video" from the camera to further process it for object detection and attribute extraction. *Id.*, 9:23-31.

b. "detecting a plurality of attributes of the object by analyzing the video from said single camera, the plurality of attributes including at least one of a physical attribute and a temporal attribute, each attribute representing a characteristic of the detected object" [1.2]

202. The '923 patent specification does not define or use the term "physical attribute." The word "physical" is only used once in the specification where it describes that a "physical subject" can be an "object." Ex. 1001, 3:27-29. A POSITA would understand that the word "physical" means that it can be perceived through the senses, *i.e.*, observable (because the system uses a video sensor). *See* Ex. 1037, 3. The specification uses the language "observable attributes" and includes, classification, size, shape, color, texture, position, velocity, speed, internal motion, motion, salient motion, feature of a salient motion, scene change, feature of a scene change etc., all of which are observable characteristics of an object. *See, e.g., id.*, 7:6-8:15. Event discriminators are described in terms of these attributes. *Id.*, 7:5:6.

203. The '923 patent specifically lists "position," "position of an object in image space as a function of time," and "an approximate position of an object in a three-dimensional representation of the environment as a function of time" as an observable attribute. *Id.*, 7:9, 7:51-54. Thus, these attributes are physical.

204. The term "physical attribute" can also include "spatial attributes." The '923 patent describes that spatial attributes include crossing a line, entering an area, or crossing a line from the left. *Id.*, 8:65-67. These characteristics require referring to an object's position in image space as a function of time.

205. *Kellogg* discloses detecting a plurality of attributes of the object by analyzing the video. For example, the "software estimates the positions and heights of people" from the field of view. Ex. 1003, 77. *Kellogg* teaches several physical attributes, including an object's classification (*e.g.*, person, cube), centroid (position of the center of the object), orientation (position of the object), bounding box (size, shape, and position of the object). Ex. 1003, 24. Additional physical attributes can be detected for certain types of objects, such as the volume of a detected cube (*id.*, 25) or the height of a person (*id.*, 71).

206. The term "temporal attributes" is also not defined in the specification of the '923 patent. Nonetheless, examples of the term include "every 15 minutes," "between 9:00 p.m. to 6:30 a.m.," "less than 5 minutes," "longer than 30 seconds," "over the weekend," and "within 20 minutes of." Ex. 1001, 8:32-36.

207. *Kellogg* discloses that at least the temporal attributes "*TemporalInterval*" representing the interval time(s) for which an object is present, "*VMTime*," representing the single time that an object is present (*i.e.*, a timestamp),

and "*RelativeTemporalObject,"* representing the existence of an object in relation to other objects may be detected. Ex. 1003, 36-41.

208. *Kellogg*'s physical and temporal attributes represent characteristics of the detected object because the physical attributes are observable traits of the detected object and the temporal information of a detected object is always associated with the object in a video.

c. "selecting a new user rule after detecting the plurality of attributes" [1.3]

209. The language "selecting a new user rule" is not used in the '923 patent specification. Indeed, the '923 patent does not even use the word "rule" in this context. However, the '923 patent discloses, "[d]uring tasking, the operator ***selects*** an area representing the space around the desired retail display." Ex. 1001, 15:26-28 (emphasis added). And the patent further describes that "[a]s a discriminator, the operator ***defines*** that he or she wishes to monitor people-sized objects that enter the area." *Id.*, 15:28-32 (emphasis added). The patent appears to use the words "select" and "define" interchangeably in the context of a user rule.

210. *Kellogg* discloses a system tasking mechanism to allow users to create or define *ad hoc* queries using a set of attributes, *i.e.*, selecting a new user rule. *Kellogg* provides many examples of creating user-defined queries ***after*** detecting the plurality of attributes.

211. One example is the following query that "tracks all objects that ***came*** within 3 units of a given object on its trajectory during a certain set of valid times," which hereinafter is referred to as the "Approach Query." Ex. 1003, 62-63 (emphasis added). Use of the past tense "came" further emphasizes that a user of *Kellogg's* system is creating a new rule seeking to extract a new event of interest from previously detected attributes.

```
Select p from Person
  during %times-1
  where p in
    (Select q from Person
      where q centroid within 3 of %spatiotemporal-spec
      during %times-2)
```

212. The Approach Query specifies a combination of a set of physical and temporal attributes together with object classification attributes for identifying an "Approach Event" between person p and persons q by searching for any persons q that came within a certain distance of person p at a given time period. Ex.1003, 62-63. For example, the query searches the centroid of the two objects and their trajectories, and the distance between the moving centroids. *Id.* The query also searches for trajectories within certain periods, *e.g.*, time-1 and time-2. *Id.* The query also applies the object classification attribute, *i.e.*, person, for both objects. *Id.* These attributes are detected before a user creates the Approach Query.

213. To the extent Patent Owner argues that the "new user rule" requires setting a response (as it did in the Related IPRs), this is incorrect because neither the claims nor the specification require one. The '923 patent unequivocally deems it "***optional***." Ex. 1001, 8:37, 8:56-58 (emphasis added). The scope of a "response" in the '923 patent is very broad. *Id.*, 8:37-49. For example, it includes activating a visual or audio alert on the system display, directing the computer system, or even saving data to a medium. *Id.* These responses do not require alerting the user in any way. A POSITA would understand that even a simple return of a query result is consistent with a "response" as used in the patent.

214. Even assuming the term "new user rule" requires setting a response—which is not the case because the '923 patent clearly states that it is optional—*Kellogg* still discloses the term. For example, *Kellogg's* system also allows a user to enter a "delay specification that indicates how long an object must remain in that region before the system triggers an alarm." Ex. 1003, 79-80. This feature allows a user to specify a "response" that would trigger only after an object remains in the region for a certain amount of time. *Id.*, 80.

215. *Kellogg* recognizes that "[a] query language provides flexibility and expressiveness but can be hard to use," so more user-friendly interfaces "might be more suitable." *Id.*, 54. Thus, *Kellogg* also discloses two types of user-friendly, graphical interfaces for selecting rules. The first, shown in Figure 4-7 of *Kellogg*,

is a graphical query interface that allows the user to select a region or locations of interest for a query.

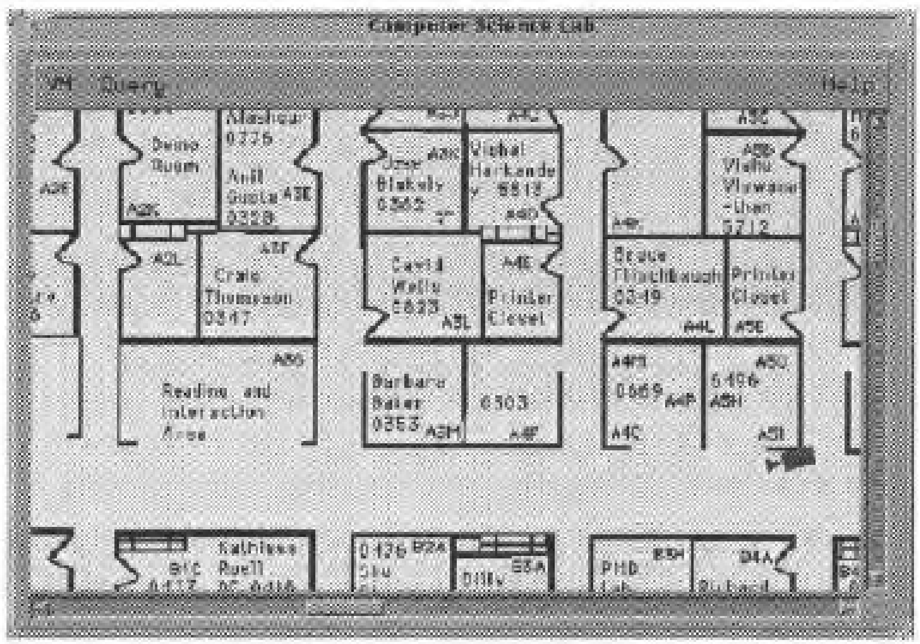


Figure 4-7: Graphical query interface viewing region

216. The second query interface, shown in annotated Figure 4-8 of *Kellogg*, allows a user to type in a time interval (between a "start" and "stop" time) or push a button for "real-time" monitoring, and to select the classes of objects of interest (e.g., animal, furniture, person, etc.). Ex. 1003, 52, 79.

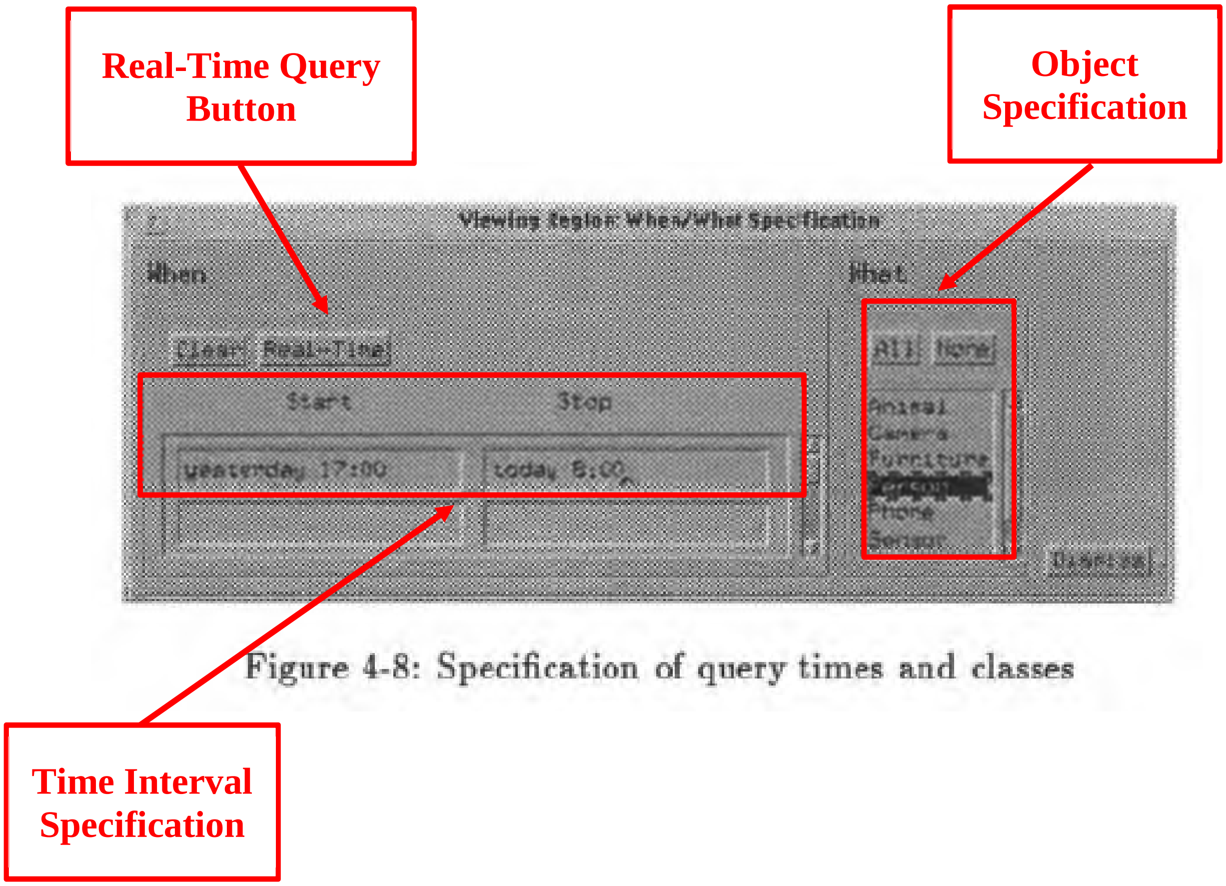


Figure 4-8: Specification of query times and classes

d. "after detecting the plurality of attributes and after selecting the new user rule, identifying an event of the object that is not one of the detected attributes of the object by applying the user rule to the plurality of detected attributes" [1.4a]

217. As discussed in Section V(C)(2), the limitation, namely, "identifying an event of the object that is not one of the detected attributes of the object" means "the user defined event comprises a minimum of two attributes."

218. The Approach Query, as discussed above, identifies an "Approach Event" between the person p and persons q. Ex. 1003, 62-63. And doing so, the query is applied to at least two different attributes from the plurality of detected

attributes, *e.g.*, centroid trajectories of objects p and q, object classifications (person) of objects p and q, time-1 and time-2. *Id.* These are non-limiting examples. The query also analyzes whether the distance between the centroid trajectories of objects p and q to is within 3 units. *Id.* Accordingly, the event being identified is not just one of the detected attributes, but rather a collection or a set of attributes.

219. In another example, a user may want to find all people who intersect (*i.e.*, overlap) a user-specified area of interest (hereinafter the "Intersect Event"), and the query may take the form of "Select p from Person where p intersects %rectangle" (hereinafter the "Intersect Query"). Ex. 1003, 55. Based on this query, at least two attributes would be searched to find objects: the class "Person" and position (*e.g.*, object polygon vertices or bounding boxes) that overlap the rectangle. *Id.* As discussed earlier, a position of the object is a physical attribute. *See* Section VIII(A)(1)(b). Accordingly, the Intersect Event is not one of the detected attributes of the object.

220. It is clear that the query is applied after the plurality of attributes are detected and after creating the query because the entire concept of *Kellogg* is to record basic attributes so that a search can later be created based on any arbitrary subset of the recorded attributes. *See, e.g.*, Ex. 1003, 53 ("The visual memory provides a powerful and expressive mechanism for retrieving information. … It is

also designed to meet a wide variety of retrieval needs, providing flexibility in specifying objects of interest.”); *id.*, 78 (“The scene monitoring prototype includes a graphical interface through which users can query the visual memory to retrieve information stored by the tracking software. A user establishes regions, times, and object types of interest, and the visual memory retrieves the corresponding objects.”); *see also id.*, 54.

e. “wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes” [1.4b]

221. As discussed in Section V(D), the limitation “the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes” requires that the prior art have the ability to search only the attributes themselves. Prior art, for example, that ***always*** require searches to process object-oriented abstractions, rather than just the attributes themselves, would not disclose this limitation.

222. *Kellogg* has the ability to search only the attributes themselves and *Kellogg* does not require traversing a tree structure of abstractions to search the detected attributes. *Kellogg* discloses a specific embodiment where the attribute data is stored in a “bucket index.” Ex. 1003, 83. *Kellogg* explains that a “bucket index simply maintains a list of all the objects stored in the visual memory.” *Id.* In

other words, the list is merely a collection of attributes without any higher-level organizational structure. *Kellogg* further discloses, "[a] bucket index answers a query by retrieving all the objects in its list and checking them against the query specification," demonstrating that the system can search only the attributes themselves. *Id.*

f. "wherein the plurality of attributes that are detected are independent of which event is identified" [1.5]

223. As discussed in Section V(C)(3), this limitation requires that attribute detection is not impacted or affected by the event detection process. *Kellogg*'s Approach Query, as shown below, meets this limitation because it identifies the "Approach" event by searching multiple detected and, pre-collected, activity, physical and temporal attributes. For example, the query searches the pre-collected centroid trajectories of the two moving objects, and calculates the distance between the moving centroids. These moving trajectories must be within certain time periods, *e.g.*, time-1 and time-2, which are also pre-collected. The query also applies the object classification attribute, *i.e.*, person, for both objects. *Id.*

```
Select p from Person
  during %times-1
  where p in
     (Select q from Person
       where q centroid within 3 of %spatiotemporal-spec
       during %times-2)
```

224. None of the attribute detection process is affected by the identification of this Approach event. For example, the system collects the basic centroid attributes of the two objects regardless of the distance parameter set after the fact by the user. Moreover, a user can adjust the distance parameter *ad hoc*, thus specifying what an "Approach" event means to the user in the same video sequence but none of the detected attributes, including the distance between the moving centroids, will be affected.

225. Other detected attributes can be mixed with these detected attributes to identify a completely different event. For example, an additional physical attribute such as "size" (Ex. 1003, 24) can be mixed into the Approach Query to identify persons q having a certain height. This will not affect the attribute detection process because the size attribute will have been pre-collected together with other basic attributes such as the centroid trajectory or the classification attribute.

226. The centroid attributes can also be used in identifying events other than the "Approach" event, such as identifying objects whose centroids are located in an arbitrary space using the "Intersects Query." *See* Ex. 1003, 55. The newly identified event would in no way affect the pre-collected attributes.

227. In fact, *Kellogg* discloses that the queries are forensic in nature where "[a] large amount of [stored] information could be established *prior to* application

execution." Ex. 1003, 9 (emphasis added); *see also, id.,* 19 (scene monitoring systems update the visual memory without regard to the application that may later use the data, which "do not need to know how [the database] achieves its results"). Alternatively, the queries may be applied in real-time, with the query "constantly poll[ing] the database for new information." *Id.*, 79. In either mode, *Kellogg's* attributes are detected and stored in the visual memory database without being affected by the event detection process.

g. "wherein the step of identifying the event of the object identifies the event without reprocessing the video" [1.6]

228. *Kellogg* discloses that events are identified by applying query specifications (user rules) to the data stored in the database (detected attributes). This does not require any reprocessing of the video because *Kellogg's* system has already stored the object attributes in its database, so it needs only to search the stored attributes to find objects engaged in activities that meet the query specification. Ex. 1003, 53-54, 79. *Kellogg* also explains that its system creates attributes such as the "positions and heights of people" while it engages in "real-time processing" of video, which "yields enough information to test the visual memory's performance to provide interesting data for queries to retrieve," without reprocessing the video. *Id*., 77.

229. This is consistent with standard database and query systems. *Kellogg*'s system queries the database of attributes to identify an event, but the original data (*i.e.*, the video) is not reprocessed.

h. "wherein the event of the object refers to the object engaged in an activity" [1.7]

230. In the '923 patent, an activity "refers to one or more composites of actions of one or more objects." Ex. 1001, 3:31-33. *Kellogg's* Approach Query identifies an "Approach" event of person p with any person q. Ex. 1003, 62-63. The objects p and/or q are each engaged in an action of moving along a certain trajectory. *Id.* Accordingly, the "Approach" event refers to an object engaged in an activity.

2. Claims 2, 4, 7[2][3], 13, 16, 23, 25, 28[2], 34 and 38

231. *Kellogg* discloses that the queries do not require all detected attributes as recited in claims 2, 4, 7[2], 13, 16, 23, 25, 28[2],34 and 38.

- "selecting the new user rule comprises selecting a subset of the plurality of attributes for analysis" (claim 2 [depends on claim 1], claim 23 [depends on claim 22])
- "no analysis is performed on at least some of the detected attributes to detect an event" (claim 4[depends on claim 1])
- "analyzing only a subset of the attributes stored in the memory" (claim 7[2] [depends on claim 1], claim 28[2] [depends on claim 22])

[3] [2] refers to the second element of the claim.

- "analyzing, of the plurality of attributes, only a selected subset of the plurality of attributes" (claim 13 [depends on claim 9], claim 34 [depends on claim 30])
- "analyzing a selection of individual ones of the detected plural attributes" (claim 16 [depends on claim 9], claim 38 [depends on claim 30])
- "do not cause the computer system to perform an analysis on at least some of the detected attributes to detect an event" (claim 25 [depends on claim 22])

232. *Kellogg*'s system detects heights of people as an attribute. Ex. 1003, 77. But the Approach Query, below, does not search for heights of person p or persons q. Accordingly, *Kellogg* discloses queries that do not require all detected attributes as recited in claims above. A POSITA would understand that *Kellogg* discloses the above limitations.

```
Select p from Person
  during %times-1
  where p in
     (Select q from Person
       where q centroid within 3 of %spatiotemporal-spec
       during %times-2)
```

3. Claims 3, 7[1][4], 17, 24, 28[1] and 39

233. *Kellogg* discloses the additional features of claims 3, 7[1] 17, 24, 28[1] and 39. The additional features added by these claim limitations all require

[4] [1] refers to the first element of the claim.

that the detected attributes are stored in the system, and claims 17 and 39 further require that it happens prior to selecting a set of attributes to define a user rule.

- "the plurality of attributes that are detected are defined in a device prior to a selection of a subset of the plurality of attributes" (claim 3 [depends on claim 1], claim 24 [depends on claim 22])
- "storing the detected plurality of attributes in memory" (claim 7[1] [depends on claim 1], claim 28[1] [depends on claim 22])
- "the plural attributes detected by the means for detecting are defined in the video device independent of a selection of the detected plural attributes" (claim 17 [depends on claim 9], claim 39 [depends on claim 30])

234. As explained in *Kellogg*, "computer vision algorithms for a security system could analyze data provided by various cameras and ***store*** information in the visual ***memory***." Ex. 1003, 10 (emphasis added). "Applications could ***then retrieve*** this data to track objects, watch for suspicious events, and ***respond to user queries***." *Id.*, 10 (emphases added). This demonstrates that *Kellogg*'s system stores the detected attributes in memory, specifically prior to a user defining a user rule. *Kellogg* also discloses that the "query mechanism works on two levels, on disk [*i.e.*, longer-term storage] and in [working] memory." *Id.*, 54; *see also* 68. Either way, attributes are stored in memory which is located in a device or video device. Accordingly, *Kellogg* discloses the above limitations, as recited in the claims.

4. Claims 5, 6, 15, 21, 26, 27 and 37

235. *Kellogg* discloses the additional features of claims 5, 6, 15, 21, 26, 27 and 37. The additional features of claims 5, 6, 15, 21, 26, 27 and 37 require that the new user rule is applied to plural physical and/or plural temporal attributes.

- "plurality of attributes include plural physical attributes and the method comprises applying the new user rule to a plural number of physical attributes" (claim 5 [depends on claim 1], claim 26[5] [depends on claim 22])
- "plurality of attributes include plural temporal attributes and the method comprises applying the new user rule to a plural number of temporal attributes" (claim 6 [depends on claim 1], claim 27[6] [depends on claim 22])
- "analyzing at least two selected physical attributes of the plurality of attributes" (claim 15 [depends on claim 9], claim 37[7] [depends on claim 30])
- "a video device … which detects plural physical attributes and plural temporal attributes of the detected object upon analyzing the video; and then, selecting the new user rule to provide an analysis of a

---

[5] A POSITA would understand that there is no meaningful difference between claims 5 and 26.

