

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SAMSUNG ELECTRONICS CO., LTD.; and
SAMSUNG ELECTRONICS AMERICA, INC.
Petitioners

v.

IMAGE PROCESSING TECHNOLOGIES, LLC
Patent Owner

Patent No. 8,805,001

**DECLARATION OF DR. JOHN C. HART
IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 8,805,001**

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND AND EXPERIENCE	1
A.	Qualifications	1
B.	Previous Testimony	4
III.	TECHNOLOGICAL BACKGROUND.....	5
IV.	THE '001 PATENT	11
V.	SUMMARY OF OPINIONS	20
VI.	LEVEL OF ORDINARY SKILL IN THE ART	21
VII.	CLAIM CONSTRUCTION.....	22
VIII.	THE PRIOR ART TEACHES OR SUGGESTS EVERY STEP AND FEATURE OF THE CHALLENGED CLAIMS OF THE '001 PATENT	22
A.	Overview Of The Prior Art References	22
1.	Alton L. Gilbert et al., A Real-Time Video Tracking System, PAMI-2 No. 1 IEEE Transactions on Pattern Analysis and Machine Intelligence 47 (Jan. 1980) (“Gilbert”) (Ex. 1005)	22
2.	U.S. Patent No. 5,521,843 (“Hashima”) (Ex. 1006)	30
3.	U.S. Patent No. 5,150,432 (“Ueno”) (Ex. 1007).....	37
B.	Ground 1: Gilbert In View Of Hashima Teaches Or Suggests Every Step And Feature Of Claims 1-4.....	41
1.	Reasons To Combine Gilbert And Hashima	41
2.	Claim 1	47
3.	Claim 2.....	60
4.	Claim 3.....	60
5.	Claim 4.....	62
6.	Detailed Application Of Gilbert And Hashima To Claims 1-4	64

C.	Ground 2: Hashima And Ueno Teaches Or Suggests Every Step And Feature Of Claims 1-4	80
1.	Reasons To Combine Hashima And Ueno.....	80
2.	Claim 1	83
3.	Claim 2.....	93
4.	Claim 3.....	94
5.	Claim 4.....	95
6.	Detailed Application Of Hashima And Ueno To Claims 1-4.....	97
D.	Ground 3: Ueno In View Of Gilbert Teaches Or Suggests Every Step And Feature Of Claims 1-4	114
1.	Reasons To Combine Ueno And Gilbert.....	114
2.	Claim 1	119
3.	Claim 2.....	131
4.	Claim 3.....	131
5.	Claim 4.....	133
6.	Detailed Application Of Ueno And Gilbert To Claims 1-4.....	136
IX.	CONCLUSION.....	152

1. I, John C. Hart, declare as follows:

I. INTRODUCTION

2. I have been retained by Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. (collectively, “Petitioner”) as an independent expert consultant in this proceeding before the United States Patent and Trademark Office (“PTO”).

3. I have been asked to consider whether certain references teach or suggest the features recited in Claims 1-4 of U.S. Patent No. 8,805,001 (“the ’001 Patent”) (Ex. 1001), which I understand is allegedly owned by Image Processing Technologies, LLC (“Patent Owner”). My opinions and the bases for my opinions are set forth below.

4. I am being compensated at my ordinary and customary consulting rate for my work.

5. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other interest in this proceeding.

II. BACKGROUND AND EXPERIENCE

A. Qualifications

6. I have more than 25 years of experience in computer graphics and image processing technologies. In particular, I have devoted much of my career to

researching and designing graphics hardware and systems for a wide range of applications.

7. My research has resulted in the publication of more than 80 peer-reviewed scientific articles, and more than 50 invited papers, and talks in the area of computer graphics and image processing.

8. I have authored or co-authored several publications that are directly related to target identification and tracking in image processing systems. Some recent publications include:

- P.R. Khorrami, V.V. Le, J.C. Hart, T.S. Huang. A System for Monitoring the Engagement of Remote Online Students using Eye Gaze Estimation. Proc. IEEE ICME Workshop on Emerging Multimedia Systems and Applications, July 2014.
- V. Lu, I. Endres, M. Stroila and J.C. Hart. Accelerating Arrays of Linear Classifiers Using Approximate Range Queries. Proc. Winter Conference on Applications of Computer Vision, Mar. 2014.
- M. Kamali, E. Ofek, F. Iandola, I. Omer, J.C. Hart Linear Clutter Removal from Urban Panoramas. Proc. International Symposium on Visual Computing. Sep. 2011.