[6] A POSITA would understand that there is no meaningful difference between claims 6 and 27.

[7] Claim 37 recites the limitation in a means-plus-function format. The corresponding structure is the same structure identified in Section VIII(A)(15)(e), which performs the additional function. *See* Section V(E)(4).

> combination of the plural physical attributes and the plural temporal attributes to detect the event" (claim 21 [depends on claim 20])

236. *Kellogg* discloses many events that involve a new user rule applied to a plural number of physical attributes of an object. As discussed in Section VIII(A)(1)(b), a physical attribute includes ***classification***, size, shape, color, texture, ***position***, velocity, speed, internal motion, motion, salient motion, feature of a salient motion, scene change, feature of a scene change etc., all of which are observable characteristics of an object.

```
Select p from Person
  during %times-1
  where p in
     (Select q from Person
       where q centroid within 3 of %spatiotemporal-spec
       during %times-2)
```

237. *Kellogg*'s Approach Query, shown above, identifies an "Approach Event" between person p and person(s) q by searching for any persons q that came within a certain distance of person p at a given time period. Ex. 1003, 62-63.

238. A plural number of physical attributes of an object are searched in the Approach Query. Specifically, the query searches at least the following physical attributes: (1) the classification "person" of the two objects (2) the centroid of the two objects, (3) the centroids' trajectories, and (4) the distance between the moving centroids. *Id.* A centroid is essentially the position of an object, which the '923

patent identifies as a physical attribute. *See* Section VIII(A)(1)(b). The query also applies the object classification "person" for both objects, which is also a physical attribute. Ex. 1003, 62-63. Accordingly, *Kellogg* discloses plural physical attributes.

```
Select p from Person
  during %times-1
  where p in
     (Select q from Person
       where q centroid within 3 of %spatiotemporal-spec
       during %times-2)
```

239. A plural number of temporal attributes of an object are searched in the Approach Query. For example, the query applies two different temporal attributes, times-1 and times-2. *Id.* Time-1 is associated with person p and time-2 is associated with any person q. *Id.*

240. *Kellogg* additionally discloses applying a new user rule to a plural number of temporal attributes. The "latest-during" query can be used to detect, for example, the last object to leave the scene, which applies the query (user rule) to the *TemporalInterval* or *VMtime* attribute to detect that the object is present during the specified query time, and the *RelativeTemporalObject* attribute to determine which object was present after all of the other objects.

241. Finally, for the reasons explained above, *Kellogg*'s Approach Query analyzes a combination of the plural physical attributes and the plural temporal attributes to detect the Approach event.

5. Independent Claim 8

242. *Kellogg* discloses the method of claim 8, including elements [8.1] to [8.7].

a. "detecting first and second objects in a video from a single camera" [8.1]

243. *Kellogg* discloses detecting an "object" in a video from a single camera for reasons explained in Section VIII(A)(1)(a) for limitation [1.1]. Compared to limitation [1.1], this limitation requires detecting "first and second objects," which is also disclosed by *Kellogg*.

244. *Kellogg* discloses a software that "tracks ***people*** walking in its ***field of view***." Ex. 1003, 77 (emphasis added). The plural word "people" emphasizes that more than one objects are detected. Indeed, the Approach Query, as discussed above, identifies an "Approach Event" between the detected person p and detected person(s) q, *i.e.*, the first and second objects, respectively. Ex. 1003, 62-63. The query specifically analyzes whether the distance between the centroid trajectories of persons p and q come within 3 units, thus identifying an approach event between

the first and second objects. *Id.* Without the capability of detecting a first and second object in a video the Approach Query would be of no use.

245. *Kellogg* also contemplates detecting at least two different objects in its field of view. *Id.*, 79 (monitoring both people and vehicles at an outdoor alarm region). *Kellogg* also teaches monitoring three objects (each represented as small squares) from a single camera (on the right) as shown in Figure 4-9.

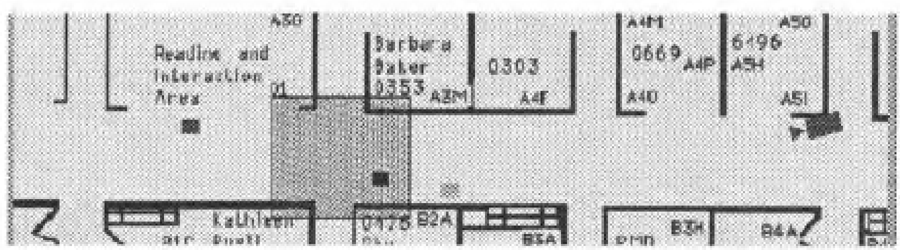


Figure 4-9: Graphical query results

246. As shown in Figure 3-5, *Kellogg* detects the spatial relationship among different objects, *i.e.*, first and second objects, and keeps track of them in a single scene. *Id.*, 30-31. Furthermore, *Kellogg's* "real-time scene monitoring system could set up an index to track the centroids of moving ***objects***." *Id.*, 21 (emphasis added).

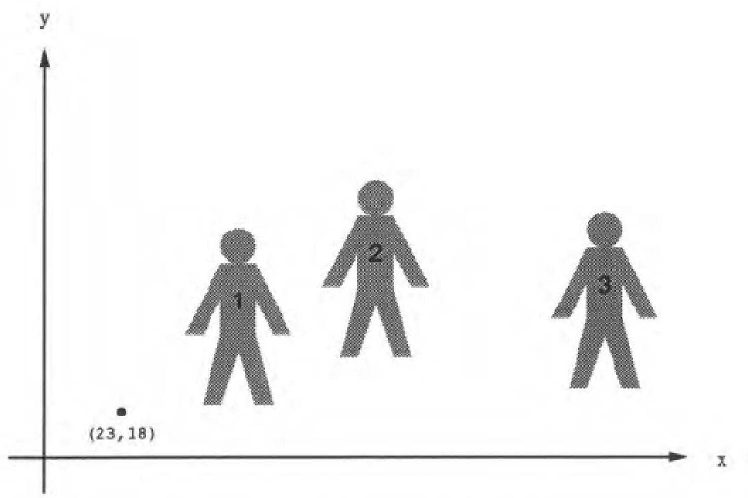


Figure 3-5: Relative spatial objects

b. "detecting a plurality of attributes of each of the detected first and second objects by analyzing the video from said single camera, each attribute representing a characteristic of the respective detected object" [8.2]

247. *Kellogg* discloses detecting a plurality of attributes of each detected object by analyzing the video. *See* Section VIII(A)(1)(b). Moreover, the "software estimates the ***positions and heights of people***" from the field of view. Ex. 1003, 77 (emphasis added). The plural word "people" further emphasizes that the plurality of attributes are detected for each of the detected first and second objects. These attributes correspond to characteristics of the detected objects. *See* Section V(A).

c. "selecting a new user rule" [8.3]

248. As explained in Section VIII(A)(1)(c), *Kellogg* discloses this element.

d. "after detecting the plurality of attributes, identifying an event that is not one of the detected attributes of the first and second objects by applying the new user rule to the plurality of detected attributes" [8.4a]

249. As explained in Section VIII(A)(1)(d), *Kellogg* discloses this element.

e. wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes" [8.4b]

250. As explained in Section VIII(A)(1)(e), *Kellogg* discloses this element.

f. "wherein the plurality of attributes that are detected are independent of which event is identified" [8.5]

251. As explained in Section VIII(A)(1)(f), *Kellogg* discloses this element.

g. "wherein the step of identifying an event of the object comprises identifying a first event of the first object interacting with the second object by analyzing the detected attributes of the first and second objects, the first event not being one of the detected attributes" [8.6]

252. The specification of the '923 patent does not specifically disclose identifying an event of a first object interacting with the second object by analyzing the detected attributes of the first and second objects. Nor does the specification define the term "interacting." While the '923 patent discloses an activity attribute called "moving to interact with another object" this is an attribute and not an event of itself. Ex. 1001, 7:42-46. Nonetheless, the '923 patent contemplates an "interacting" event embodiment that looks for "two objects

com[ing] together." *Id.*, 8:62-64. A POSITA would understand that an event identifying two objects coming together would meet a first object interacting with a second object.

253. As discussed in Section VIII(A)(1)(d), the Approach Query identifies an "Approach" event, *i.e.*, a first event, between the first object p and the second object q. Ex. 1003, 62-63. The claimed limitation "first object interacting with the second object" broadly encompasses an embodiment where one object comes within an arbitrary distance, *e.g.*, 3 units, of another object on its motion trajectory. *Id.* This is consistent with the '923 patent's disclosure of an interacting event as discussed above. Ex. 1001, 8:62-64. The query searches collected attributes such as the centroid trajectories of the two objects p and q, and the distance between the moving centroids. Ex. 1003, 62-63. The query also searches for trajectories within certain periods, *e.g.*, time-1 for p and time-2 for q. *Id.* The query also applies the object classification attribute, *i.e.*, person, for both objects p and q. *Id.* The query searches for detected attributes of the first and second objects. And because identifying the "Approach" event comprises searching for a minimum of two attributes, the first event is not one of the detected attributes.

254. A POSITA would have understood that the distance between the two objects p and q, *i.e.*, 3 units, in *Kellogg*'s Approach Query is merely an exemplary distance. *Kellogg* describes that a unit is represented by a point or coordinate, so

within 3 units could mean that the objects are in very close proximity. *Id.*, 25. In any event, a POSITA would have readily understood that the distance can be adjusted as he or she sees fit, including 0 units, which would clearly mean that the two objects would be physically contacting each other. Both scenarios would be well within the claim limitation that requires first object interacting with the second object.

h. "wherein the event of the object refers to the object engaged in an activity" [8.7]

255. As explained in Section VIII(A)(1)(h), *Kellogg* discloses this element.

6. Independent Claim 9

256. Claim 9 recites means-plus-function elements ([9.1]-[9.2] and [9.4]-[9.8]) with functions substantially corresponding to the steps of claim 1.

a. A video device comprising

257. *Kellogg* discloses a video device comprising a computer-based visual memory system (*i.e.*, a "video device") that includes video cameras, image processing software, a visual memory database, and a user interface as shown in Figure 4-1. Ex. 1003, 68-70, 77-80.

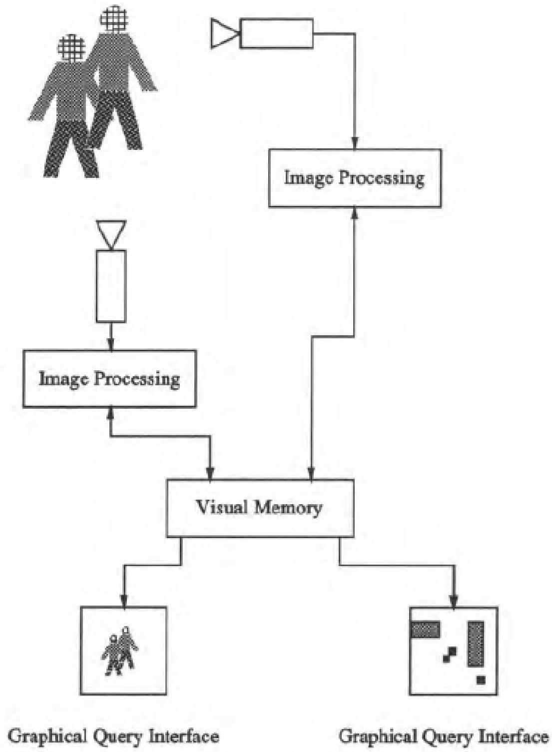


Figure 4-1: Scene monitoring prototype

b. "means for detecting an object in a video from a single camera" [9.1]

258. *Kellogg* discloses the corresponding structures of this means-plus-function element (*see* Section V(E)(1)) in view of the evidence provided herein in addition to Section VIII(A)(1)(a).

259. *Kellogg's* system is a computer-based system and includes a computer or processing device performing the recited functions of these elements. Ex. 1003, 68-70, 77-80.

260. The '923 patent admits that object detection algorithms are well known and expressly states that "any" detection algorithm can be used. Ex. 1001,

9:30-41; 10:27-30. Accordingly, *Kellogg* teaches motion and/or change detection algorithms and computer vision algorithms, within the scope of that described in the ’923 patent, for detecting an object in a video and detecting attributes of the object. Ex. 1001, 9:33-41. In other words, in order to detect the attributes disclosed in *Kellogg*, motion and/or change detection algorithms are required. And the ’923 patent has affirmatively stated that any such algorithm is suitable structure for the disclosed and claimed system.

261. *Kellogg* reviews several known systems that detect and store attributes of objects. Ex. 1003, 13-17. *Kellogg’s* system employs the image processing software developed at Texas Instruments for detecting and tracking objects. *Id.*, 77. This is within the scope of that described in the ’923 patent which recognizes that any conventional object detection algorithm would work well. Ex. 1001, 9:33-41.

c. “means for detecting a plurality of attributes of the object by analyzing the video from said single camera, the plurality of attributes including at least a physical attribute and a temporal attribute, each attribute representing a characteristic of the detected object” [9.2]

262. *Kellogg* discloses the corresponding structures of this means-plus-function element (*see* Section V(E)(2)) in view of the evidence provided herein in addition to Section VIII(A)(1)(b). The ’923 patent briefly discloses this process as

using the same conventional algorithms as the object detection and tracking process. Ex. 1001, 10:49-51. *Kellogg's* system fits within this generic disclosure. *Kellogg's* system is a computer-based system and includes a computer or processing device performing the recited functions of these elements. Ex. 1003, 68-70, 77-80. *Kellogg* reviews several known systems that detect and store attributes of objects. *Id.*, 13-17. Also, *Kellogg's* system relies on the image processing software developed at Texas Instruments for detecting attributes of the object. *Id.*, 77. Accordingly, conventional attribute detection techniques are used in *Kellogg*'s system are exactly within the scope of what is disclosed by the '923 patent. Ex. 1001, 10:11-22, 10:27-30, 10:39-41, 10:44-47.

d. "a memory storing the plurality of detected attributes" [9.3]

263. For the reasons in Section VIII(A)(3), *Kellogg* teaches this element.

e. "means for selecting a new user rule after the plurality of detected attributes are stored in memory" [9.4]

264. *Kellogg* discloses the corresponding structures of this means-plus-function element (*see* Section V(E)(3)) in view of the evidence provided herein in addition to Section VIII(A)(1)(c). The user interface and query mechanisms of *Kellogg*'s system provide the means for selecting a new user rule that defines an event. A POSITA would understand *Kellogg's* user interface includes a display or

monitor (Figs. 4-1, 4-7, 4-8 and related text) and input devices such as a mouse or keyboard for a user to enter queries. It is clear that the query mechanism is tasked after the plurality of attributes are stored in memory because the entire concept of *Kellogg* is to record basic attributes so that a search can later be created based on any arbitrary subset of the recorded attributes. Ex. 1003, 10, 53, 54, 78.

f. "means for identifying an event of the object that is not one of the detected attributes of the object by applying a selected new user rule to the plurality of attributes stored in memory" [9.5]; "for identifying the event independent of when the attributes are stored in memory" [9.6]; "and for identifying the event without reprocessing the video" [9.7]; "wherein the applying the new rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes." [9.8]

265. The corresponding structure of these means-plus-function elements (*see* Section V(E)(4)) is disclosed by *Kellogg's* query mechanisms as shown in Section VIII(A)(1)(d)-(f).

g. "wherein the event of the object refers to the object engaged in an activity" [9.9]

266. As explained in Section VIII(A)(1)(h), *Kellogg* teaches this element.

7. Claims 10 and 31

267. Claims 10 and 31 depend from claims 9 and 30, respectively. *Kellogg* discloses the additional feature of "a video camera operable to obtain the video." Ex. 1003, 68 (In the "real-time scene monitoring prototype…image processing

using video ***cameras*** tracks objects and stores information about them in the visual memory.") (emphasis added).

8. Claims 11 and 32

268. Claims 11 and 32 depend from claims 9 and 30, respectively. *Kellogg* discloses these claims.

a. "in real time"

269. Claims 11 and 32 add the following features that require identifying a first event in real time.

- means for identifying a first event of the object in real time by analyzing, of the plurality of attributes, only a first selected subset of the plurality of attributes (claim 11)

- means for identifying a first event in real time by analyzing, of the plurality of attributes, only a first selected subset of the plurality of attributes (claim 32)

270. The corresponding structure of the means-plus-function limitations in claims 11 and 32 is the structure identified in Sections VIII(A)(6)(f) and VIII(A)(15)(e), respectively, each of which performing the additional function of identifying a first event in real time. *See* Section V(E)(4).

271. The '923 patent discloses "real time" in several different embodiments. For example, "[t]he video surveillance system of the invention operates automatically, detects and ***archives*** video primitives of objects in the scene, and detects event occurrences in ***real time*** using event discriminators**.** In

addition, ***action*** is taken in ***real time***." Ex. 1001, 9:14-18 (emphasis added); *see also id.,* 2:48-50 ("An object of the invention is to produce a ***real time alarm*** based on an automatic detection of an event from video Surveillance data.") (emphasis added); *see also id.,* 11:34-35 ("the information can be viewed by the operator at any time, including ***real time***.") (emphasis added); *see also id.*, 9:25-26 ("video primitives are extracted in ***real time*** from the source video.") (emphasis added). Consequently, the term ***real time*** is used in the '923 patent to pertain to time-sensitive processes, i.e. those that depend upon timeliness and temporal continuity to properly perform their intended functions.

272. Importantly, claims 11 and 32 depend from claims 9 and 30, respectfully, both of which specifically require that the "means for identifying" applies the selected new user rule to the attributes ***stored in memory***." Thus, while claims 11 and 32 recite identifying an object in "real time" it must still query or filter for the stored attributes. This is consistent with the '923 patent's disclosure that the detected attributes are first ***archived*** and event identification occurs in real time based on the detected attributes. *See id.*, 9:14-18. Nowhere in the specification indicates that event identification occurs in real time based on attributes that are not also stored or archived. Thus, in view of the '923 patent claims and disclosure storing the attributes is not inconsistent with real-time operation.

273. *Kellogg*'s system allows real-time scene monitoring, and implements various real-time operations. Ex. 1003, 10; Fig. 4-8 (annotated). In particular, *Kellogg* discloses that "it provides the keyword 'now' to signify a ***real-time query***, one that constantly polls the database for new information." *Id.*, 79. This is consistent with the '923 patent's real time detection and what the claims 11 and 32 require. *Kellogg* also describes that "[t]he input for the scene monitoring prototype comes from ***real-time*** processing of CCD camera images." *Id.*, 77. Thus, a POSITA would understand that *Kellogg's* system is capable of identifying an event of an object in real time.

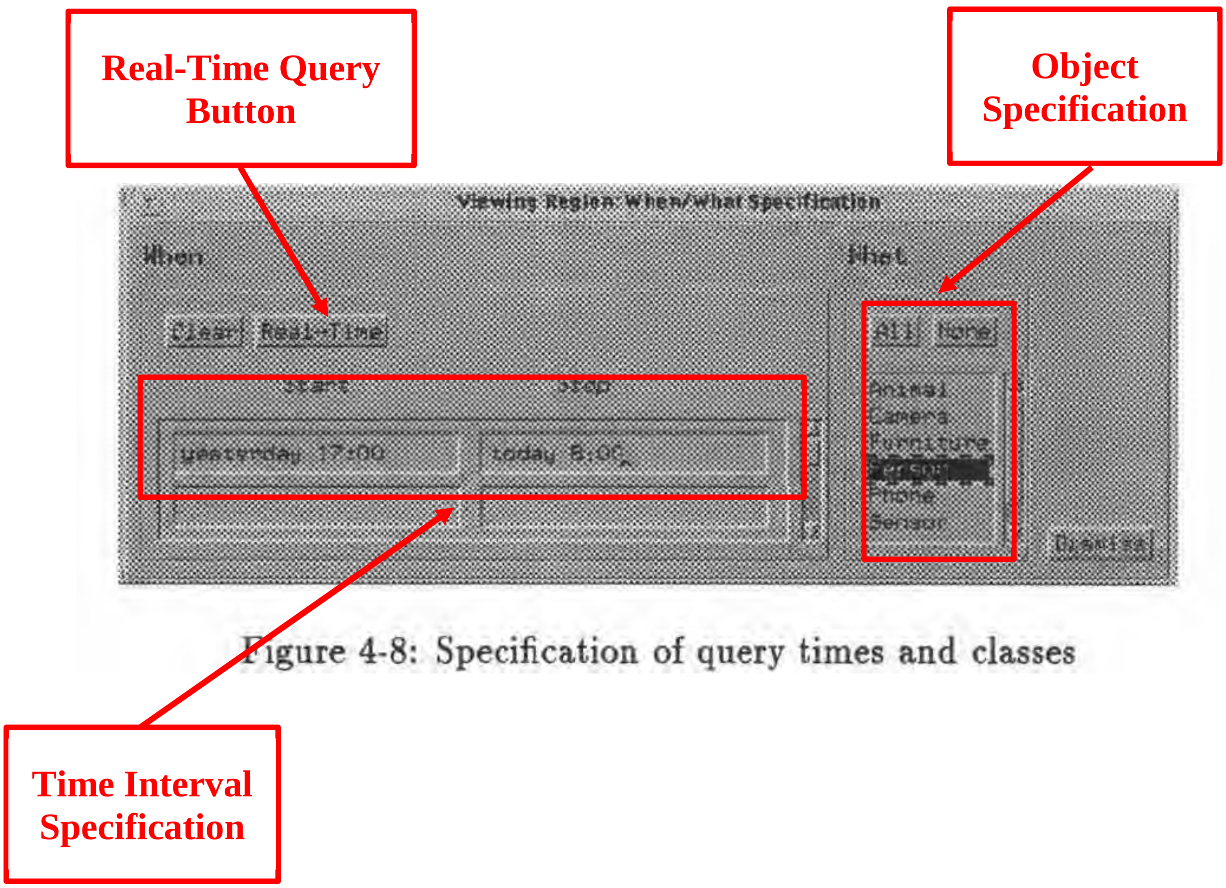


Figure 4-8: Specification of query times and classes

b. The events being identified

274. Claim 11 requires the limitation "means for identifying a first event of the object"; whereas claim 32 requires "means for identifying an event of the first object interacting with the second object." These limitations are disclosed as explained in Sections VIII(A)(6)(f) and VIII(A)(15)(f), respectively.

c. "analyzing, of the plurality of attributes, only a first selected subset of the plurality of attributes"

275. *Kellogg*'s Approach Query does not require analyzing all detected attributes as explained in Section VIII(A)(2), therefore it meets the limitation.

9. Claims 12 and 33

276. Claims 12 and 33 depend on claims 11 and 32, respectively. Claim 12 adds the limitation "the means for identifying an event of the object comprises means for identifying a second event of the object by analyzing, of the plurality of attributes, only a second selected subset of the plurality of attributes that have been archived." Claim 33 adds the limitation "the means for identifying an event of the first object interacting with the second object comprises means for identifying a second event by analyzing, of the plurality of attributes, only a second selected subset of the plurality of attributes which have been archived."

277. While claims 11 and 32 recite a means-plus-function limitation requiring the capability of identifying an event relating to an object in ***real-time***, these claims recite a means-plus-function limitation requiring the capability of

further identifying an event relating to that same object based on ***stored or archived*** attributes. The corresponding structure of this means-plus-function limitation in claims 12 and 33 is the structure identified in Sections VIII(A)(6)(f) and VIII(A)(15)(e), respectively, each of which performing the additional function of identifying an event based on ***stored or archived*** attributes. *See* Section V(E)(4).

278. *Kellogg* discloses this feature. For example, *Kellogg* discloses a scene monitoring system that alerts the user of an event in an alarm region. Ex. 1003, 78-79. As shown in annotated Fig. 4-8 below, the system provides a user interface for the user to set parameters to define queries to detect an event. *Id.* Specifically, the user has an option to select the "real-time query" button that would dictate the system to "constantly poll the database for new information." *Id.*; *see also id.*, 79 ("it provides the keyword 'now' to signify a real-time query").

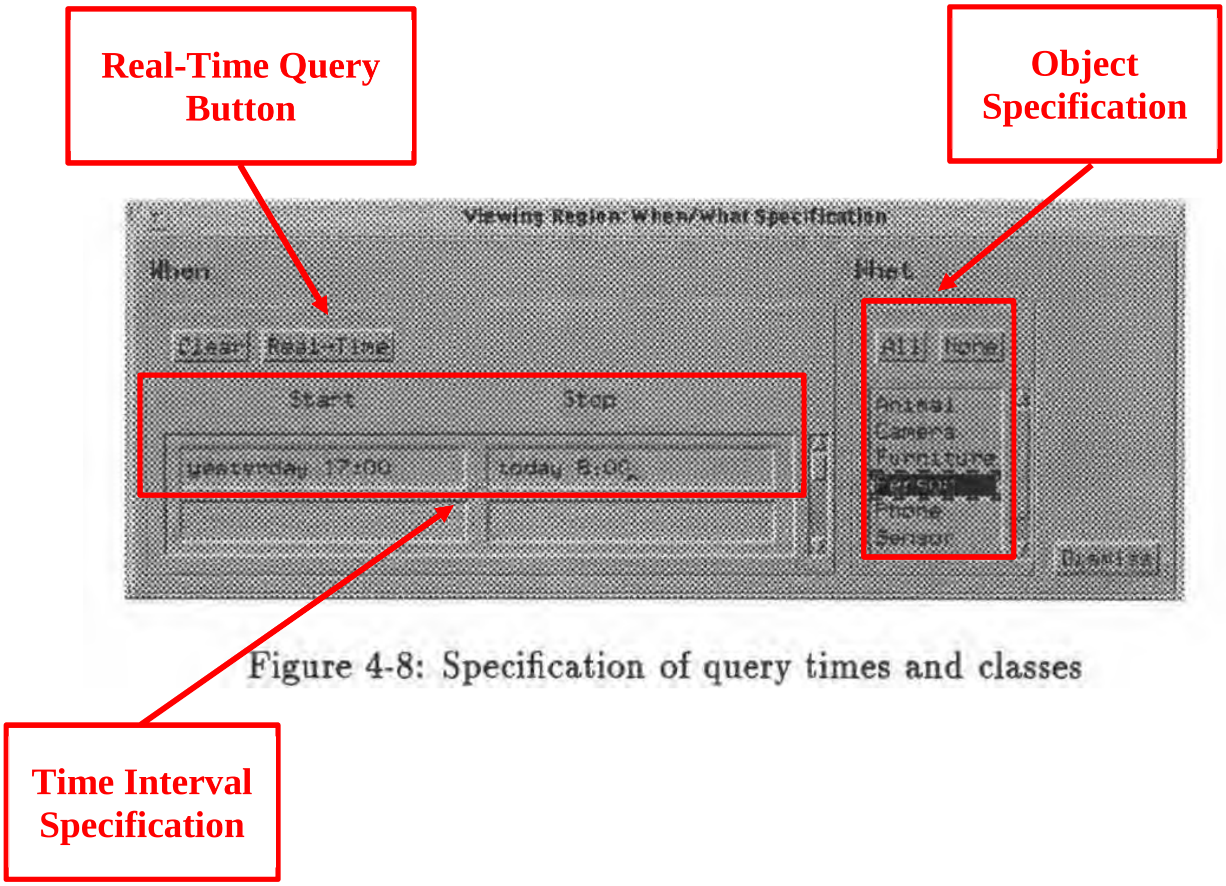


Figure 4-8: Specification of query times and classes

279. In contrast, when the button is not selected, the user can input specific time intervals such as from "yesterday 17:00" to "today 8:00" so that the system searches for stored or archived data of the alarm region. *Id.*, Fig. 4-8 (annotated); *see also id.*, 79. The system can still identify an event relating to that same object identified in the real-time query because the user can select the same object of interest using the Object Specification box on the right in the user interface. *Id.*, Fig. 4-8 (annotated). Accordingly, *Kellogg* discloses these claims.

10. Claims 14 and 35

280. Claims 14 and 35 depend from claims 9 and 30, respectively. These claims require the additional features of the memory being "configured to store at

least some of the plurality of attributes for at least two months." These claims also recite a means-plus-function limitation requiring analyzing some of the attributes that are stored for at least two months. The corresponding structure of this means-plus-function limitation in claims 14 and 35 is the same structure identified in Sections VIII(A)(6)(f) and VIII(A)(15)(e), respectively, each of which performing the additional function of analyzing some of the attributes that are stored for at least two months. *See* Section V(E)(4). Accordingly, the means-plus-function limitation is disclosed by *Kellogg*.

281. Claims 14 and 35 are apparatus claims directed to structures, e.g., memories. This reads on conventional computer non-volatile memory, which is designed to retain the stored information indefinitely, as users cannot enjoy the full benefit of a storage system if the data they have decided to retain is deleted without warning. These claims only require that the memory "is configured to store…for at least two months." This configuration is a functional requirement, which would be met by any conventional non-volatile memory at the time of the '923 patent. Accordingly, *Kellogg's* disclosed memory would meet this limitation as it is configured to store data indefinitely and therefore it is capable of performing the two-month storage function. *See, e.g.*, Ex. 1003, 54.

282. Nonetheless, *Kellogg* discloses the "configured to store at least some of the plurality of attributes for at least two months" feature. For example, *Kellogg*

discloses a scene monitoring system that allows a user to search past events by entering a specific time interval or a certain time point and an operator "…" to search for all information after that time point. Ex. 1003, 79. A user can input these temporal parameters using the interface as shown in annotated Fig. 4-8 below. The system does not limit the amount of time that a user can go back in time. *Kellogg* discloses a specific example where it searches information from "3/8/93 8:00" until "today 13:00." *Id.*, 79. Based on the publication date of *Kellogg* (May 1993) being two months after "3/8/93" a POSITA would have understood that *Kellogg* teaches the capability of storing attributes for at least two months.

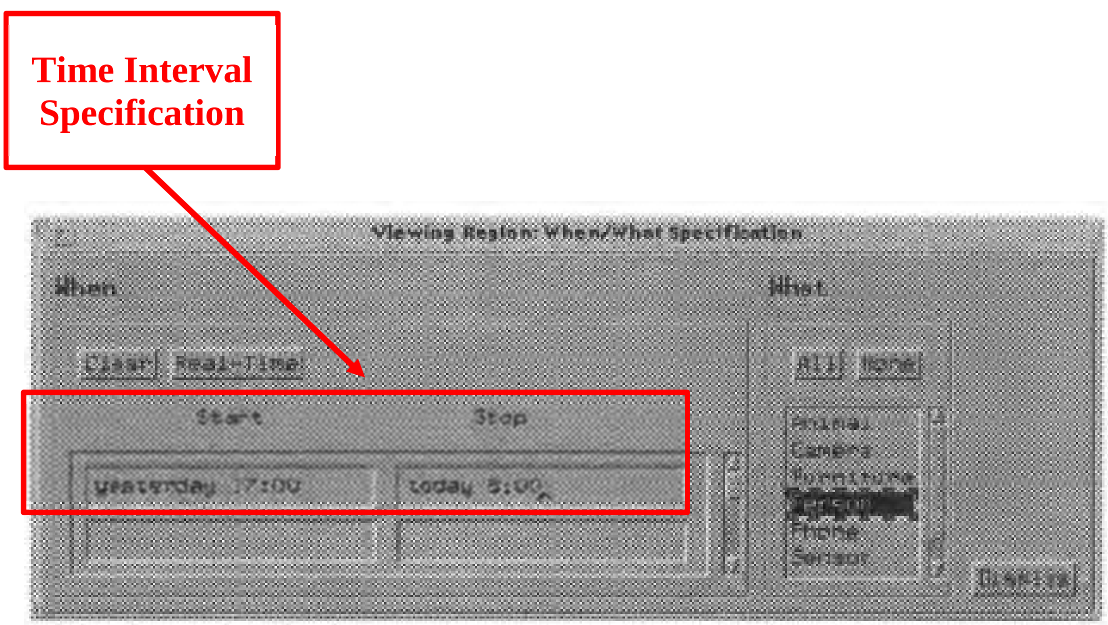


Figure 4-8: Specification of query times and classes

11. Claims 18 and 40

283. Claims 18 and 40 depend from claims 9 and 30, respectively, and add the feature "the video surveillance device is a computer system configured as a video surveillance device." The '923 patent specification describes that the purpose of a "video surveillance system … is for monitoring a location for, for example, market research or security purposes." Ex. 1001, 4:47-49.

284. *Kellogg*'s system is configured as a video surveillance device that can be used for security purposes. For example, *Kellogg* discloses, "[t]he primary goal of the thesis is to design a visual memory architecture that meets the requirements of ***various computer vision application***. A secondary goal is to implement a visual memory prototype to support a ***real-time scene monitoring prototype***." Ex. 1003, 10 (emphasis added). Further, *Kellogg* discloses "the ***scene monitoring system*** alerts the user to events in ***alarm regions*** … [which] can be established all over the map, allowing the user to monitor a number of disjoint regions without having to watch them all." *Id.*, 77-78. For each alarm region, users can enter "a delay specification that indicates how long an object must remain in that region before the system triggers an alarm." *Id.*, 78-79. *Kellogg*'s system can also display the video surveillance results using "playback, event report, and trail trace," where Figure 4-9 (below) "shows part of a playback window, with one object inside an alarm region and two other objects also being monitored." *Id.*, 80.

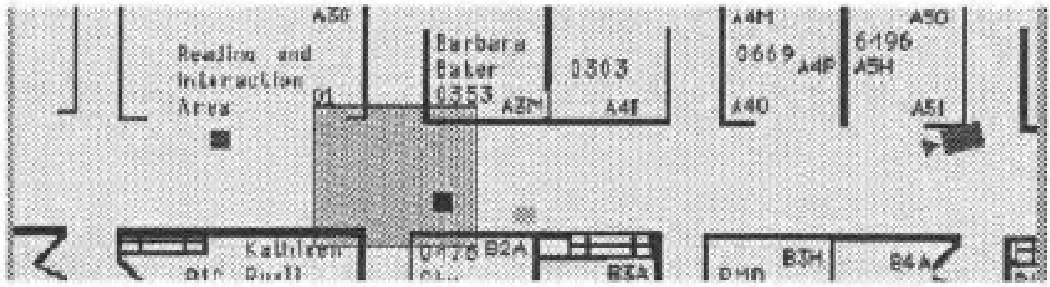


Figure 4-9: Graphical query results

12. Claims 19 and 41

285. Claims 19 and 41 depend from claims 9 and 30, respectively. *Kellogg* discloses the video device further comprising "video sensors," which include video cameras. Ex. 1001, 6:8-12. Figure 4-1 depicts *Kellogg*'s system using more than one video cameras.

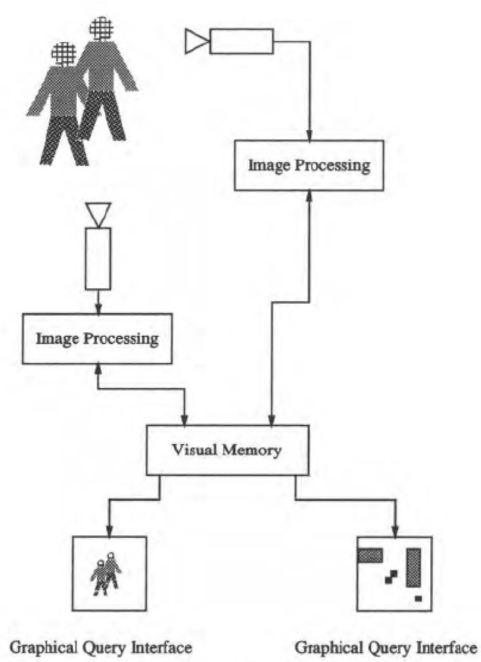


Figure 4-1: Scene monitoring prototype

13. Independent Claim 20

286. Claim 20 is substantially identical to claim 1. Claim 20 recites "at least a physical attribute and a temporal attribute," instead of reciting "at least one of a physical attribute and a temporal attribute" as in claim 1.

287. Claim 20 also recites "a combination of the attributes" instead of "a plurality of detected attributes." This does not meaningfully distinguish claim 20 from claim 1 because the recited combination does not refer to any specific combination, and the word "combination" simply implies that there is more than one as in "a plurality." For the reasons in Section VIII(A)(1), *Kellogg* anticipates claim 20.

288. *Kellogg* discloses the term "video device" for reasons explained in Section VIII(A)(6)(a).

14. Independent Claims 22 and 29

289. Claims 22 and 29 are essentially claims 1 and 8, respectively, in system form where they are directed to a "non-transitory computer-readable storage medium." A computer-readable-medium having software instructions stored therein would generally meet the added feature. *Kellogg's* scene monitoring prototype is a software (Ex. 1003, 77), and it is embodied in a computer readable medium, such as memory, hard drive or removable storage media. Accordingly,

these claims are anticipated by *Kellogg* for reasons set forth in Sections VIII(A)(1) and VIII(A)(5).

15. Independent Claim 30

290. Much like claim 9 as explained in Section VIII(A)(6), claim 30 recites means-plus-function elements ([30.1]-[30.2] and [30.4]-[30.7]) with functions substantially corresponding to the steps of claim 8.

a. A video device comprising

291. For the reasons in Section VIII(A)(6)(a), *Kellogg* discloses the preamble.

b. "means for detecting first and second objects in a video from a single camera" [30.1]

292. For the reasons in Sections VIII(A)(5)(a) and VIII(A)(6)(b), Kellogg discloses element [30.1].

c. "means for detecting a plurality of attributes of the object by analyzing the video from said single camera, each attribute representing a characteristic of the respective detected object" [30.2]

293. For the reasons in Sections VIII(A)(5)(b) and VIII(A)(6)(c), *Kellogg* discloses element [30.2].

d. "a memory storing the plurality of detected attributes" [30.3]

294. For the reasons in Sections VIII(A)(6)(d) and VIII(A)(3), *Kellogg* discloses this element.

e. "means for identifying an event of the first object interacting with the second object by applying a selected new user rule to the plurality of attributes stored in memory" [30.4]; "for identifying the event independent of when the attributes are stored in memory, the event not being one of the detected attributes" [30.5]; "wherein the applying the selected new user rule to the plurality of attributes stored in memory comprises applying the selected new user rule to only the plurality of attributes stored in memory." [30.6]

295. Compared to claim 9, claim 30 does not recite "and for identifying the event without reprocessing the video," so claim 9 is inclusive. Accordingly, for the reasons in Sections VIII(A)(5)(c)-(g) and VIII(A)(6)(e)-(f), *Kellogg* discloses elements [30.4]-[30.6].

f. "wherein the event of the object refers to the object engaged in an activity" [30.8]

296. For the reasons in Sections VIII(A)(5)(h) and VIII(A)(6)(g), *Kellogg* discloses this element.

16. Claim 36

297. Claims 36 depends from claim 30 and requires the additional feature of "wherein the means for identifying an event includes means for identifying the event without reprocessing the video." The corresponding structure of this means-plus-function limitation is the same structure identified in Section VIII(A)(15)(e), which performs the additional function. *See* Section V(E)(4). Accordingly, *Kellogg* discloses claim 36.

B. *Kellogg* in view of *Brill* Renders Claims 1-41 Obvious

1. Claims 1-7

298. To the extent Patent Owner argues that *Kellogg* does not disclose detecting an object in a video from a "single" camera as required in claim 1, *Brill* teaches such feature as demonstrated in Figure 10. Ex. 1004, 12.

Figure 10: A series of simple events

299. In addition, if Patent Owner argues that *Kellogg*'s disclosure is inadequate or a particular object detection algorithm is required (though not claimed), such as the background-model based technique described in the '923 patent (Ex. 1001, 9:39-48), *Brill* teaches such an object detection technique in detail (Ex. 1004, 6-9). Indeed, the '923 patent specification states that any change detection algorithm for detecting changes from a background can be used. Ex. 1001, 9:39-41. It would have been obvious to implement *Brill*'s object detection technique in *Kellogg*'s system for object tracking for reasons set forth in Section VI(E).

300. In addition, if Patent Owner argues that the recited "new user rule" requires setting a response, and that *Kellogg* does not disclose setting a response, *Brill* does so. For example, in selecting a new user rule for *Brill*, a user can set "actions" such as "beep," "log," or "popup," which are responses once an event is identified. Ex. 1004, Figure 11. Also, if Patent Owner argues that *Kellogg* does not disclose analyzing plural temporal attributes (*see* Section VIII(A)(4)) in the Approach Query, it would have been obvious for a POSITA to include another temporal attribute such as "days of week" in view of *Brill's* user interface to allow a user to further narrow down the search to specific days of the week. Ex. 1004, Figure 11.

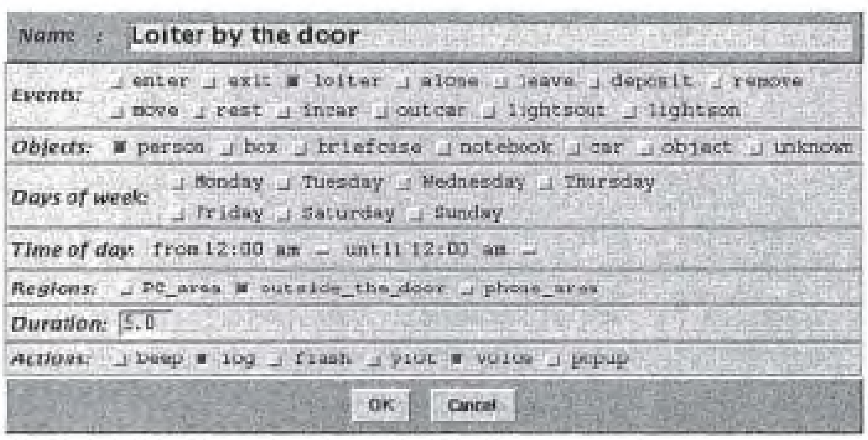


Figure 11: Selecting a type of simple event

301. A POSITA would have found it obvious to implement *Kellogg*'s system with the teachings of *Brill* to provide enhancements or achieve particular design objectives in *Kellogg*'s system, while yielding predictable results.

2. Claim 8

302. To the extent Patent Owner argues that *Kellogg* does not disclose detecting “first” and “second” objects in a video from a “single” camera in claim 8, *Brill* teaches this feature because its system detects a person (*i.e.*, a first object) and a briefcase object (*i.e.*, a second object) in a video from a single camera as demonstrated in Figure 10. Ex. 1004, 12; *see also* 6-9 (detecting people enter and exiting a car).

Figure 10: A series of simple events

303. In addition, if Patent Owner argues that *Kellogg* does not disclose the element “wherein the step of identifying an event of the object comprises identifying a first event of the first object interacting with the second object by analyzing the detected attributes of the first and second objects, the first event not being one of the detected attributes” in claim 8, *Brill* teaches this.

304. As discussed in Section VIII(5)(g), identifying two objects coming together meets a first object interacting with a second object. *Brill*’s system identifies a person picking up a briefcase and removing it from the scene, *i.e.*, the

first object interacting with the second object, and it analyzes the detected attributes such as the object type (*e.g.*, person, briefcase), activity attribute such as "remove," and the location of each object. Ex. 1004, 12-13; *see also* 6-9 (detecting people entering and exiting a car).

305. In addition, if Patent Owner argues that the recited "new user rule" requires setting a response, and that *Kellogg* does not disclose setting a response, *Brill* does so. For example, in selecting a new user rule for *Brill*, a user can set "actions" such as "beep," "log," or "popup," which are responses once an event is identified. Ex. 1004, Figure 11.