9. From 2008-2012, as a Co-PI of the \$18M Intel/Microsoft Universal Parallelism Computing Research Center at the University of Illinois, I led the

AvaScholar project for visual processing of images that included face identification, tracking and image histograms.

10. I am a co-inventor of U.S. Patent No. 7,365,744.

11. I have served as the Director for Graduate Studies for the Department of Computer Science, an Associate Dean for the Graduate College, and I am currently serving as the Executive Associate Dean of the Graduate College at the University of Illinois. I am also a professor in the Department of Computer Science at the University of Illinois, where I have served on the faculty since August 2000. As a professor I have taught classes on image processing and graphics technology and have conducted research into specific applications of these technologies.

12. From 1992 to 2000, I worked first as an Assistant Professor and then as an Associate Professor in the School of Electrical Engineering and Computer Science at Washington State University.

13. From 1991-1992, I was a Postdoctoral Research Associate at the Electronic Visualization Laboratory at the University of Illinois at Chicago, and at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign.

14. I earned a Doctor of Philosophy in Electrical Engineering and Computer Science from the University of Illinois at Chicago in 1991.

15. I earned a Master's Degree in Electrical Engineering and Computer Science from the University of Illinois at Chicago in 1989.

16. I earned a Bachelor of Science in Computer Science from Aurora University in 1987.

17. I have been an expert in the field of graphics and image processing since prior to 1996. I am qualified to provide an opinion as to what a person of ordinary skill in the art ("POSA") would have understood, known, or concluded as of 1996.

18. Additional qualifications are detailed in my curriculum vitae, which I understand has been submitted as Exhibit 1003 in this proceeding.

B. Previous Testimony

19. In the previous five years, I have testified as an expert at trial or by deposition or have submitted declarations in the following cases:

20. *Certain Computing or Graphics Systems, Components Thereof, and Vehicles Containing Same*, Inv. No. 337-TA-984.

21. *ZiiLabs Inc., Ltd v. Samsung Electronics Co. Ltd. et al.*, No. 2:14-cv-00203 (E.D. Tex. Feb. 4, 2016).

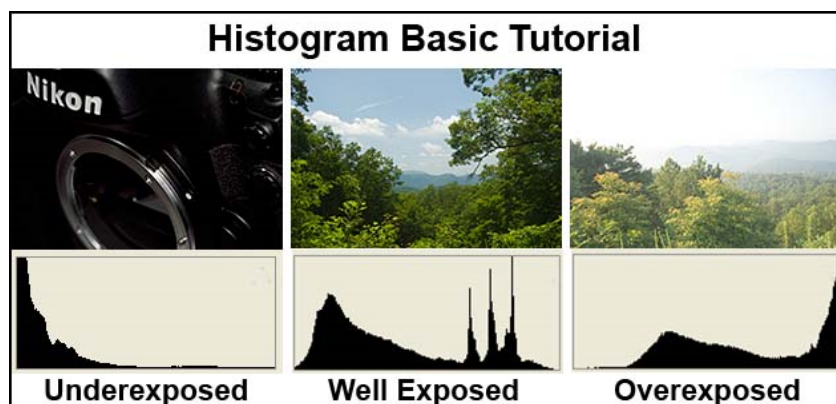
22. *Certain Consumer Electronics with Display and Processing Capabilities*, Inv. No. 337-TA-884.

III. TECHNOLOGICAL BACKGROUND

23. Image processing systems have long used histograms as a mathematical tool to identify and track image features and to adjust image properties. The use of histograms to identify and track image features dates back to well before 1997. *See, e.g.*, Alton L. Gilbert et al., “A Real-Time Video Tracking System,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 2, January 1980 (Ex. 1005); D. Trier, A. K. Jain and T. Taxt, “Feature Extraction Methods for Character Recognition-A Survey”, Pattern Recognition, vol. 29, no. 4, 1996, pp. 641–662 (Ex. 1008 at 10) (citing M. H. Glauberman, “Character recognition for business machines,” Electronics, vol. 29, pp. 132-136, Feb. 1956 (Ex. 1009)).