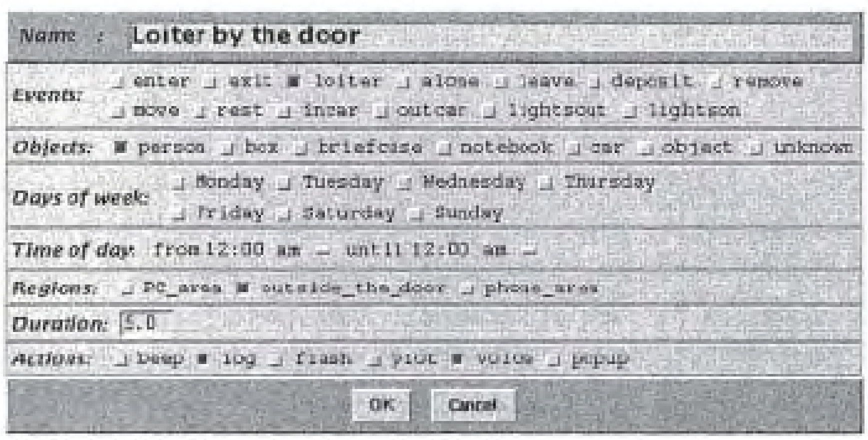


Figure 11: Selecting a type of simple event

306. A POSITA would have found it obvious to implement *Kellogg*'s system with the teachings of *Brill* to achieve particular design objectives in *Kellogg*'s system, such as monitoring interaction of two objects, while yielding predictable results.

3. Claims 9-19

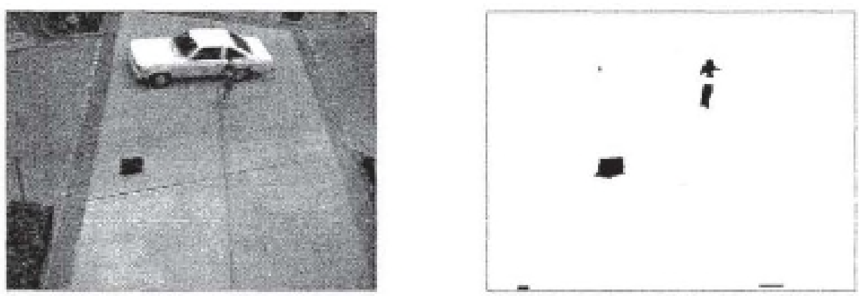
307. To the extent Patent Owner argues that a "video device" is not disclosed in *Kellogg*, *Brill* discloses smart cameras wherein the "attributes to be detected are defined in [the] device prior to the selection of a subset of the plurality of attributes." *Brill*'s AVS system is a distributed system where smart cameras analyze the video to detect objects and attributes of objects (including activity attributes). Ex. 1004, 5, Fig. 1. The smart cameras then transmit the attributes to the VSS, which analyzes the attributes and other data based on user-defined events (*i.e.*, user rules) and infers event occurrences. *Id.* The smart camera can also signal an alarm when an event occurs. *Id.*, 6.

308. In addition, if Patent Owner argues that *Kellogg* does not disclose the corresponding structure of the limitation "means for detecting an object in a video from a single camera" in claim 9, *Brill* teaches such structure.

309. A POSITA would understand the corresponding structure is a computer system or equivalent video processing system utilizing conventional motion and/or change detection algorithms to detect objects. *See* Section V(E)(1). Further, the '923 patent describes that the motion and change detection algorithms operate in parallel and can be performed in any order or concurrently. *Id.*, 9:31-33. The '923 patent specifically discloses that any such algorithm would work. *Id.*,

9:34-35, 39-41. And the '923 patent itself cites to other references for those algorithms. *Id.*, 9:35-37, 44-48.

310. *Brill* teaches a change detection algorithm that corresponds to "block 52" of the '923 patent, which could be "[a]ny change detection algorithm." Ex. 1001, 9:39-48. This algorithm detects an object "if one or more pixels in a frame are deemed to be in the foreground." *Id.* *Brill*'s system employs this algorithm as demonstrated in Figure 4(b) where the person is deemed to be in the foreground. Ex. 1004, 7.



(a) (b)

Figure 4: (a) Current video image. (b) Foreground difference image

311. *Brill* also teaches a motion detection algorithm that corresponds to "block 51" of the '923 patent. Ex. 1001, 9:33-38 ("Any motion detection algorithm for detecting movement between frames at the pixel level can be used for this block."). *Brill*'s system employs this algorithm as demonstrated in Figure 16 where the algorithm detects movement of the two people between frames at the pixel level. Ex. 1004, 15-16. *Brill* explains, "[t]he brightness [of Figs. 16(b) and

(d)] indicates the probability that the person's image intersects the given ***pixel***, which is highest in the middle of the region, and falls off towards the edge." Ex. 1004, 15.

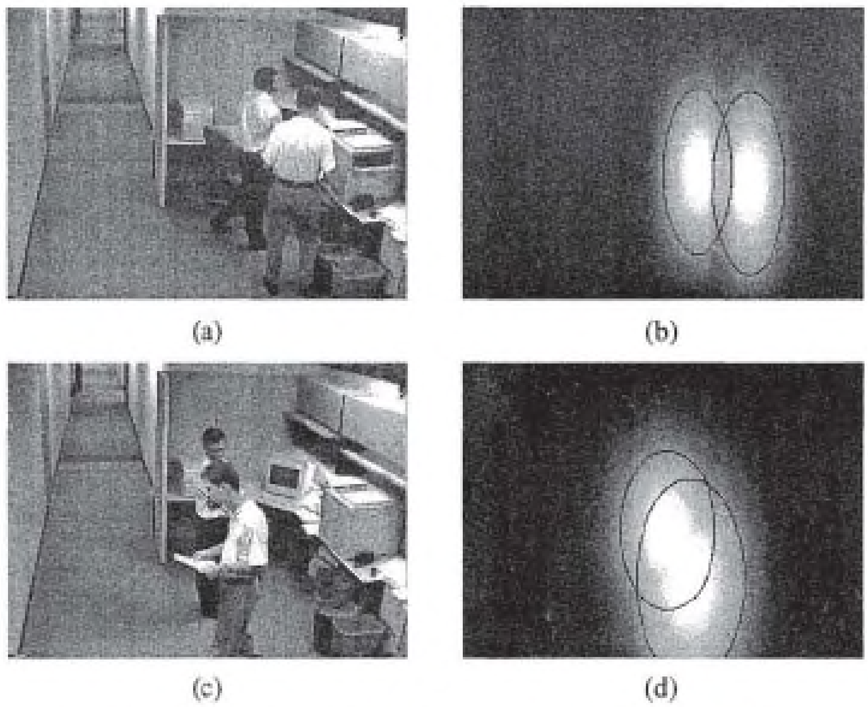


Figure 16: P-template images for partially occluding people

312. In addition, if Patent Owner argues that the recited "new user rule" requires setting a response, and that *Kellogg* does not disclose setting a response, *Brill* does so. For example, in selecting a new user rule for *Brill*, a user can set "actions" such as "beep," "log," or "popup," which are responses once an event is identified. Ex. 1004, Figure 11.

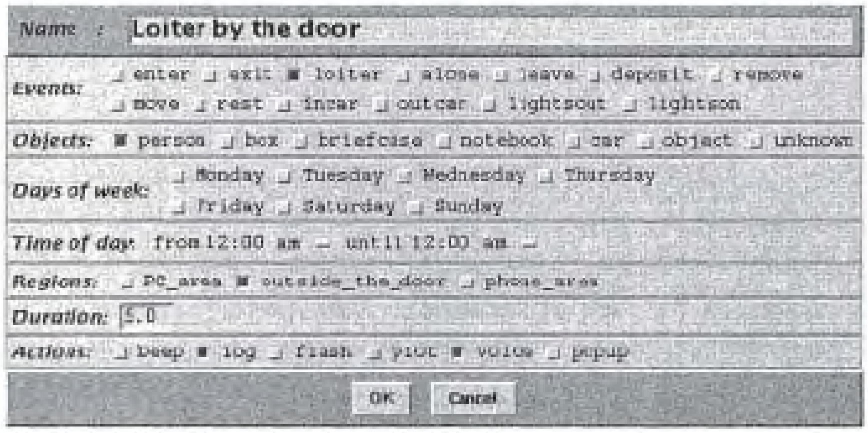


Figure 11: Selecting a type of simple event

313. A POSITA would have found it obvious to implement *Kellogg*'s system with the teachings of *Brill* to achieve enhanced object detection, while yielding predictable results.

4. Claims 12 and 33

314. To the extent Patent Owner argues that *Kellogg* does not disclose claims 12 and 33 these claims are obvious over the combination of *Kellogg* and *Brill*.

315. In addition to Sections VIII(B)(3) and VIII(B)(9), this Section applies to claims 12 and 33. Claims 12 and 33 depend on claims 11 and 32, respectively. Claim 12 adds the limitation "the means for identifying an event of the object comprises means for identifying a second event of the object by analyzing, of the plurality of attributes, only a second selected subset of the plurality of attributes that have been archived." Claim 33 adds the limitation "the means for identifying

an event of the first object interacting with the second object comprises means for identifying a second event by analyzing, of the plurality of attributes, only a second selected subset of the plurality of attributes which have been archived."

316. While claims 11 and 32 require the capability of identifying an event relating to an object in real-time, these claims require the capability of further identifying an event relating to that same object based on stored or archived attributes.

317. *Kellogg* expressly teaches the use of an object identifier, "OID," including searching based on OID. *See*, Ex. 1003, 50. Thus, once an object is identified in real-time, *Kellogg* could use the OID of the object to perform a second event search on that same object by using the OID.

318. Similarly, *Brill* explains that "the system monitors and records the movements of humans in its field of view," and that "[f]or every person that it sees, it creates a log file that summarizes important information about the person, including a snapshot taken when the person was close to the camera and (if possible) facing it." Ex. 1004, 18. "When the person leaves the scene, the log entry is saved to a file. Each log entry records the time when the person entered the scene and a list of coordinate pairs showing their position in each video frame." *Id*.; *see also id*. at Fig. 1:

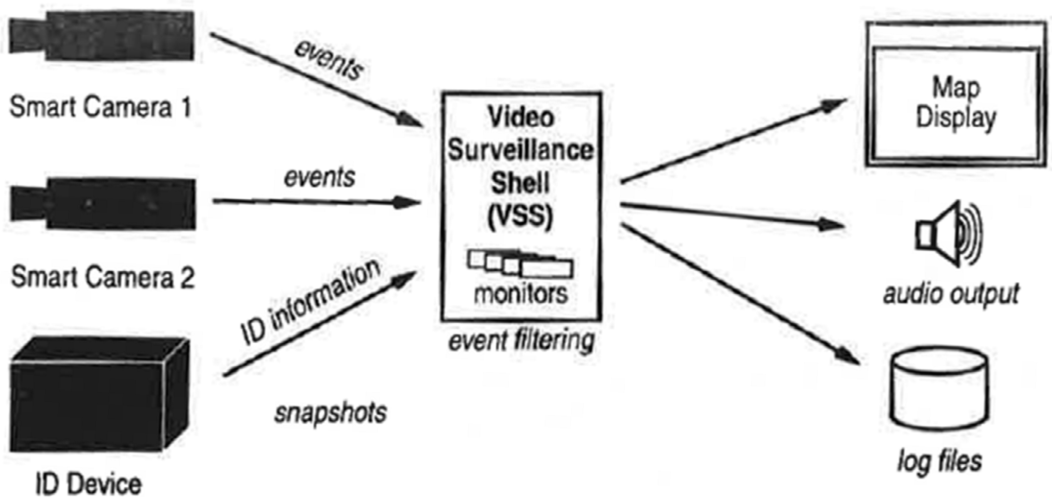


Figure 1: AVS system diagram

319. *Brill's* queries can then be run on the entries stored in the log files, with respect to a particular ID of the object that corresponds to a person of interest that was the subject of a real-time event. *Id*. at 11.

320. To the extent Patent Owner argues that the recited "second selected subset of the plurality of attributes" must be different from the "first selected subset" of claims 11 and 32[8], a "second event" will necessarily be associated with different attributes compared to the real-time "first event" because at the very least

[8] Nowhere in the '923 patent specification discloses a "second event" or a "second selected subset of the plurality of attributes," much less the "second selected subset of the plurality of attributes" being different from the "first selected subset."

the second event will be based on the particular ID of the object. Regardless, *Kellogg* teaches that the query used to determine each particular event can be a subset of attributes that has no relation to any other event. *See* Section VIII(A)(2).

321. The requirement of claim 33 that the second event relate to the first object interacting with the second object is, as discussed above with respect to claim 32, taught by both *Kellogg* and *Brill*. *See* Section VIII(A)(8).

5. Claims 14 and 35

322. In addition to Sections VIII(B)(3) and VIII(B)(9), this Section applies to claims 14 and 35. Claims 14 and 35 depend from claims 9 and 30, respectively. These claims recite the functional feature of the memory being "configured to store at least some of the plurality of attributes for at least two months." To the extent Patent Owner argues that the additional feature is not disclosed by *Kellogg*, these claims are still obvious over the combination of *Kellogg* and *Brill.*

323. As explained above with respect to claims 12 and 33, both *Kellogg* and *Brill* teach identifying events by analyzing stored attributes. *See* Section VIII(B)(4). Neither reference places any limit on how long these attributes can remain in storage prior to their analysis. This is because the amount of time information is stored is a trivial, non-technical matter of design choice, particularly with respect to an arbitrary, relatively short amount of time, such as two months.

324. Of course, conventional computer non-volatile memory is designed to retain the stored information indefinitely. Thus, a memory used in *Kellogg* and *Brill* would satisfy the claim limitation. The limitation does not depend on any technicality but rather on the user's intended use of the system. If the user decides to store the attributes for more than two months, the user can simply store the memory some place safe such that the stored attributes are not accidently deleted or overwritten.

325. Indeed, this limitation requires no technical change to the structures disclosed in *Kellogg* and *Brill. Kellogg* teaches a "playback" function where a user can view the monitoring results in a later time point. Ex. 1003, 80. *Kellogg* places no time limit on how long later point in time could be and it could be two months, three months or a year. Instead, all that is required to meet this limitation is that the user of such a system not delete the stored attributes that are collected for more than two months.

326. The ordinary knowledge of a POSITA would encourage storing attributes for at least two months in a video surveillance system like that disclosed in *Kellogg* and *Brill* designed for security purposes. Ex. 1003, 10, 77-78; Ex. 1004, 4-5. For example, a POSITA would know that many crimes can go easily unnoticed for more than two months. A POSITA would know also that repeat crimes can happen within a longer interval than two months. The POSITA's

knowledge would have been supplemented by *Dimitrova*'s teaching of using such surveillance systems in a museum setting. For example, *Dimitrova* discloses that its system can be used in a museum setting to identify someone who damaged or stole an expensive Miro painting. Ex. 1006, 11. A POSITA would be motivated to have the data stored for more than two months in the event that the information is needed for evidence in a prosecution, which could take more than two months.

327. Similarly, *Kellogg* identifies the utility of its system for sports data Ex. 1003, 53. It is well known that archived data about sports personnel is relevant for well over two months and should therefore be stored for a longer period of time. For example, in 1999 Kevin McHale was inducted into the Basketball Hall of Fame. Ex. 1039, 1. Mr. McHale's induction to the Hall of Fame was due to his time playing as a professional in the 1980s. *Id.* Thus, a POSITA, and any sports fan, would be readily motivated to store sports footage attribute data for more than two months, indeed for years, in order to readily identify the highlights of famous players, such as Kevin McHale, who was being celebrated for his achievements on the court over a decade later. *Id.*

6. Claims 18 and 40

328. In addition to Sections VIII(B)(3) and VIII(B)(9), this Section applies to claims 18 and 40. Claims 18 and 40 depend from claims 9 and 30, respectively, and adds the feature "the video surveillance device is a computer system

configured as a video surveillance device." To the extent Patent Owner argues that the additional feature is not disclosed by *Kellogg*, these claims are still obvious over the combination of *Kellogg* and *Brill*.

329. *Brill*'s system is called the "Autonomous Video Surveillance (AVS)" system that "processes live video streams from ***surveillance*** cameras to automatically produce a real-time map-based display of the locations of people, objects and events in a monitored region." Ex. 1004, 4 (emphasis added); *see also id.*, 5 ("AVS system incorporates multiple cameras to enable ***surveillance*** of a wider area than can be monitored via a single camera.") (emphasis added). The Video Surveillance Shell (VSS) in Figure 1 is a computer system that integrates the information from the surveillance cameras and displays it on a map. *Id.*, 4. Reading *Brill*, a POSITA would have readily configured *Kellogg's* system as a video surveillance device because *Kellogg* also contemplates its system to be used in scene monitoring. Ex. 1003, 77-80; *see* Section VIII(A)(11).

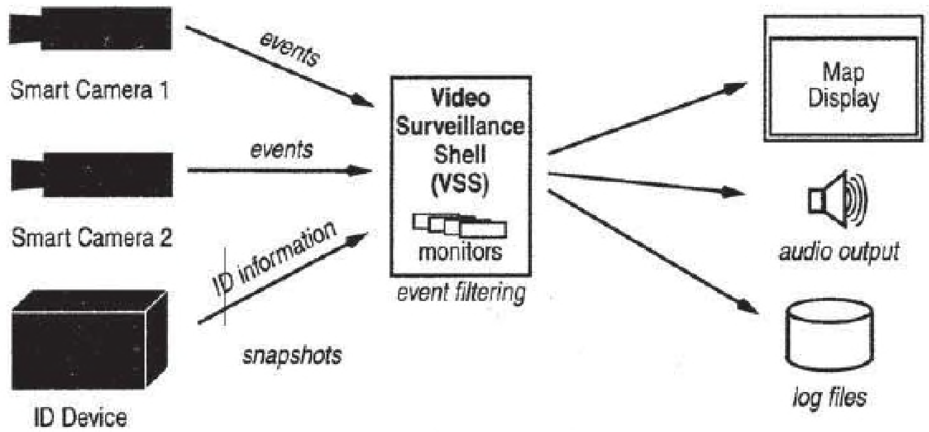


Figure 1: AVS system diagram

7. Claims 20-21

330. To the extent Patent Owner argues that a "video device" is not explicitly disclosed in *Kellogg*, *Brill* discloses smart cameras wherein the "attributes to be detected are defined in [the] device prior to the selection of a subset of the plurality of attributes." *Brill*'s AVS system is a distributed system where smart cameras analyze the video to detect objects and attributes of objects (including activity attributes). Ex. 1004, 5, Fig. 1. The smart cameras then transmit the attributes to the VSS, which analyzes the attributes and other data based on user-defined events (*i.e.*, user rules) and infers event occurrences. *Id.* The smart camera can also signal an alarm when an event occurs. *Id.*, 6.

331. In addition, if Patent Owner argues that *Kellogg* does not disclose "selecting the new user rule to provide an analysis of a combination of the plural physical attributes and the plural temporal attributes to detect the event," *Brill*

explicitly discloses this. For example, in selecting a new user rule, a user can select a "regions" and an "object type" (*i.e.*, together plural physical attributes), and a "days of week" and a "time of day" (*i.e.*, together plural temporal attributes). Ex. 1004, Figure 11.

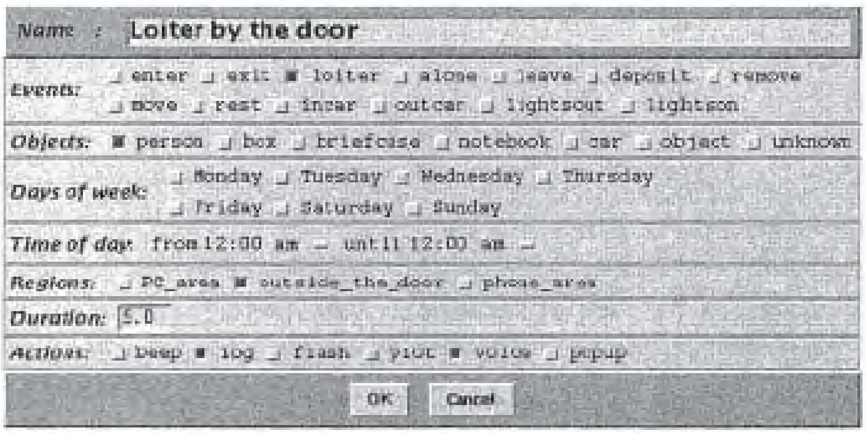


Figure 11: Selecting a type of simple event

332. In addition, if Patent Owner argues that the recited "new user rule" requires setting a response, and that *Kellogg* does not disclose setting a response, *Brill* does so. For example, in selecting a new user rule for *Brill*, a user can set "actions" such as "beep," "log," or "popup," which are responses once an event is identified. Ex. 1004, Figure 11.

333. It would have been obvious to implement *Kellogg's* system to detect events based on the attributes taught by *Brill*, in view of the overlapping teachings (including the detection of temporal and physical attributes in both references).

8. Claims 22-29

334. Claims 22 and 29 are essentially claims 1 and 8, respectively, in a system form directed to a "non-transitory computer-readable storage medium." The added feature does not add any technicality to claims 1 and 8, and it would have been obvious to a POSITA at the priority date to have a computer-readable-medium having software instructions stored therein.

335. To the extent Patent Owner argues *Kellogg* lacks a disclosure of a "non-transitory computer-readable storage medium," *Brill* teaches it. *Brill's* AVS system is a "software" (Ex. 1004, 6), and it is embodied in a computer readable medium, such as memory, hard drive or removable storage media.

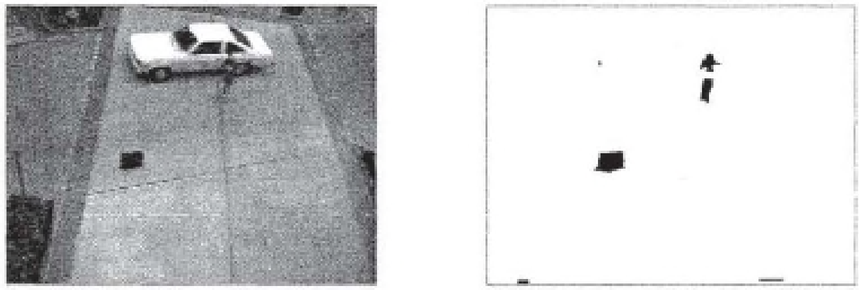
9. Claims 30-41

336. To the extent Patent Owner argues that a "video device" is not explicitly disclosed in *Kellogg*, *Brill* discloses smart cameras wherein the "attributes to be detected are defined in [the] device prior to the selection of a subset of the plurality of attributes." *Brill*'s AVS system is a distributed system where smart cameras analyze the video to detect objects and attributes of objects (including activity attributes). Ex. 1004, 5, Fig. 1. The smart cameras then transmit the attributes to the VSS, which analyzes the attributes and other data based on user-defined events (*i.e.*, user rules) and infers event occurrences. *Id.* The smart camera can also signal an alarm when an event occurs. *Id.*, 6.

337. If the Patent Owner argues that *Kellogg* does not disclose the corresponding structure of "means for detecting first and second objects in a video from a single camera" in claim 30, *Brill* teaches such structure.

338. A POSITA would understand the corresponding structure is a computer system or equivalent video processing system utilizing conventional motion and/or change detection algorithms to detect objects. *See* Section V(E)(1). Further, the '923 patent describes that the motion and change detection algorithms operate in parallel and can be performed in any order or concurrently. *Id.*, 9:31-33. The '923 patent specifically discloses that any such algorithm would work. *Id.*, 9:34-35, 39-41. And the '923 patent itself cites to other references for those algorithms. *Id.*, 9:35-37, 44-48.

339. *Brill* teaches a change detection algorithm that corresponds to "block 52" of the '923 patent, which could be "[a]ny change detection algorithm." Ex. 1001, 9:39-48. This algorithm detects an object "if one or more pixels in a frame are deemed to be in the foreground." *Id.* *Brill*'s system employs this algorithm as demonstrated in Figure 4(b) where the person is deemed to be in the foreground. Ex. 1004, 7.



(a) (b)

Figure 4: (a) Current video image. (b) Foreground difference image

340. *Brill* also teaches a motion detection algorithm that corresponds to "block 51" of the '923 patent. Ex. 1001, 9:33-38 ("Any motion detection algorithm for detecting movement between frames at the pixel level can be used for this block."). *Brill*'s system employs this algorithm as demonstrated in Figure 16 where the algorithm detects movement of the two people between frames at the pixel level. Ex. 1004, 15-16. *Brill* explains, "[t]he brightness [of Figs. 16(b) and (d)] indicates the probability that the person's image intersects the given ***pixel***, which is highest in the middle of the region, and falls off towards the edge." Ex. 1004, 15.

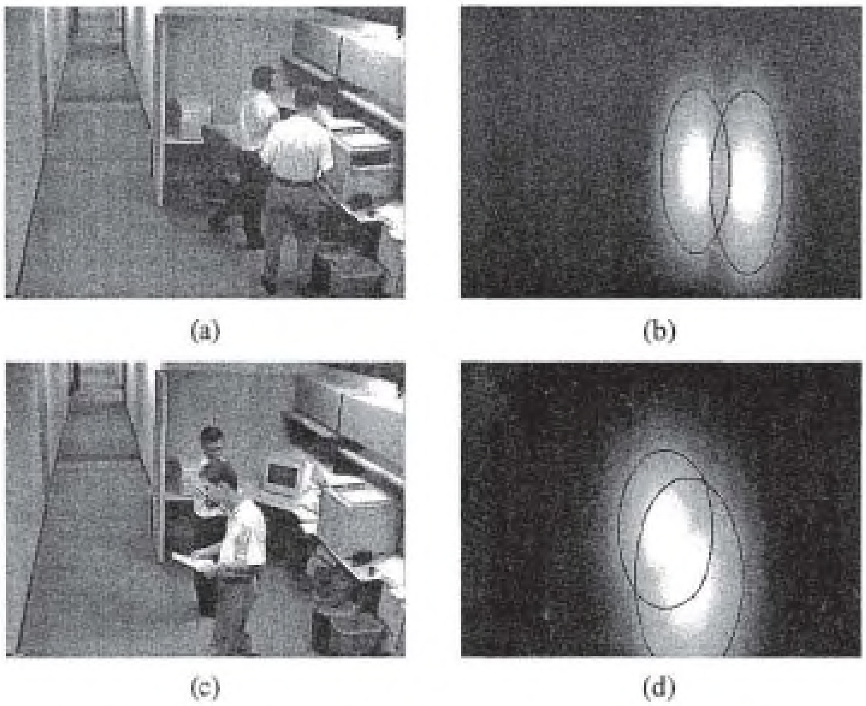


Figure 16: P-template images for partially occluding people

341. In addition, if Patent Owner argues that the recited "new user rule" requires setting a response, and that *Kellogg* does not disclose setting a response, *Brill* does so. For example, in selecting a new user rule for *Brill*, a user can set "actions" such as "beep," "log," or "popup," which are responses once an event is identified. Ex. 1004, Figure 11.

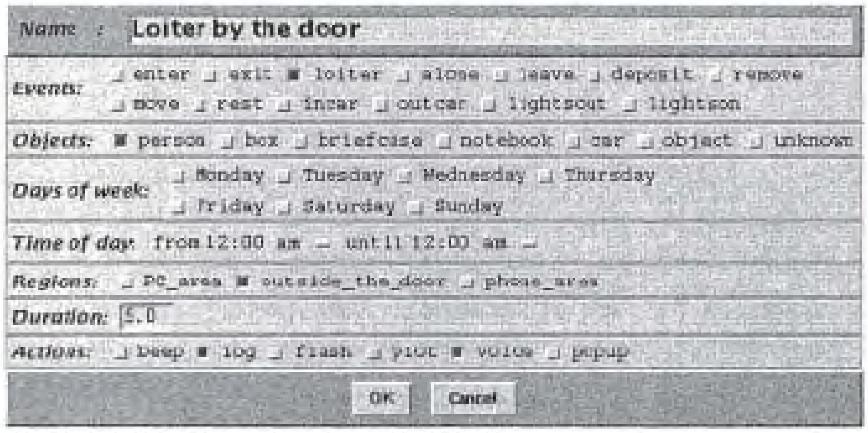


Figure 11: Selecting a type of simple event

342. Lastly, if Patent Owner argues that *Kellogg* does not disclose "means for identifying an event of the first object interacting with the second object by applying a selected new user rule to the plurality of attributes stored in memory," *Brill* teaches it. As discussed in Section VIII(A)(5)(g), identifying two objects coming together meets a first object interacting with a second object. *Brill*'s system identifies a person picking up a briefcase and removing it from the scene, *i.e.*, the first object interacting with the second object, and it analyzes the detected attributes such as the object type (*e.g.*, person, briefcase), activity attribute such as "remove," and the location of each object. Ex. 1004, 12-13; *see also* 6-9 (detecting people entering and exiting a car).

343. A POSITA would have found it obvious to implement *Kellogg*'s system with the teachings of *Brill* to achieve enhanced object detection, while yielding predictable results. Accordingly, claims 30-41 are rendered obvious by *Kellogg* in view of *Brill*.

C. Dimitrova in view of Brill Renders Claims 1-41 Obvious

1. Independent Claim 1

a. "detecting an object in a video from a single camera" [1.1]

344. As shown in Fig. 5, the '923 patent describes that object detection (51, 52) occurs before attribute detection (57). Ex. 1001, 9:30-10:57. The '923 patent provides two embodiments used in detecting objects (*e.g.*, detecting objects via motion (51) and detecting objects via change (52)), both of which take into consideration that objects in a scene move or change over time. *Id.*, 9:33, 9:39, Fig. 5. But, '923 the patent does not purport to have invented object detection. In fact, the '923 patent admits that many object detection methods were known. *Id.*, 9:35-38, 9:44-48. The '923 patent does not limit the object detection to a specific algorithm and allows a person skilled in the art to use "any" such algorithm available. *Id.*, 9:34-35, 9:39-41. Moreover, the claim limitation does not even require using any algorithm. The limitation merely requires that an object is detected and nothing more.

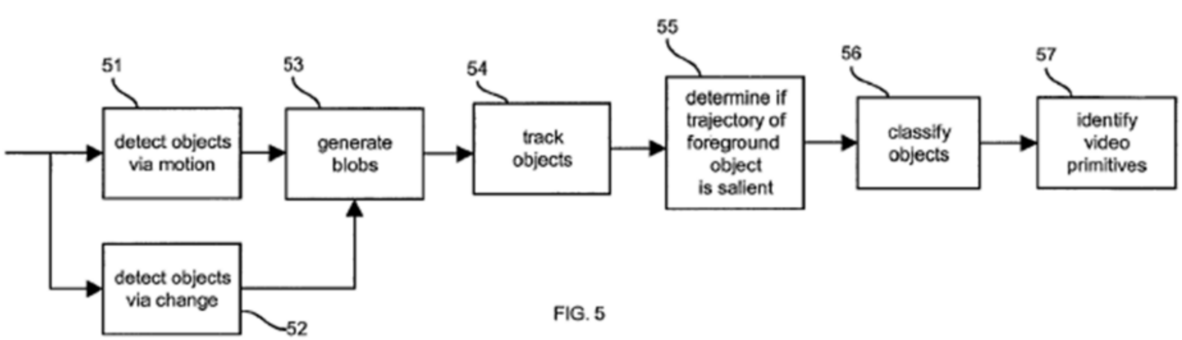


FIG. 5

345. *Dimitrova* discloses detection of objects and their motion in a video obtained from a single camera. For example, Figure 4 shows detecting and tracing a moving yacht in three frames of video sequences. *See also* Ex. 1006, 8-19. In another example, Figure 5 shows multiple frames of a small toy in motion from a single camera, and Figure 6 shows the detection and tracking result of the moving toy. *Dimitrova* also describes that its "***object recognition*** ideas have been influenced by" several prior art methods for detecting objects. *Id.*, 8 (emphasis added).

346. *Dimitrova* capitalizes on the fact that object detection often relies on motion detection, looking for changes in the pixels between frames of a video to detect a moving object in the camera's view as separate from the static background. The trajectories, or direction of change, of clusters of pixels can be calculated. *Id.*, 13. For rigid objects, clusters of pixels with similar trajectories can be grouped and identified as an object. For non-rigid objects (*e.g.*, a person), more complicated algorithms are needed, but *Dimitrova*'s technique of using motion to detect an object remains the same. *Id*. This method of detecting objects is an accurate way to detect objects because spurious pixel changes are ignored (*e.g.*, changing light conditions, etc.) while objects of interest are identified.

347. To the extent the Patent Owner argues that the limitation somehow requires detecting an object from a single camera that would be an incorrect

reading because the limitation only requires that an object is detected in a ***video***. Also, such a reading would be inconsistent with the specification where the only disclosed embodiment is that the "computer system 11 obtains source video" from the camera to further process it for object detection and attribute extraction. *Id.*, 9:23-31.

b. "detecting a plurality of attributes of the object by analyzing the video from said single camera, the plurality of attributes including at least one of a physical attribute and a temporal attribute, each attribute representing a characteristic of the detected object" [1.2]

348. The '923 patent specification does not particularly define or use the term "physical attribute." The word "physical" is only used once in the specification where it describes that a "physical subject" can be an "object." Ex. 1001, 3:27-29. A POSITA would understand that the word "physical" means that it can be perceived through the senses, *i.e.*, observable (because the system uses a video sensor). *See* Ex. 1037, 3. The specification uses the language "observable attributes" and includes, classification, size, shape, color, texture, position, velocity, speed, internal motion, motion, salient motion, feature of a salient motion, scene change, feature of a scene change etc., all of which are observable characteristics of an object. *See, e.g., id.*, 7:6-8:15. Event discriminators are described in terms of these attributes. *Id.*, 7:5:6.

349. The '923 patent specifically lists "position," "position of an object in image space as a function of time," and "an approximate position of an object in a three-dimensional representation of the environment as a function of time" as an observable attribute. *Id.*, 7:9, 7:51-54. Thus, these attributes are physical attributes.

350. The term "physical attribute" can also include "spatial attributes." The '923 patent describes that spatial attributes include crossing a line, entering an area, or crossing a line from the left. *Id.*, 8:65-67. These characteristics require referring to an object's position in image space as a function of time.

351. *Dimitrova* discloses detecting a plurality of attributes of the object by analyzing the video. For example, the system uses "motion analysis" techniques on the video source to detect basic activity attributes of an object such as strolling, walking or hurrying. Ex. 1006, 17, 25. *Dimitrova*'s activity attributes are consistent with the activity and/or motion attributes (*e.g.*, entering, exiting, stopping, appearing, disappearing or moving from one place to another) disclosed in the '923 patent, some of which need to be detected over a period of time. Ex. 1001, 3:30-33, 7:8-10; 7:37-46, 63-67.

352. *Dimitrova* teaches at least one physical attribute, including an object's shape, object skeleton, centroid, and bounding box in the multiresolution hierarchy of Figure 7. Ex. 1006, 18-20. Other physical attributes include: category (*e.g.*,

"person" or "pet") (Ex. 1006, 19); relative size (*e.g.*, "big") (*id.*), color (e.g., "brown") (*id.*); parts representing a human figure, *e.g.*, head, torso, arms and legs (*id.*, 20).

353. *Dimitrova* also discloses that a trajectory attribute can be detected by "trac[ing] object through 20 encoded frames." Ex. 1006, 13. A trajectory of an object is also considered a physical attribute—"a feature of a salient motion"—in the '923 patent. Ex. 1001 7:8-10; 7:47-49.

354. *Dimitrova* can also detect internal motion activity attributes such as "waving," much like the '923 patent. Ex. 1006, 25; Ex. 1001, 7:8-10, 7:34-36 ("video primitives include … an internal motion" such as "person having swinging arms and legs").

355. The term "temporal attributes" is also not defined in the specification of the '923 patent. Nonetheless, examples of the term include "every 15 minutes," "between 9:00 p.m. to 6:30 a.m.," "less than 5 minutes," "longer than 30 seconds," "over the weekend," and "within 20 minutes of." Ex. 1001, 8:32-36. It does not have to be in a particular form.

356. One example of temporal attributes in *Dimitrova* is the video frame numbers, namely a "first frame" and a "last frame," referring to the first and last frames in which the object is detected. Ex. 1006, 20, 32. A frame number represents a particular moment in time in a video, and the frame numbers advance

as time advances. Indeed, *Dimitrova* refers to these as "time pointers (*e.g.*, the frame counter)." Ex. 1006, 32. Based on the disclosure of *Dimitrova*, a POSITA may at once envisage any number of queries that apply to a plural number of the attributes. For example:

- Find "all objects" that enter a scene at "time X" and leave at "time Y," in other words, objects whose first frame is time X and whose last frame is time Y (temporal attributes "first frame" and "last frame").
- Find all objects that did not leave or remained in the scene between "time X" and "time Y," in other words, objects whose first frame is before time X and whose last frame is after time Y (temporal attributes "first frame" and "last frame").Using the frame numbers, all objects that enter the scene at the time of "first frame" and exit the scene at the time of "last frame," or all objects that did not leave or remained in the scene between "first frame" and "last frame" can be identified.

357. The frame numbers also serve as temporal attributes in identifying events that involve multiple objects at the same time. For example, a user may request "retriev[al of] all the sequences in which a tall person is waving while the president walks," which hereinafter is referred to as the "Parade Query." Ex. 1006, 25. This query identifies an event that can be called a "parade" event. The

"parade" event comprises recognizing characteristics of multiple objects (1) any tall people waving; and (2) while the president walks. Many tall people waving can be detected, and each of those objects has an associated first and last video frame in which they appear. Some people could be waving before or after the president walks. So in order to identify the intended "parade" event those frame numbers must be compared to the frame numbers associated with the president walking. *Dimitrova's* system employs operators called "Temporal Combination Functions," which includes operators like "before, meets, simultaneously, starts, finishes." Ex. 1006, 25. The Parade Query applies an operator "while" to the first and last frame of each object, *i.e.*, plural temporal attributes, to search for temporal concurrency of tall persons waving and the president walking. Accordingly, *Dimitrova*'s video frame numbers are temporal attributes.

358. Moreover, a POSITA would understand that "temporal" is not just limited to absolute time, but instead it also encompasses relative time or time intervals. A POSITA would know also that *Dimitrova*'s video frame numbers can be used to infer relative time or time intervals based on the frame rate, explained as follows. The frame rate, which is usually expressed as the number of frames per second, is known by design or by a setting in the system. In a typical video image acquisition system, as a POSITA would know, frames are incremented at the frame rate by a digital clock, internal to the system, that sequences operation. The time

between successive frames is the inverse of the frame rate, and may be expressed as seconds or milliseconds per frame. In typical applications, the digital clock is driven by an electronic oscillator that emits a stable train of pulses to establish timing, similarly as in any digital clock such as a digital wrist watch. This pulse train is divided down in frequency by digital circuitry to generate timing pulses at the frame rate. These pulses are used to initiate the acquisition of each frame. Thus, sequential frame numbers correspond to time intervals between frames. For example, the ’923 patent’s temporal attribute “less than 5 minutes” (Ex. 1001, 8:35) can be equally searched in *Dimitrova’s* system using frame numbers (i.e. frame counts) and the frame acquisition rate for the video stream. The increase in frame number between the beginning and end of a time interval (i.e. the change in frame count from the “first frame” to the “last frame”) is simply calculated, as a POSITA would know, by multiplying the specified time interval by the frame rate. For example, if the frame rate for a given system is, say, 10 frames per second, then a time interval of 5 minutes (i.e. 5x60=300 seconds) would correspond to 300x10=3000 frames. By noting the frame number at the beginning of the interval to be timed (“first frame”), it would be known that 5 minutes had passed in the video stream when the frame number reaches the starting frame number plus 3000. Conversely, the time interval between two frames is calculated as the difference between the frame numbers of the two frames divided by the frame rate. In the above example, for a frame rate of 10 frames per second, the time interval in the video acquisition stream between two frames that are 3000 frames apart is 3000/10=300 seconds, or 5 minutes.

359. *Dimitrova*'s physical and temporal attributes represent characteristics of the detected object because the physical attributes are observable traits of the detected object and the temporal information of a detected object is always associated with the object in a video.

360. To the extent Patent Owner argues that *Dimitrova's* frame numbers do not adequately disclose temporal attributes, *Brill* does as disclosed in the '923 patent. Ex. 1001, 8:32-36 ("every 15 minutes," "between 9:00 p.m. to 6:30 a.m.," "less than 5 minutes," "longer than 30 seconds," "over the weekend," and "within 20 minutes of."). For example, a "loiter by the door" event can "triggered when a person loiters in the area near the door for ***more than 5 seconds***." Ex. 1004, 13 (emphasis added). This event is defined using plural physical attributes such as "person" and "near the door," an activity attribute "loiter" and a temporal attribute "more than 5 seconds." *Id.* As a result, there is at least one of a physical attribute and a temporal attribute. In addition, *Brill's* system can detect events based on "plural attributes," including "plural physical attributes" (Ex. 1004, 11 (*e.g.*, person, box, car, region, etc.)); "plural temporal attributes" (*id.*, 13, Figure 11 (*e.g.*, time of day, day of week)); and "a physical attribute and a temporal attribute" (*id.*).

It would have been obvious to employ the additional temporal attributes such as "time of day" in detecting the "loiter by the door" event such that the event is narrowed to a certain time of day. This would have resulted in analyzing plural temporal attributes.

361. In another example, an "arrival at work" event could be defined using the Figure 11 interface to select "person" and "briefcase" objects "outside the door" between "8:50 am and 9:10 am" on "Monday" morning. *Id.* The last two attributes are temporal attributes, consistent with the temporal attributes disclosed in the '923 patent. Ex. 1001, 4:63-5:5.

Figure 11: Selecting a type of simple event

c. "selecting a new user rule after detecting the plurality of attributes" [1.3]

362. The language "selecting a new user rule" is not used in the '923 patent specification. Indeed, the '923 patent does not even use the word "rule" in this

context. However, the ’923 patent discloses, “[d]uring tasking, the operator ***selects*** an area representing the space around the desired retail display.” Ex. 1001, 15:26-28 (emphasis added). And the patent further describes that “[a]s a discriminator, the operator ***defines*** that he or she wishes to monitor people-sized objects that enter the area.” *Id.*, 15:28-32 (emphasis added). The patent appears to use the words “select” and “define” interchangeably in the context of a user rule.

363. *Dimitrova* discloses a system tasking mechanism to allow users to create *ad hoc* queries based on plurality of attributes that are already detected. Ex. 1006, 20 (“operators may be used in a relational form, mostly in a table lookup mode, or may be embedded into a more-elaborate query language”); *see also id.*, 28-29 (“The visual query given in Figure 10 will select those video sequences from the repository in which the player's trajectory is similar to the one drawn by the user”).

364. Specifically, the Parade Query specifies a combination of set of detected activity, physical and temporal attributes for identifying a “parade” event. For example, the query searches detected activity attributes such as “waving” and “walking,” and physical or object associated attributes such as “tall,” “president” and “person.” Ex. 1006, 25. As explained in the previous element, the query also employs parameters such as “while” together with frame numbers to detect temporal concurrency.

365. To the extent Patent Owner argues that the "new user rule" requires setting a response (as it did in the Related IPRs), this is incorrect because neither the claims nor the specification require one. The '923 patent unequivocally deems it "***optional***." Ex. 1001, 8:37, 8:56-58 (emphasis added). The scope of a "response" in the '923 patent is very broad. *Id.*, 8:37-49. For example, it includes activating a visual or audio alert on the system display, directing the computer system, or even saving data to a medium. *Id.* These responses do not require alerting the user in any way. A POSITA would understand that even a simple return of a query result is consistent with a "response" as used in the patent.