24. A digital image is represented by a number of picture elements, or pixels, where each pixel has certain properties, such as brightness, color, position, velocity, etc., which may be referred to as domains. For each pixel property or domain, a histogram may be formed. A histogram is a type of bar chart used for counting the number of items meeting certain criteria. In image processing, histograms count the number of pixels in the image having certain characteristics, where each bar counts the pixels that share a specific value or value range for that characteristic. For example, if the luminance (brightness) of each pixel were represented by an 8-bit value, a luminance histogram would be a bar chart

indicating, for each luminance value from 0 to 255, the count of the number of pixels in the image with that luminance value. As shown below, a luminance histogram may reveal certain properties of an image, such as whether it is properly exposed, based on whether an excessive number of pixels fall on the dark end or light end of the luminance range.

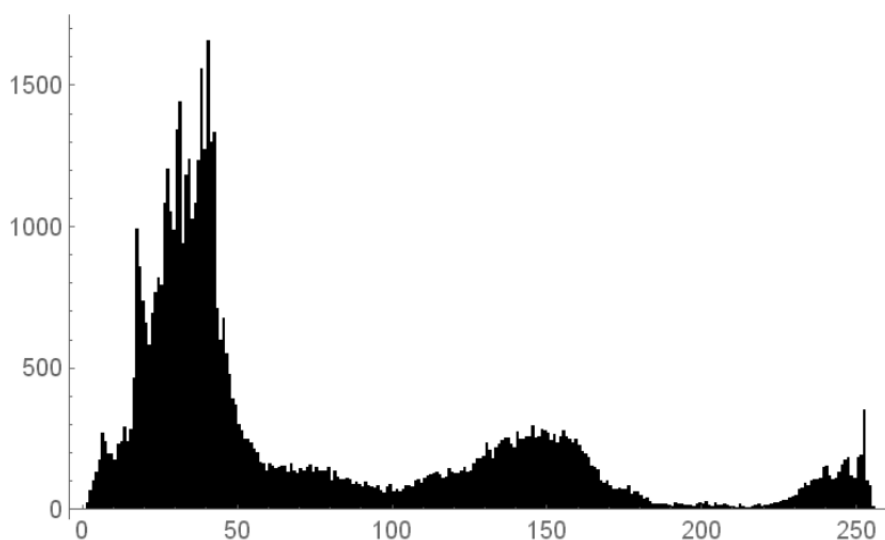
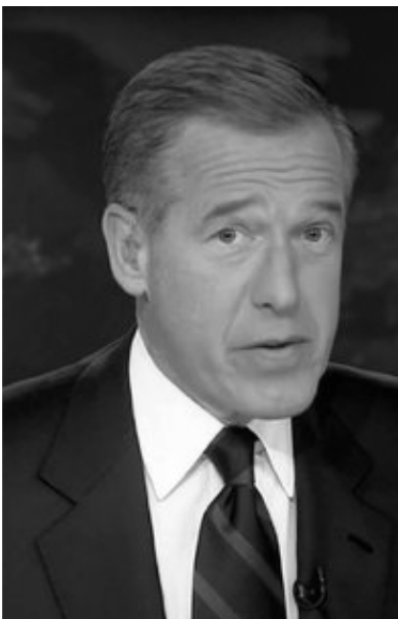


25. Histograms of other pixel properties can also be formed. For example, the figure below illustrates two histograms formed by counting the number of black pixels having each X-coordinate value (i.e., the X-coordinate domain) and the number having each Y-coordinate value (i.e., the Y-coordinate domain).



26. Such histograms are sometimes called “projection histograms” because they represent the image projected onto each axis. In the example above, the image was pure black and white, but projection histograms of a greyscale image can also be formed in a similar manner by defining a luminance threshold and projecting, for example, only those pixels that have a luminance value lower than 100.