366. Nevertheless, *Dimitrova* teaches that its computer-based system can be used in video surveillance where it describes that "[a]pplications such as ***automated surveillance*** may require retrieval of either video sequences or objects contained in these sequences based on the object trajectories." *Id.*, 11 (emphasis added); *see also id.*, 27 ("The automation of this whole process is possible for strictly limited application domains such as industrial ***monitoring***, domain specific video editing, camera ***surveillance***, and others.") (emphasis added). A POSITA would have been motivated to optionally include a response to be triggered when the rule identifies an event because it would better serve the purpose of a video surveillance system. And a POSITA would have looked at *Brill* which specifically teaches such feature. For example, in selecting a new user rule for *Brill*, a user can

set "actions" such as "beep," "log," or "popup," which are responses once an event is identified. Ex. 1004, Figure 11.

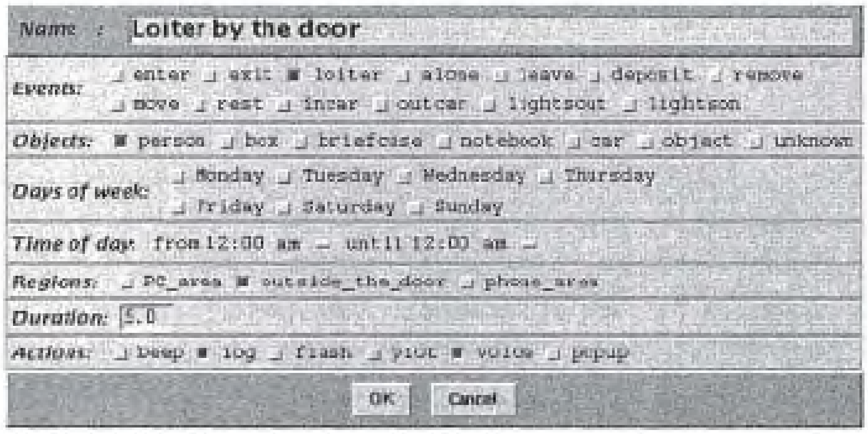


Figure 11: Selecting a type of simple event

367. A POSITA would have found it obvious to implement *Dimitrova*'s system with the teachings of *Brill* to provide enhancements or achieve particular design objectives in *Dimitrova*'s system, while yielding predictable results. *See* Section VI(F).

d. "after detecting the plurality of attributes and after selecting the new user rule, identifying an event of the object that is not one of the detected attributes of the object by applying the user rule to the plurality of detected attributes," [1.4a]

368. As discussed in Section V(C)(2), the limitation "identifying an event of the object that is not one of the detected attributes of the object" means "the user defined event comprises a minimum of two attributes." In *Dimitrova*, the Parade Query identifies a "parade event." And in doing so, the query applies two or more attributes, *e.g.*, "waving," "walking," "tall," "president," "person" and four frame

numbers (two from the tall person and two from the president), from the plurality of detected attributes. Ex. 1006, 25.

369. It is also clear that the Parade Query is used to identify events ***after*** the plurality of attributes are detected and ***after*** the new user rule defining the Parade Query is selected because the Parade Event is created after-the-fact and applied to existing recorded attributes. Ex. 1006, 33 and 34 (disclosing after-the-fact created queries that search archived attributes extracted from NBA footage).

370. To the extent Patent Owner argues that *Dimitrova*'s activity attributes "walking" or "waving" are user-defined events, this is wrong for the reasons discussed in Section V(C) regarding Argument (3). "Walking" or "waving" are not events because the claimed event is required to be applied to a minimum of two detected attributes. Moreover, the user rules disclosed in *Dimitrova* define events that are far more complex than the single activity attributes of "walking" or "waving." The Parade Query includes the ***two*** activity attributes, walking and waving, each applied to a different object. And it further includes the specific classification attributes of person, tall, and president. And it further includes a temporal attribute requiring that the two objects meet the specified criteria at the same time. Defining a user rule comprising a complex collection of attributes the "Parade Query" meets the limitations of the claim.

371. It is clear that the query is tasked after the plurality of attributes are detected and after creating the query because the entire concept of *Dimitrova* is to record basic attributes so that a search can later be created based on any arbitrary subset of the recorded attributes. *See, e.g.*, Ex. 1006, 24-25, 29.

e. "wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes" [1.4b]

372. As discussed in Section V(D), the limitation "the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes" requires searching only the detected attributes. Prior art, for example, that ***always*** require searches to process object-oriented abstractions rather than just the attributes themselves would not disclose this limitation. This interpretation of the limitation is based solely on the prosecution history as explained in Section V(D).

373. *Dimitrova* discloses this limitation by providing a query operator called "exact" (using the symbol "!"). Ex. 1004, 21. This operator allows the system to retrieve only objects that have exactly the same detected attributes as the user specified in the query. *Id.* This operator prevents the system from returning higher-level abstractions based on the queried attributes, which would merely represent approximations of the searched attributes.

374. Furthermore, while *Dimitrova* discloses schemas that define logical relationships between objects and attributes, those schemas do not define the storage structure of the data in the *Dimitrova* database. Ex. 1006, 22-23. Rather, *Dimitrova* merely teaches that the attributes are stored as OMV triplets, without requiring any further hierarchy. *Id.*, 19. Thus, *Dimitrova* can apply its user rules to only the attributes, *e.g.*, the OMV triplet data, and does not require searching that involves higher-level abstractions like the *Day* references Patent Owner was trying to distinguish with this limitation.

f. "wherein the plurality of attributes that are detected are independent of which event is identified" [1.5]

375. As discussed in Section V(C), this limitation corresponds to Patent Owner's Argument (2), the proper construction of which requires that the event detection process does not alter the attribute detection process. *Dimitrova*'s system meets this limitation because it allows a user to create sophisticated *ad hoc* event definitions based on various basic attributes that were previously collected by the system and stored in the database. *See, e.g.,* Ex. 1006, 25 ("Using the above combinators and the OMV structure, many new types of queries that refer to the contents of video sequences can be specified"); *see also id.*, 29 ("The visual query given in Figure 10 will select those video sequences from the repository in which the player's trajectory is similar to the one drawn by the user and display the name

and the position of the player.”). These new event definitions do not and cannot alter the attribute detection process.

376. For example, in the context of the Parade Query, the attributes collected are not based on, or effected by, the tasking of the Parade Query. Instead, *Dimitrova* records basic activity attributes by analyzing the video and the Parade Query is defined ***later*** using those pre-existing attributes. Ex. 1006, 25. This is demonstrated by the fact that *Dimitrova* uses many of the same basic attributes to define completely different user rules, such as the “Pet Query,” which is defined as “retrieve all the video sequences in which a pet walks and makes a trajectory $t_{1,}$.” *Id.*, 20. This Pet Query uses the activity attribute “walking” from the Parade Query but it further uses the “pet” classification and a “trajectory.” If *Dimitrova* attributes were not “independent” of the identified event, the system would not be reusing “walking” attributes in each query.

377. Moreover, by mixing attributes, a user can identify a completely new arbitrary event such as a Walking the Dog, which comprises the attributes a person walks while a pet walks. Defining queries in this way, as disclosed in *Dimitrova*, allows a multitude of arbitrary events that could be defined based on *Dimitrova’s* collected attributes. Importantly, none of these new event definitions will affect the attribute detection process because the attributes are basic attributes the system

collects in any type of video source prior to a user selecting or creating a new query based on a combination of those collected attributes.

378. The independence of *Dimitrova's* attributes from its event is further shown by the exemplary event "retrieve objects that have a motion trajectory whose point of origination is the main gallery door and terminate at the Joan Miro's picture on the opposite wall" defined by *Dimitrova*. Ex. 1006, 11. The hypothetical here is that the Miro painting has been damaged and the museum wants to figure out who did it. *Id.* Of course, there was no indication in advance that the Miro would be damaged, so the system was not preset to watch out for this hypothetical event. Rather, *Dimitrova* describes an after-the-fact query to allow identification of anyone with trajectories to the Miro painting. The trajectory of each object, its origination point, and its termination point are previously collected by the system no matter what kind of event is identified by the user. *Id.*

379. Patent Owner would be wrong if it continues to argue that *Dimitrova*'s Parade Query merely references "an already determined/detected event with respect to variably selected locations and/or time," as it did in the related IPRs (IPR2018-00140, Paper No. 11, 36 (September 4, 2018)). Even if one were to incorrectly consider "walking" an event, the Parade Query requires much more than merely indexing walking to a time or location. As discussed above, the Parade Query relies on multiple activity attributes, "walking" and "waving," of

multiple classified objects, a tall person and the president. Even the simple Pet Query, which relies on a pet object classification attribute, the walking activity and a trajectory t_1 attribute, includes far more than the single activity attribute "walking." Thus, it is clearly an event.

380. Finally, while this limitation should not be construed to require independence from knowledge or regard for a predefined list of events, see Section V(C), *Dimitrova* has no such list.

g. "wherein the step of identifying the event of the object identifies the event without reprocessing the video" [1.6]

381. *Dimitrova* discloses that events are identified by querying a database of attributes rather than processing the entire video. Ex. 1006, 24-27 (applying a retrieval predicate to the OMV triplets stored in the database). Because the attributes are already stored in the database *Dimitrova's* query does not require or involve reprocessing the video.

h. "wherein the event of the object refers to the object engaged in an activity" [1.7]

382. *Dimitrova* discloses this element because its Parade Query identifies a "parade" event by searching for activity attributes "waving" and "walking". Both objects of the Parade Query are engaged in an activity as demonstrated by these activity attributes.

2. Claims 2, 4, 7[2][9], 13, 16, 23, 25, 28[2], 34 and 38

383. *Dimitrova* discloses that the queries do not require all detected attributes as recited in claims 2, 4, 7[2], 13, 16, 23, 25, 28[2],34 and 38.

- "selecting the new user rule comprises selecting a subset of the plurality of attributes for analysis" (claim 2 [depends on claim 1], claim 23 [depends on claim 22])
- "no analysis is performed on at least some of the detected attributes to detect an event" (claim 4[depends on claim 1])
- "analyzing only a subset of the attributes stored in the memory" (claim 7[2] [depends on claim 1], claim 28[2] [depends on claim 22])
- "analyzing, of the plurality of attributes, only a selected subset of the plurality of attributes" (claim 13 [depends on claim 9], claim 34 [depends on claim 30])
- "analyzing a selection of individual ones of the detected plural attributes" (claim 16 [depends on claim 9], claim 38 [depends on claim 30])
- "do not cause the computer system to perform an analysis on at least some of the detected attributes to detect an event" (claim 25 [depends on claim 22])

384. *Dimitrova* explains that object attributes such as color (e.g., "brown") (Ex. 1006, 19) or other motion attributes such as trajectory, velocity, torsion, etc. (*id.*, 20) are detected, but none of these are analyzed in the Parade query. The same is true for the Pet Query or the Miro example in Section VIII(A)(1)(e) – none

[9] [2] refers to the second element of the claim.

of them require analysis of the color, trajectory, velocity, or torsion attributes. Therefore, the queries do not require all detected attributes.

3. Claims 3, 7[1][10], 17, 24, 28[1] and 39

385. *Dimitrova* discloses the additional features of claims 3, 7[1] 17, 24 and 28[1]. The additional features added by these claim limitations all require that the detected attributes are stored in the system, and some of these claims further require that it happens prior to selecting a set of attributes to define a user rule.

- "the plurality of attributes that are detected are defined in a device prior to a selection of a subset of the plurality of attributes" (claim 3 [depends on claim 1], claim 24 [depends on claim 22])
- "storing the detected plurality of attributes in memory" (claim 7[1] [depends on claim 1], claim 28[1] [depends on claim 22])
- "the plural attributes detected by the means for detecting are defined in the video device independent of a selection of the detected plural attributes" (claim 17 [depends on claim 9], claim 39 [depends on claim 30])

386. *Dimitrova* discloses a "general architecture for video ***database retrieval***" and explains that "[t]he extracted spatial and motion characteristics are ***stored*** [as] conceptual data" for later retrieval. Ex. 1006, 24 (emphasis added); *see also id.*, 26-27. *Dimitrova's* "Insertion Module" extracts attributes including "basic spatial properties," "generic activities descriptions," "trajectories," and

[10] [1] refers to the first element of the claim.

"motions," and stores them in the "Video Storage Server" which is part of a video device. Ex. 1006, Fig. 9. A POSITA would have understood that the arrow pointing from the Insertion Module to the Video Storage Server demonstrates this process. The Video Storage Server is "a disk array serving as a repository." Ex. 1006, 27. Thus, it meets recited memory of claims 7 and 28. The later retrieval process occurs through a separate module called the "Interactive Query Module," which independently queries the previously-stored attributes. Ex. 1006, 26-29; Figure 9; *see also id.*, 21.

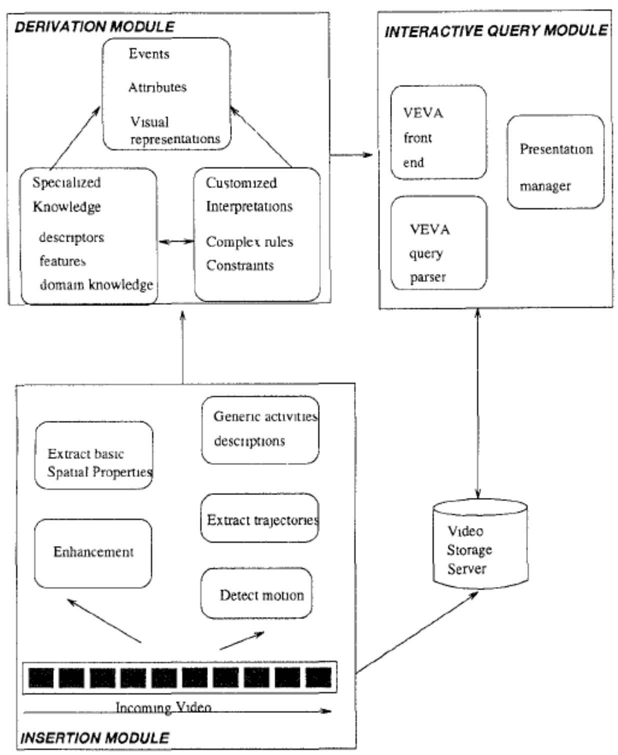


Fig. 9. An architecture for video classification and retrieval.

387. *Dimitrova* further explains that EVA, the query language that was the basis for the prototype VEVA system, "is the interface to a multimedia database

system capable of *storage*, retrieval, management, analysis, and delivery of objects of various media types, including text, audio, images, and moving pictures." *Id.*, 21 (emphasis added). The VEVA prototype queries databases storing video attributes. *Id.*, 27-29; Figure 9. As shown in Figure 10 (reproduced below), *Dimitrova's* VEVA prototype included databases (*i.e.*, stored data) of attributes for basketball videos, travel videos, and others.

4. Claims 5, 6, 15, 21, 26, 27 and 37

388. The additional features of claims 5, 6, 15, 21, 26, 27 and 37 require that the new user rule is applied to plural physical and/or plural temporal attributes.

- "plurality of attributes include plural physical attributes and the method comprises applying the new user rule to a plural number of physical attributes" (claim 5 [depends on claim 1], claim 26[11] [depends on claim 22])
- "plurality of attributes include plural temporal attributes and the method comprises applying the new user rule to a plural number of temporal attributes" (claim 6 [depends on claim 1], claim 27[12] [depends on claim 22])

[11] No meaningful difference between claims 5 and 26.

[12] No meaningful difference between claims 6 and 27.

- "analyzing at least two selected physical attributes of the plurality of attributes" (claim 15 [depends on claim 9], claim 37[13] [depends on claim 30])

389. The additional features applying the new user rule to more than one physical attributes, and/or more than one temporal attributes. In *Dimitrova*, the Parade Query "retrieve[s] all the sequences in which a tall person is waving while the president walks." Ex. 1006, 25. The query identifies a "parade event." *Id.* And in doing so, the query is applied to the following attributes: "waving," "walking," "tall," "president," "person" and four frame numbers from the plurality of detected attributes. *Id.*

390. As discussed in Section VIII(C)(1)(b), the '923 patent's physical attribute includes, among other things, ***classification***, ***size***, shape, color, texture, position, velocity, speed, ***internal motion***, motion, salient motion, feature of a salient motion, scene change, feature of a scene change etc., all of which are observable characteristics of an object.

391. The Parade Query analyzes a plurality of physical attributes. For example, the query analyzes object classification attribute "person" and a relative

[13] Claim 37 recites the limitation in a means-plus-function format. The corresponding structure is the structure identified in Section VIII(A)(15)(e), which performs the additional function. *See* Section V(E)(4).

size attribute "tall." The query also analyzes internal motion attributes "waving" and "walking," much like the '923 patent. Ex. 1006, 25; Ex. 1001, 7:8-10, 7:34-36 ("video primitives include … an internal motion" such as "person having swinging arms and legs").

392. The Parade Query also analyzes a plurality of temporal attributes. The four frame numbers associated with the two objects (first and last frames for each object) serve as temporal attributes for reasons explained in Section VIII(C)(1)(b).

393. Accordingly, the Parade Query analyzes plural physical and plural temporal attributes.

394. To the extent Patent Owner argues that *Dimitrova's* frame numbers do not adequately disclose temporal, *Brill* does as explained in Section VIII(C)(1)(b). In addition, *Brill* teaches analyzing plural physical and/or plural temporal attributes as explained in Section VIII(C)(1)(b). Accordingly, the combination of *Dimitrova* and *Brill* discloses the limitations.

5. Independent Claim 8

a. "detecting first and second objects in a video from a single camera" [8.1]

395. *Dimitrova* discloses detecting an "object" in a video obtained from a single camera for reasons explained in Section VIII(C)(1)(a) for limitation [1.1].

Compared to limitation [1.1], this limitation requires detecting "first and second objects," which is also disclosed by *Dimitrova*.

396. *Dimitrova* explains, "we need the ability to classify ***objects*** appearing in ***a video sequence*** based on their characteristics and features such as shape or color, as well as ***their*** movements." Ex. 1006, 3 (emphasis added); *see also* 11 ("Applications such as automated surveillance may require retrieval of either video sequences or ***objects*** contained in these ***sequences*** based on the object trajectories.) (emphasis added). The plural words emphasize that more than one objects are detected.

397. *Dimitrova* also teaches this element as demonstrated by Figs. 4-6. Specifically, in Fig. 5, a "small toy" (highlighted in yellow), *i.e.*, first object, is detected and traced for each frame in the sequence as shown by the "traced trajectory" in Fig. 6. Ex. 1006, 13. While both cups in Fig. 5, *i.e.*, second and third objects, appear stationary those are still detected because *Dimitrova* "keep[s] track of the zero motion…to describe stationary objects" from the scene. *Id.*, 11. As shown in the frames in Fig. 5, the small toy and the cups are all detected in a video, sent from a single camera.

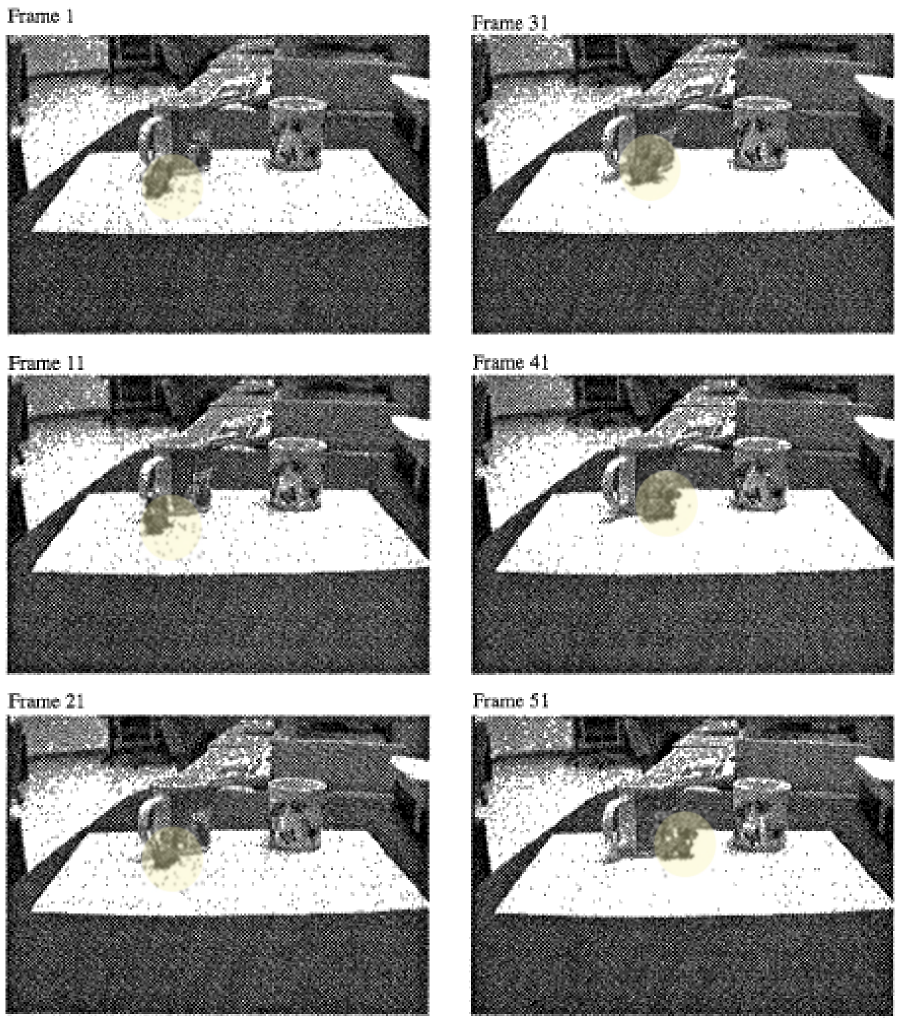
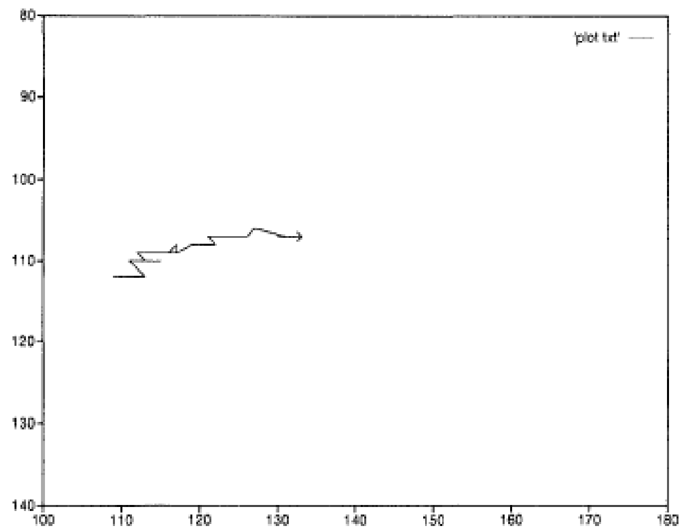


Fig. 5. Snapshots from a video sequence.