27. A more complex greyscale image is shown below, along with its luminance histogram (black = 0; white = 255):

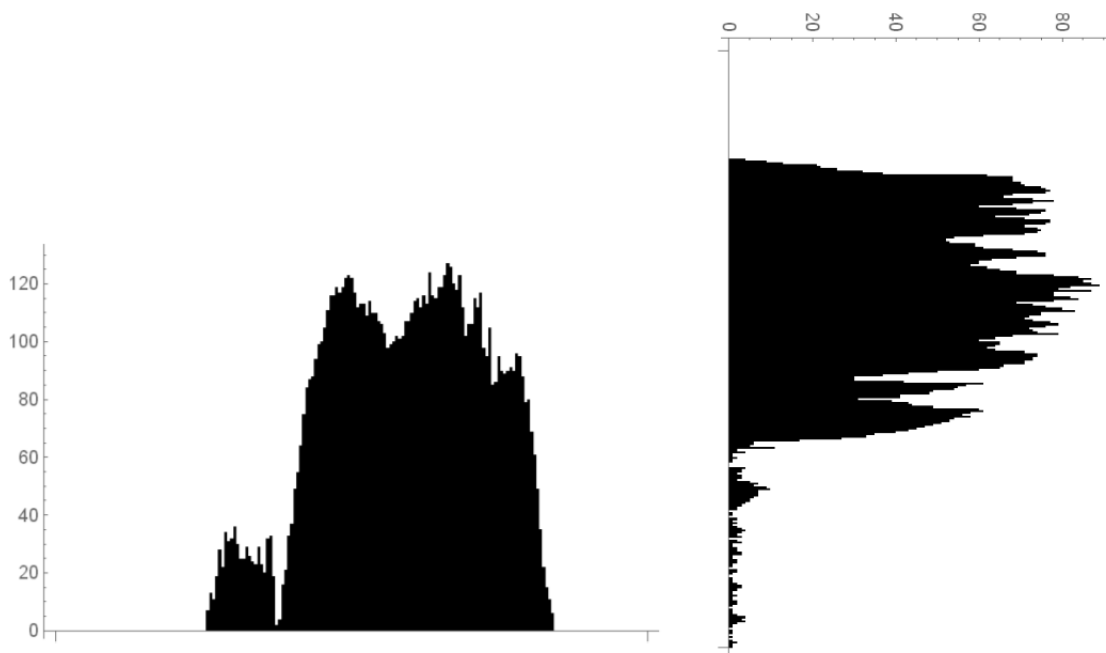


28. Here, the peak in the dark luminance region (luminance = 0-50) corresponds to the dark suit and tie and relatively dark background. The peak in the light luminance region (luminance > 230) corresponds to the white shirt, while the central peak (between luminance 130 and 170) corresponds to the medium brightness of the face. If one were to select only the subset of pixels with

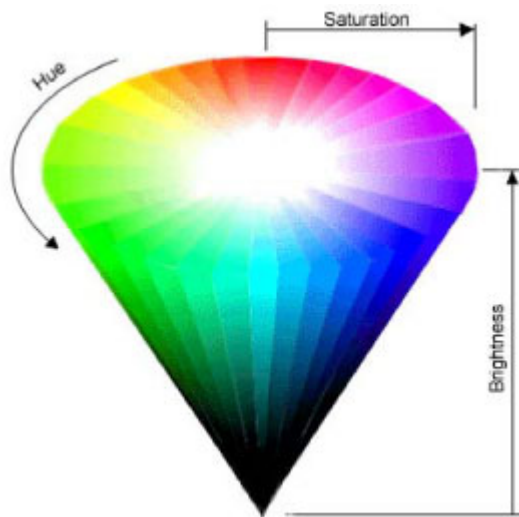
brightness between 130 and 170 and plot them according to their x and y positions, one would get the following image:



29. Taking projection histograms of this subset of pixels with luminance between 130 and 170, then, provides an indication of location of the face in the image. On the left, below, is a projection of this subset of pixels onto the x axis, and on the right is a similar projection onto the y axis.



30. Histograms may also be formed of pixel color properties in much the same way. Color is typically represented by three values: hue, saturation and luminance. Hue (aka “tone”) is an angle ranging from 0° to 360° around a color wheel that indicates which “color” is being represented, e.g. 0° = red, 60° = yellow, 120° = green, 180° = cyan, 240° = blue, and 300° = magenta. Saturation, which may also range from 0 to 255, represents how “brilliant” the color is. For example, if a color with a saturation of 255 represents red, then a saturation of 128 would represent pink and a saturation of 0 would represent gray. Luminance ranges from 0 to 255 represent the “brightness” of the color. If luminance = 0, then the color is black, regardless of the other values. Given a color image, the luminance values of the pixels would yield the “black-and-white” or grayscale version of the image.