(a) Trace of the macroblock (14,14) using all frames



(b) Trace of the macroblock (14,14) using every other frame

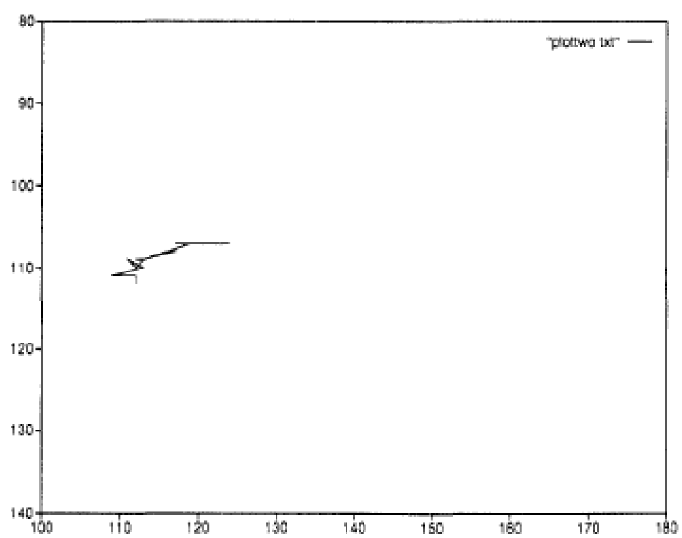


Fig. 6. Traced trajectories in the Walfky video sequence. (a) all frames; (b) only every other frame is used.

398. *Dimitrova*'s Parade Query is another example where first and second objects, *i.e.*, the tall person and the president, have been detected in a video. Ex. 1006, 25.

399. To the extent Patent Owner argues that *Dimitrova* does not disclose detecting "first" and "second" objects in a video sent from a "single" camera, *Brill* teaches this element. For example, *Brill's* system detects a person (*i.e.*, a first object) entering a monitored area and a briefcase (*i.e.*, a second object) being picked up and removed from the scene in a video from a single camera as demonstrated in Figure 10. Ex. 1004, 12.

Figure 10: A series of simple events

400. *Brill* also teaches detecting people entering and exiting a car in a video from a single camera. *Id.*, 6-9, Figure 6.

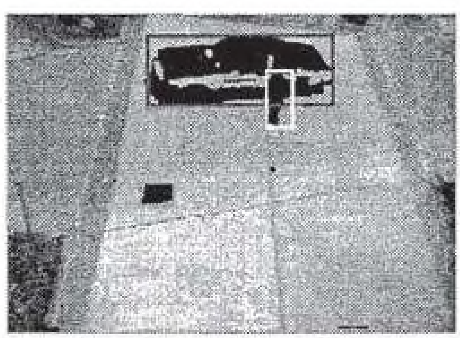


Figure 6: Final output of system

b. "detecting a plurality of attributes of each of the detected first and second objects by analyzing the video from said single camera, each attribute representing a characteristic of the respective detected object" [8.2]

401. As explained in Section VIII(C)(1)(b), the combination of *Dimitrova* and *Brill* discloses this element. Specifically, *Dimitrova*'s Parade Query demonstrates that a plurality of attributes of the first and second objects, *i.e.*, attributes of the person waving and the president, are already detected.

c. "selecting a new user rule" [8.3]

402. As explained in Section VIII(C)(1)(c), the combination of *Dimitrova* and *Brill* discloses this element.

d. "after detecting the plurality of attributes, identifying an event that is not one of the detected attributes of the first and second objects by applying the new user rule to the plurality of detected attributes," [8.4a]

403. As explained in Section VIII(C)(1)(d), *Dimitrova* discloses this element.

e. "wherein the applying the new user rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes" [8.4b]

404. As explained in Section VIII(C)(1)(e), *Dimitrova* discloses this element.

f. "wherein the plurality of attributes that are detected are independent of which event is identified" [8.5]

405. As explained in Section VIII(C)(1)(f), *Dimitrova* discloses this element.

g. "wherein the step of identifying an event of the object comprises identifying a first event of the first object interacting with the second object by analyzing the detected attributes of the first and second objects, the first event not being one of the detected attributes" [8.6]

406. The specification of the '923 patent does not specifically disclose identifying an event of a first object interacting with the second object by analyzing the detected attributes of the first and second objects. Nor does the specification define the term "interacting."

407. *Dimitrova* discloses this element. The Parade Query identifies a first event where a tall person (*i.e.*, the first object) "interacts" with the president (*i.e.*, the second object) by waving while the president walks. Ex. 1006, 25. This event looks for an interaction between the person and the president because the person is acting upon, *i.e.*, waving, while the president walks. This is identified by analyzing the detected attributes of the first object, namely "person," "tall" and "waving," that are temporally concurrent with the attributes of the second object, namely "president" and "walking."

408. *Dimitrova* discloses the limitation "the first event not being one of the detected attributes" as explained in Section VIII(C)(1)(d) with respect to the corresponding limitation "identifying an event of the object that is not one of the detected attributes of the object."

409. To the extent Patent Owner argues that *Dimitrova* does not disclose "identifying a first event of the first object ***interacting*** with the second object," *Brill* does. For example, *Brill*'s system identifies an event of a person picking up a briefcase and removing it from the scene, *i.e.*, the first object interacting with the second object, and it analyzes the detected attributes such as the object type (*e.g.,* person, briefcase), activity attributes such as "enter," "remove" and "exit." Ex. 1004, 12-13; *see also* 6-9 (person exiting and entering a car). A POSITA would have readily implemented this embodiment in *Dimitrova* because *Dimitrova* contemplates video surveillance and monitoring. *See*, *e.g.*, Ex. 1006, 6, 11, 27.

410. Although *Brill* labels the individual activity attributes of "enter," "remove," and "exit" as "events," they are not "events" as defined by the Patent Owner here. *See* Section V(C)(2), Argument (3). At a minimum, there must be two attributes to define an event. *Id.* But, those attributes are merely single activity attributes, much like the "motion" attributes disclosed in the '923 patent which includes "appearance of an object, disappearance of an object, a vertical movement of an object, a horizontal movement of an object…." Ex. 1001, 7:37-

40. Accordingly, *Brill's* disclosure of "enter," "remove," and "exit" are each activity *attributes* that are combined to define the complex event.

h. "wherein the event of the object refers to the object engaged in an activity" [8.7]

411. As explained in Section VIII(C)(1)(h), *Dimitrova* discloses this element. To the extent *Dimitrova* does not disclose "identifying a first event of the first object ***interacting*** with the second object," *Brill* does as explained in Section VIII(C)(5)(g). *Brill's* event of a person picking up a briefcase and removing it from the scene involves multiple activity attributes as explained in Section VIII(C)(5)(g). Accordingly, *Brill's* event refers to the object engaged in an activity.

6. Independent Claim 9

412. Claim 9 recites means-plus-function elements ([9.1]-[9.2] and [9.4]-[9.8]) with functions substantially corresponding to the steps of claim 1.

a. A video device comprising

413. *Dimitrova* discloses a video classification and retrieval system (*i.e.*, a "video device") that includes three modules—the insertion, derivation, and interactive query modules. Ex. 1006, 25-27, Fig. 9. *Dimitrova*'s system is a computer-based system and necessarily includes a computer or processing device with software or a set of instructions to perform the recited functions of the means-plus-function elements of claim 9.

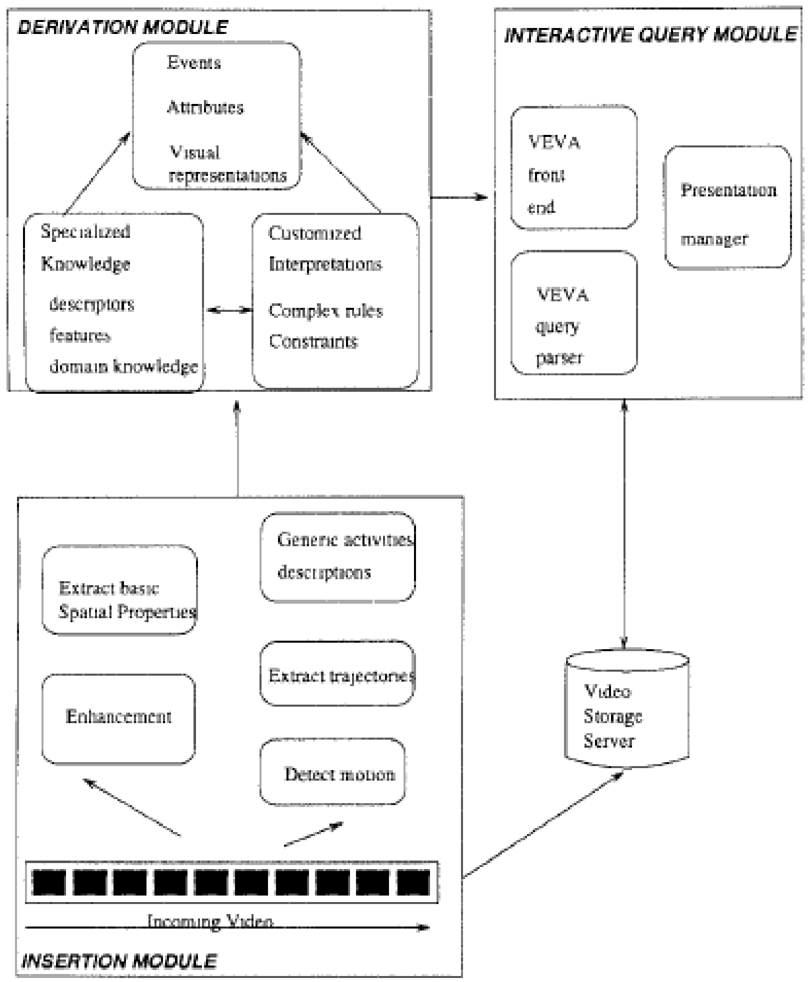


Fig. 9. An architecture for video classification and retrieval.

414. In my opinion, and in view of the well-developed state of the art by 2000, a POSITA would have known how to implement *Dimitrova*'s computer-based system with conventional computer hardware and software in view of *Dimitrova*'s disclosure. For example, *Dimitrova* relies upon the prior art of Khoros and the EVA query language for its system.

415. *Brill* separately discloses a "video device" as claimed. *Brill* explicitly discloses smart cameras wherein the "attributes to be detected are defined in [the]

device prior to the selection of a subset of the plurality of attributes." Ex. 1004, 5-6.

b. "means for detecting an object in a video from a single camera" [9.1]

416. *Dimitrova* and *Brill* separately disclose the corresponding structures of this means-plus-function element (*see* Section V(E)(1)) in view of the evidence provided herein in addition to Section VIII(C)(1)(a).

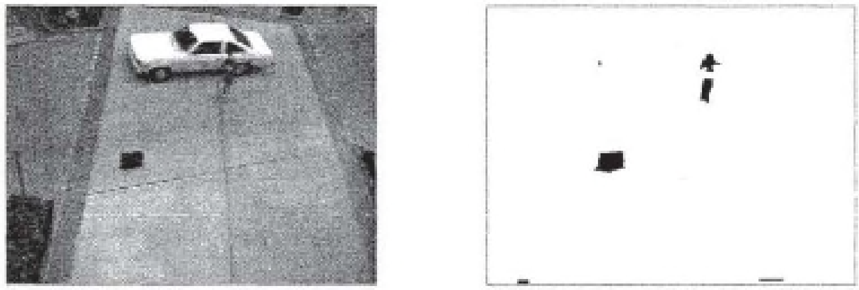
417. *Dimitrova* is a computer-based system with a computer performing the recited functions of these elements. Ex. 1006, 6. *Dimitrova* discusses several known systems and techniques in the field of "dynamic computer vision" for detecting and tracking an object in a video. *Id.*, 4, 6-8. *Dimitrova* also discloses the insertion module as the software that processes incoming video and that the insertion and derivation software modules detect objects in the video. *See, e.g., id.*, Fig. 9. Indeed, the '923 patent admits that object detection algorithms are well known and expressly states that "any" detection algorithm can be used. Ex. 1001, 9:30-41; 10:27-30.

418. Further, the use of computer hardware and software for video systems was well known in the art years before the '923 patent, and a POSITA would have known how to implement *Dimitrova*'s computer-based system and method in view of *Dimitrova*'s disclosure.

419. To the extent Patent Owner argues that *Dimitrova* does not disclose the corresponding structure of the limitation "means for detecting an object in a video from a single camera" in claim 9, *Brill* teaches such structure.

420. A POSITA would understand the corresponding structure is a computer system or equivalent video processing system utilizing conventional motion and/or change detection algorithms to detect objects. *See* Section V(E)(1). Further, the '923 patent describes that the motion and change detection algorithms operate in parallel and can be performed in any order or concurrently. *Id.*, 9:31-33. The '923 patent specifically discloses that any such algorithm would work. *Id.*, 9:34-35, 39-41. And the '923 patent itself cites to other references for those algorithms. *Id.*, 9:35-37, 44-48.

421. *Brill* teaches a change detection algorithm that corresponds to "block 52" of the '923 patent, which could be "[a]ny change detection algorithm." Ex. 1001, 9:39-48. This algorithm detects an object "if one or more pixels in a frame are deemed to be in the foreground." *Id.* *Brill*'s system employs this algorithm as demonstrated in Figure 4(b) where the person is deemed to be in the foreground. Ex. 1004, 7.



(a) (b)

Figure 4: (a) Current video image. (b) Foreground difference image

422. *Brill* also teaches a motion detection algorithm that corresponds to "block 51" of the '923 patent. Ex. 1001, 9:33-38 ("Any motion detection algorithm for detecting movement between frames at the pixel level can be used for this block."). *Brill*'s system employs this algorithm as demonstrated in Figure 16 where the algorithm detects movement of the two people between frames at the pixel level. Ex. 1004, 15-16. *Brill* explains, "[t]he brightness [of Figs. 16(b) and (d)] indicates the probability that the person's image intersects the given ***pixel***, which is highest in the middle of the region, and falls off towards the edge." Ex. 1004, 15.

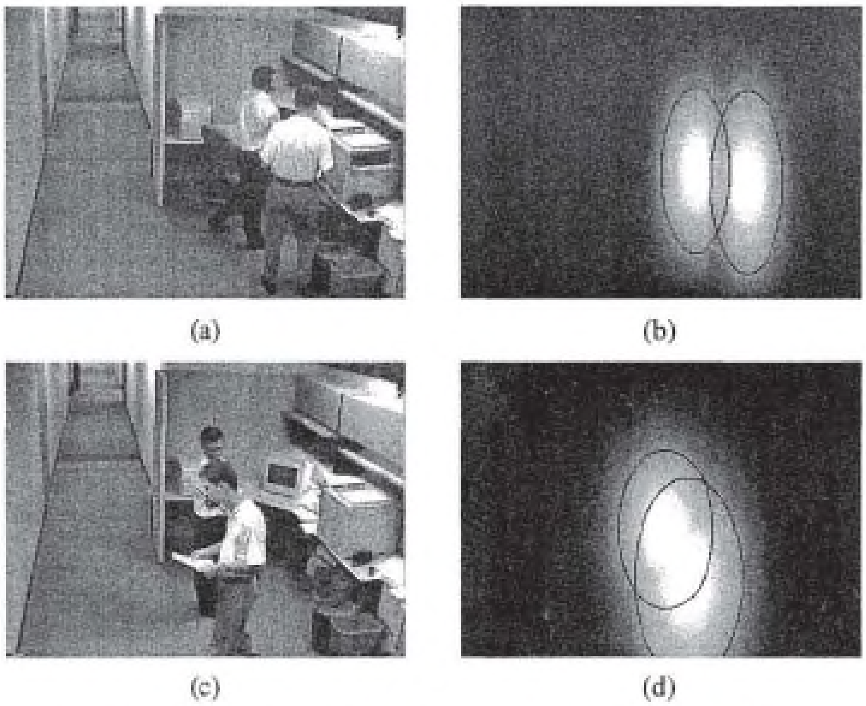


Figure 16: P-template images for partially occluding people

423. A POSITA would have found it obvious to implement *Dimitrova's* system with the teachings of *Brill* to achieve enhanced object detection, while yielding predictable results

c. "means for detecting a plurality of attributes of the object by analyzing the video from said single camera, the plurality of attributes including at least a physical attribute and a temporal attribute, each attribute representing a characteristic of the detected object" [9.2]

424. *Dimitrova* and *Brill* disclose the corresponding structures of this means-plus-function element (*see* Section V(E)(2)) in view of the evidence provided herein in addition to Section VIII(C)(1)(b). The '923 patent briefly discloses this process as using the same conventional algorithms as the object

detection and tracking process. Ex. 1001, 10:49-51. The combined system of *Dimitrova* and *Brill* fits within this generic disclosure.

425. *Dimitrova* is a computer-based system with a computer performing the recited functions of these elements. Ex. 1006, 6. *Dimitrova* also discloses that the insertion and derivation software modules provide the means for the detection of the attributes of objects. The insertion module contains the software "operators for the extraction of basic spatial properties, and operators for motion detection and the extraction of motion trajectories." *Id.*, 26. *Dimitrova* discloses that "the functionalities for spatial analysis are supplied by the Khoros computer vision environment [Rasure et al. 1990]. The extraction of image features, finding regions, and thinning operators are performed by calls to Khoros functions." *Id.*, 24. "The derivation module consists of operators for translation of the extracted features into meaningful descriptions [*i.e.*, attributes] for retrieval." *Id.*, 27.

426. *Dimitrova* discusses in detail motion recovery techniques and actual algorithms for detecting various attributes such as activity attributes and trajectory attributes. *Id.*, 13-18. This is well within, or even more than, the scope of that described in the '923 patent as the patent merely assumes that these techniques are known in the art. Ex. 1001, 10:11-22, 10:27-30, 10:39-41, 10:44-47.

d. "a memory storing the plurality of detected attributes" [9.3]

427. For the reasons in Sections VIII(C)(3), *Dimitrova* teaches this element.

e. "means for selecting a new user rule after the plurality of detected attributes are stored in memory" [9.4]

428. *Dimitrova* and *Brill* disclose the corresponding structures of this means-plus-function element (*see* Section V(E)(3)) in view of the evidence provided herein in addition to Section VIII(C)(1)(c).

429. A POSITA would understand that *Dimitrova* teaches user interfaces and I/O devices to enable a user to enter queries, consistent with that described in the '923 patent for element [9.4]. *Dimitrova* discloses a graphical user interface provided by a "visual front end" and query mechanisms of the interactive query module for creating a user rule that defines an event, which includes a display and user input devices. Ex. 1006, 27-28, Fig. 10.

f. "means for identifying an event of the object that is not one of the detected attributes of the object by applying a selected new user rule to the plurality of attributes stored in memory" [9.5]; "for identifying the event independent of when the attributes are stored in memory" [9.6]; "and for identifying the event without reprocessing the video" [9.7]; "wherein the applying the new rule to the plurality of detected attributes comprises applying the new user rule to only the plurality of detected attributes." [9.8]

430. The corresponding structure of these means-plus-function elements (*see* Section V(E)(4)) is disclosed by *Dimitrova's* query mechanisms as shown in Section VIII(C)(1)(d)-(f).

g. "wherein the event of the object refers to the object engaged in an activity" [9.9]

431. As explained in Section VIII(C)(1)(h), *Dimitrova* teaches this element.

7. Claims 10 and 31

432. Claims 10 and 31 depend from claims 9 and 30, respectively. *Dimitrova* discloses the additional feature of "a video camera operable to obtain the video." *Dimitrova* discloses, "the ***camera*** focus on a moving object." Ex. 1006, 32. A POSITA would understand that *Dimitrova* includes a video capture apparatus to capture the video that is processed by the system. *Brill* also discloses smart cameras used in the AVS system. Ex. 1004, Figure 1.

8. Claims 11 and 32

433. Claims 11 and 32 depend from claims 9 and 30, respectively.

a. "in real time"

434. Claims 11 and 32 add the following features that require identifying a first event in real time.

- means for identifying a first event of the object in real time by analyzing, of the plurality of attributes, only a first selected subset of the plurality of attributes (claim 11)
- means for identifying a first event in real time by analyzing, of the plurality of attributes, only a first selected subset of the plurality of attributes (claim 32)

435. The corresponding structure of the means-plus-function limitations in claims 11 and 32 is the structure identified in Sections VIII(C)(6)(f) and VIII(C)(15)(e), respectively, each of which performing the additional function of identifying a first event in real time. *See* Section V(E)(4).

436. This limitation requires identifying an event of the object in real time. The '923 patent discloses "real time" in several different embodiments. For example, "[t]he video surveillance system of the invention operates automatically, detects and ***archives*** video primitives of objects in the scene, and detects event occurrences in ***real time*** using event discriminators**.** In addition, ***action*** is taken in ***real time***." Ex. 1001, 9:14-18 (emphasis added); *see also id.,* 2:48-50 ("An object

of the invention is to produce a ***real time alarm*** based on an automatic detection of an event from video Surveillance data.") (emphasis added); *see also id.*, 11:34-35 ("the information can be viewed by the operator at any time, including ***real time***.") (emphasis added); *see also id.*, 9:25-26 ("video primitives are extracted in ***real time*** from the source video.") (emphasis added). Consequently, the term ***real time*** is used in the '923 patent to pertain to time-sensitive processes, i.e. those that depend upon timeliness and temporal continuity to properly perform their intended functions.

437. Importantly, claims 11 and 32 depend from claims 9 and 30, respectfully, both of which specifically require that the "means for identifying" applies the selected new user rule to the attributes ***stored in memory***." Thus, while claims 11 and 32 recite identifying an object in "real time" it must still query or filter for the stored attributes. This is consistent with the '923 patent's disclosure that the detected attributes are first ***archived*** and event identification occurs in real time based on the detected attributes. *See id.*, 9:14-18. Nowhere in the specification indicates that event identification occurs in real time based on attributes that are not also stored or archived. Thus, in view of the '923 patent claims and disclosure storing the attributes is not inconsistent with real-time operation.

438. *Dimitrova*'s system is applied in surveillance and monitoring. *See*, *e.g.*, Ex. 1006, 6, 11, 27. A POSITA would have been motivated to add a real-time identification feature to its system because it would have been essential for such purposes. One way of achieving real-time identification would have been to add a fast, repeated query feature mimicking real-time identification.

439. Based on *Brill's* teaching, a POSITA would have been further motivated to add the real-time feature. *Brill* discloses that its system "processes live video streams from ***surveillance cameras*** to automatically produce a ***real-time*** map-based display of the locations of people, objects and events in a monitored region." Ex. 1004, 4 (emphasis added); *see also id.*, 14 ("Once a simple or complex event has been defined, the AVS system immediately begins recognition of the new events in ***real time***, and taking the actions specified by the user.") (emphasis added). Specifically, *Brill's* system "recognizes and reports" a person entering a monitored area and a briefcase being picked up "in ***real time*** as illustrated in Figure 10." *Id.*, 12 (emphasis added).

Figure 10: A series of simple events

440. A POSITA would have found it obvious to implement *Dimitrova*'s system with the real-time feature based on the teachings of *Brill* to provide enhancements or achieve particular design objectives in *Dimitrova*'s system, which is to better achieve its goal as a video surveillance or monitoring system.

b. The events being identified

441. Claim 11 requires the limitation "means for identifying a first event of the object"; whereas claim 32 requires "means for identifying an event of the first object interacting with the second object." These limitations are disclosed as explained in Sections VIII(C)(6)(f) and VIII(C)(15)(e), respectively.

c. "analyzing, of the plurality of attributes, only a first selected subset of the plurality of attributes"

442. *Dimitrova*'s Parade Query does not require analyzing all detected attributes as recited as explained in Section VIII(C)(2).

9. Claims 12 and 33

443. Claims 12 and 33 depend on claims 11 and 32, respectively. Claim 12 adds the limitation "the means for identifying an event of the object comprises means for identifying a second event of the object by analyzing, of the plurality of attributes, only a second selected subset of the plurality of attributes that have been archived." Claim 33 adds the limitation "the means for identifying an event of the first object interacting with the second object comprises means for identifying a

second event by analyzing, of the plurality of attributes, only a second selected subset of the plurality of attributes which have been archived."

444. The corresponding structure of this means-plus-function limitation in claims 12 and 33 is the structure identified in Sections VIII(C)(6)(f) and VIII(C)(15)(e), respectively, each of which performing the additional function. *See* Section V(E)(4).

445. While claims 11 and 32 recite a means-plus-function limitation requiring the capability of identifying an event relating to an object in ***real-time***, these claims recite a means-plus-function limitation requiring the capability of further identifying an event relating to that same object based on ***stored or archived*** attributes. The '923 patent specification does not specifically disclose this aspect of the claims. In any event, both *Dimitrova* and *Brill* teach this additional requirement.

446. *Dimitrova* discloses a "video storage server" that is "a disk array serving as a repository of the video sequences." Ex. 1006, 27, Fig. 9. After storage, queries can be run on the attributes stored in the video storage server. For example, *Dimitrova* explains that the example query of Figure 10 "will select those video sequences from the repository" that satisfy the query's conditions and display the results. *Id.*, 29.

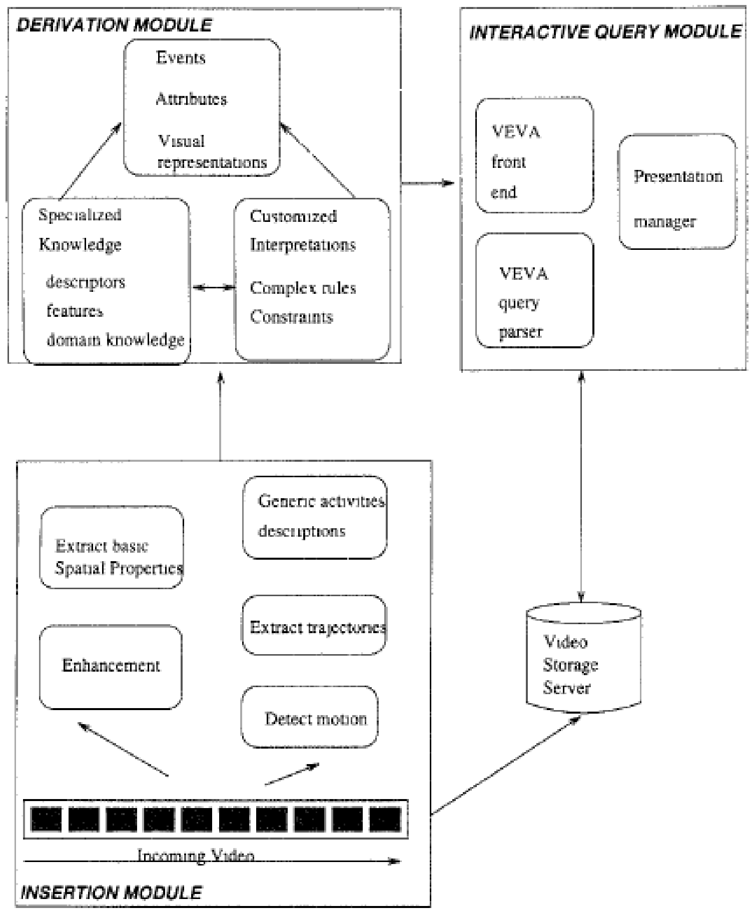


Fig. 9. An architecture for video classification and retrieval.

447. The events *Dimitrova* identifies based on the stored attributes can relate to the same objects that are the subject of real-time analyses. As *Dimitrova* explains, the attributes relating to its objects can include a unique "object ID," or other object-specific attributes such as "*name_of*," so that queries can limit their results to those relating to a specific object, including an object involved in a real-time event. Ex. 1006, 17.

448. Similarly, *Brill* explains that "the system monitors and records the movements of humans in its field of view," and that "[f]or every person that it sees,

it creates a log file that summarizes important information about the person, including a snapshot taken when the person was close to the camera and (if possible) facing it." Ex. 1004, 18. "When the person leaves the scene, the log entry is saved to a file. Each log entry records the time when the person entered the scene and a list of coordinate pairs showing their position in each video frame." *Id.*; *see also id.* at Fig. 1:

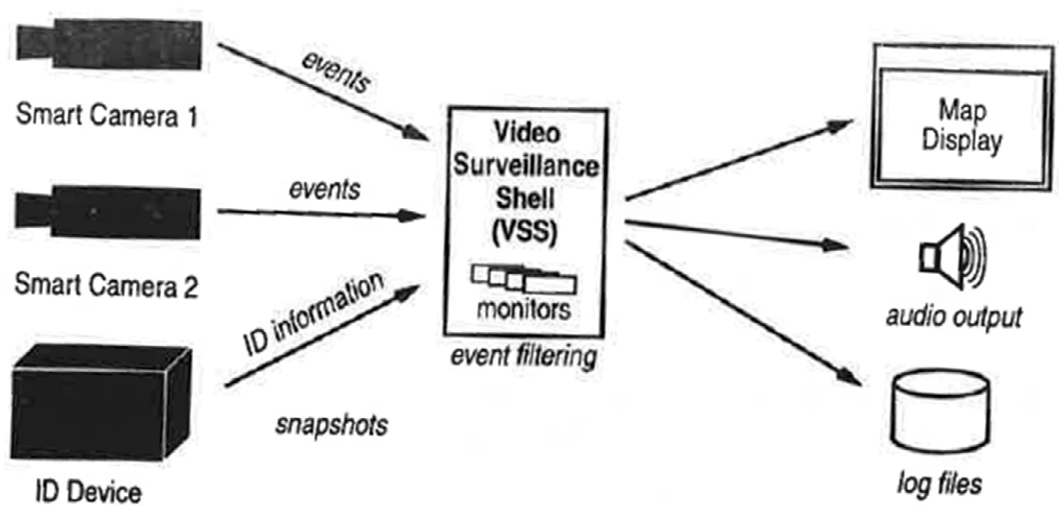


Figure 1: AVS system diagram

449. *Brill's* queries can then be run on the entries stored in the log files, with respect to a particular ID of the object that corresponds to a person of interest that was the subject of a real-time event. *Id.*, 11.

450. To the extent Patent Owner argues that the recited "second selected subset of the plurality of attributes" must be different from the "first selected

subset" of claims 11 and 32[14], a "second event" will necessarily be associated with different attributes compared to the real-time "first event" because at the very least the second event will be based on the particular ID of the object. Regardless, *Dimitrova* teaches that the query used to determine each particular event can be a subset of attributes that has no relation to any other event. *See* Section VIII(C)(2).

451. The requirement of claim 33 that the second event relate to the first object interacting with the second object is, as discussed above with respect to claim 32, taught by both *Dimitrova* and *Brill*. *See* Section VIII(C)(8).

10. Claims 14 and 35

452. Claims 14 and 35 depend from claims 9 and 30, respectively. These claims require the additional features of the memory being "configured to store at least some of the plurality of attributes for at least two months." These claims also recite a means-plus-function limitation requiring analyzing some of the attributes that are stored for at least two months. The corresponding structure of this means-plus-function limitation in claims 14 and 35 is the structure identified in Sections VIII(C)(6)(f) and VIII(C)(15)(e), respectively, which performs the additional

[14] Nowhere in the '923 patent specification discloses a "second event" or a "second selected subset of the plurality of attributes," much less the "second selected subset of the plurality of attributes" being different from the "first selected subset."

function. *See* Section V(E)(4). Accordingly, the means-plus-function limitation is disclosed by *Dimitrova*.

453. Claims 14 and 35 are apparatus claims directed to structures, e.g., memories. This reads on conventional computer non-volatile memory, which is designed to retain the stored information indefinitely, as users cannot enjoy the full benefit of a storage system if the data they have decided to retain is deleted without warning. These claims only require that the memory "is configured to store…for at least two months." This configuration is a functional requirement, which would be met by any conventional non-volatile memory at the time of the '923 patent. Accordingly, *Dimitrova* disclosed memory would meet this limitation as it is configured to store data indefinitely and therefore it is capable of performing the two-month storage function.

454. Nonetheless, the combination of *Dimitrova* and *Brill* renders the claims obvious. As explained above with respect to claims 12 and 33, both *Dimitrova* and *Brill* teach identifying events by analyzing stored attributes. *See* Section VIII(B)(9). Neither reference places any limit on how long these attributes can remain in storage prior to their analysis. This is because the amount of time information is stored is a trivial, non-technical matter of design choice, particularly with respect to an arbitrary, relatively short amount of time, such as two months.

455. Of course, conventional computer non-volatile memory is designed to retain the stored information indefinitely. Thus, a memory used in *Dimitrova* and *Brill* would satisfy the claim limitation. The limitation does not depend on any technicality but rather on the user's intended use of the system. If the user decides to store the attributes for more than two months, the user can simply store the memory some place safe such that the stored attributes are not accidently deleted or overwritten.

456. Indeed, this limitation requires no technical change to the structures disclosed in *Dimitrova* and *Brill*. *Dimitrova* allows a user to view the visual query results in a later time point. Ex. 1006, 28-29. *Dimitrova* places no time limit on how long later point in time could be and it could be two months, three months or a year. Instead, all that is required to meet this limitation is that the user of such a system not delete the stored attributes that are collected for more than two months.

457. The ordinary knowledge of a POSITA would encourage storing attributes for at least two months in a video surveillance system like that disclosed in *Dimitrova* and *Brill* designed for security purposes. Ex. 1006, 11, 27; Ex. 1004, 4, 5. For example, a POSITA would know that many crimes can go easily unnoticed for more than two months. A POSITA would know also that repeat crimes can happen within a longer interval than two months. The POSITA's knowledge would have been supplemented by *Dimitrova*'s teaching of using such

surveillance systems in a museum setting. For example, *Dimitrova* discloses that its system can be used in a museum setting to identify someone who damaged or stole an expensive Miro painting. Ex. 1006, 11. A POSITA would be motivated to have the data stored for more than two months in the event that the information is needed for evidence in a prosecution, which could take more than two months.

458. Similarly, *Dimitrova* identifies the utility of its system for sports data Ex. 1006, 22-25, 28-29, Fig. 11. It is well known that archived data about sports personnel is relevant for well over two months and should therefore be stored for a longer period of time. For example, in 1999 Kevin McHale was inducted into the Basketball Hall of Fame. Ex. 1039, 1. Mr. McHale's induction to the Hall of Fame was due to his time playing as a professional in the 1980s. *Id.* Thus, a POSITA, and any sports fan, would be readily motivated to store sports footage attribute data for more than two months, indeed for years, in order to readily identify the highlights of famous players, such as Kevin McHale, who was being celebrated for his achievements on the court over a decade later. *Id.*

459. This would motivate a POSITA to implement the *Dimitrova* system such that the data is stored for well over two months. And to take full advantage of this feature, it would be obvious to search some of the attributes stored for at least two months.

11. Claims 18 and 40

460. Claims 18 and 40 depend from claims 9 and 30, respectively, and adds the feature "the video surveillance device is a computer system configured as a video surveillance device." *Dimitrova* and *Brill* separately disclose this element.

461. *Dimitrova* describes, "[t]he research presented in this article builds on the existing results in two areas: ***dynamic computer vision*** and digital video modeling." Ex. 1006, 6 (emphasis added). *Dimitrova* is fully aware that its computer based system can be used in video surveillance where it describes that "[a]pplications such as ***automated surveillance*** may require retrieval of either video sequences or objects contained in these sequences based on the object trajectories." *Id.*, 11 (emphasis added); *see also id.*, 27 ("The automation of this whole process is possible for strictly limited application domains such as industrial ***monitoring***, domain specific video editing, camera ***surveillance***, and others.") (emphasis added). A POSITA would have understood that *Dimitrova's* system is configured as a video surveillance device.

462. *Brill*'s system is called the "Autonomous Video Surveillance (AVS)" system that "processes live video streams from ***surveillance*** cameras to automatically produce a real-time map-based display of the locations of people, objects and events in a monitored region." Ex. 1004, 4 (emphasis added); *see also id.*, 5 ("AVS system incorporates multiple cameras to enable surveillance of a

wider area than can be monitored via a single camera."). The Video Surveillance Shell (VSS) in Figure 1 is a computer system that integrates the information from the surveillance cameras and displays it on a map. *Id.*, 4. A POSITA would have understood that *Brill's* system is also configured as a video surveillance device.

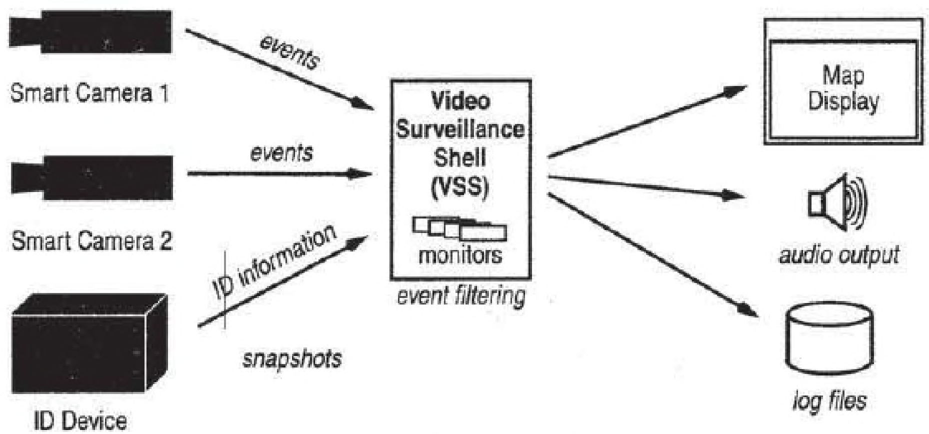


Figure 1: AVS system diagram

12. Claims 19 and 41

463. Claims 19 and 41 depend from claims 9 and 30, respectively. *Dimitrova* discloses the video device further comprising "video sensors," which include video cameras. Ex. 1001, 6:8-12. *Dimitrova*'s system uses a video camera. Ex. 1006, 32 ("For example, when we have the ***camera*** focus on a moving object, then the object appears to be stationary.") (emphasis added). *Brill* also discloses more than one smart cameras used in the AVS system. Ex. 1004, Figure 1. Accordingly, the combination of *Dimitrova* and *Brill* discloses this these claims.

13. Independent Claim 20

464. Claim 20 is substantially identical to claim 1. Claim 20 recites "at least a physical attribute and a temporal attribute," instead of reciting "at least one of a physical attribute and a temporal attribute" as in claim 1.

465. Claim 20 also recites "a combination of the attributes" instead of "a plurality of detected attributes." This does not meaningfully distinguish claim 20 from claim 1 because the recited combination does not refer to any specific combination, and the word "combination" simply implies that there is more than one as in "a plurality."

466. For the reasons in Sections VIII(C)(1), the combination of *Dimitrova* and *Brill* discloses all corresponding limitations of claim 20.

14. Independent Claims 22 and 29

467. Claims 22 and 29 are essentially claims 1 and 8, respectively, in system form where they are directed to a "non-transitory computer-readable storage medium." The added feature does not add any technicality to claims 1 and 8, and it would have been obvious to a POSITA at the priority date to have a computer-readable-medium having software instructions stored therein.

468. *Dimitrova* teaches that its "video storage server is envisioned to be a disk array." Ex. 1006, 27. A POSITA would understand that *Dimitrova's* system

is embodied in a computer readable medium, such as memory, hard drive or removable storage media.

469. *Brill* separately teaches a "non-transitory computer-readable storage medium." *Brill's* AVS system is a "software" (Ex. 1004, 6), and a POSITA would understand that it is embodied in a computer readable medium, such as memory, hard drive or removable storage media.

470. Accordingly, the combined teachings of *Dimitrova* and *Brill* disclose the limitations of these claims. *See* Sections VIII(C)(1) and VIII(C)(5).

15. Independent Claim 30

471. Much like claim 9 as explained in Section VIII(C)(6), claim 30 recites means-plus-function elements ([30.1]-[30.2] and [30.4]-[30.7]) with functions substantially corresponding to the steps of claim 8.

a. A video device comprising

472. For the reasons in Section VIII(C)(6)(a), *Dimitrova* and *Brill* separately disclose the preamble.

b. "means for detecting first and second objects in a video from a single camera" [30.1]

473. For the reasons in Section VIII(C)(5)(a) and VIII(C)(6)(b), the combination of *Dimitrova* and *Brill* discloses element [30.1].

c. "means for detecting a plurality of attributes of the object by analyzing the video from said single camera, each attribute representing a characteristic of the respective detected object" [30.2]

474. For the reasons in Sections VIII(C)(5)(b) and VIII(C)(6)(c), the combination of *Dimitrova* and *Brill* discloses element [30.2].

d. "a memory storing the plurality of detected attributes" [30.3]

475. For the reasons in Sections VIII(C)(6)(d) and VIII(C)(3), *Dimitrova* discloses this element.

e. "means for identifying an event of the first object interacting with the second object by applying a selected new user rule to the plurality of attributes stored in memory" [30.4]; "for identifying the event independent of when the attributes are stored in memory, the event not being one of the detected attributes" [30.5]; "wherein the applying the selected new user rule to the plurality of attributes stored in memory comprises applying the selected new user rule to only the plurality of attributes stored in memory." [30.6]

476. Compared to claim 9, claim 30 does not recite "and for identifying the event without reprocessing the video," so claim 9 is inclusive. Accordingly, for the reasons in Sections VIII(C)(5)(c)-(g) and VIII(C)(6)(e)-(f), *Dimitrova* discloses elements [30.4]-[30.6].

f. "wherein the event of the object refers to the object engaged in an activity" [30.8]

477. For the reasons in Sections VIII(C)(5)(h) and VIII(C)(6)(g), *Dimitrova* discloses this element.

16. Claim 36

478. Claims 36 depends from claim 30 and requires the additional feature of "wherein the means for identifying an event includes means for identifying the event without reprocessing the video." The corresponding structure of this means-plus-function limitation is the structure identified in Section VIII(A)(15)(e), which performs the additional function. *See* Section V(E)(4). Accordingly, *Dimitrova* discloses claim 36.

IX. CONCLUSION

479. For the reasons set forth above, it is my opinion that claims 1-41 of the '923 patent are (1) disclosed by *Kellogg* (2) rendered obvious by *Kellogg* in view of *Brill* and/or (3) rendered obvious by *Dimitrova* in view of *Brill.*

480. I understand that this Declaration will be filed as evidence in a petition for an *inter partes* review before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I further understand that I may be subject to cross-examination in relation to this case. If such cross-examination is

required, I will appear for cross-examination within the United States during the time allotted for cross-examination.

I declare that all statements made herein are made on information and belief and I believe them to be true. I understand that any willfully incorrect statements may be punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Respectfully submitted,

Dated: November 12, 2018 By: John R. Grindon

John R. Grindon, D.Sc.