IV. THE '001 PATENT

31. The '001 Patent, entitled “Image Processing Method,” was filed on November 17, 2009, and issued on August 12, 2014. The '001 Patent names Patrick Pirim as the sole inventor. I understand that the '001 Patent claims a priority date of July 22, 1996.

32. The '001 Patent is generally directed to tracking a target using an image processing system. For example, in the Abstract, the '001 Patent notes that the invention relates to “[a] method and apparatus for localizing an area in relative movement and for determining the speed and direction thereof.” Ex. 1001, '001 Patent at Abstract.

33. The '001 Patent uses the pixels in a frame of an image of a video signal to form one or more histograms in order to identify and track a target. *See*

e.g., Ex. 1001, '001 Patent at Claims 1-4. The input signal employed in the '001 Patent is comprised of a “succession of frames, each frame having a succession of pixels.” Ex. 1001 at 3:28-31. Although the disclosed embodiments relate primarily to a video signal, the '001 Patent also teaches that the input signal could correspond to other types of signals, for example “ultrasound, IR, Radar, tactile array, etc.” Ex. 1001, '001 Patent at 9:26-29.

34. The '001 Patent teaches constructing a histogram showing the frequency of the pixels meeting a certain characteristic. In the '001 Patent, these characteristics—such as luminance or speed—are referred to as “domains.” Histograms may be constructed in a variety of domains, for example, the '001 Patent teaches that examples of possible domains include pixel data such as “i) luminance, ii) speed (V), iii) oriented direction (D1), (iv) time constant (CO), v) hue, vi) saturation, and vii) first axis (x(m)), and viii) second axis (y(m)).” *Id.* at 4:3-6.

35. A domain can be further subdivided into classes, each class consisting of the subset of pixels with similar domain values. Figure 14a (and the accompanying text) illustrates an example of “classes” within a domain:

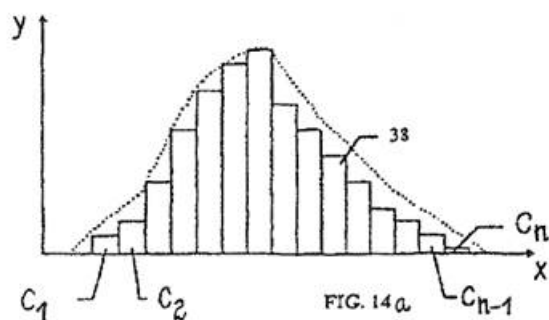
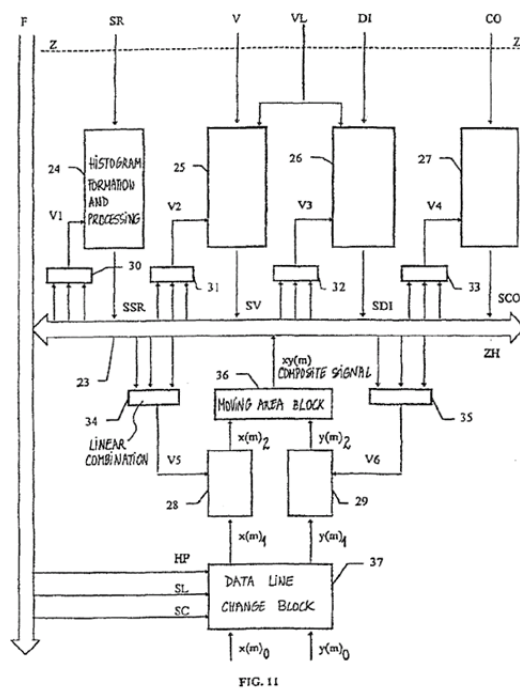


FIG. 14 *a* shows an example of the successive classes C_1 , C_2 , ..., C_{n-1} , C_n , each representing a particular velocity, for a hypothetical velocity histogram, with their being categorization for up to 16 velocities (15 are shown) in this example. Also shown is envelope **38**, which is a smoothed representation of the histogram.

Ex. 1001, '001 Patent at 20:54-59.

36. The hypothetical histogram in Figure 14a would be constructed using histogram formation block 25 in Figure 11. In this figure, various histogram processors (numbered 24-29) are shown that allow creation of histograms in various domains. Block 25 is disclosed as creating velocity histograms. *Id.* at 17:9-15.



A detailed depiction of histogram block 25 is shown in Figure 13:

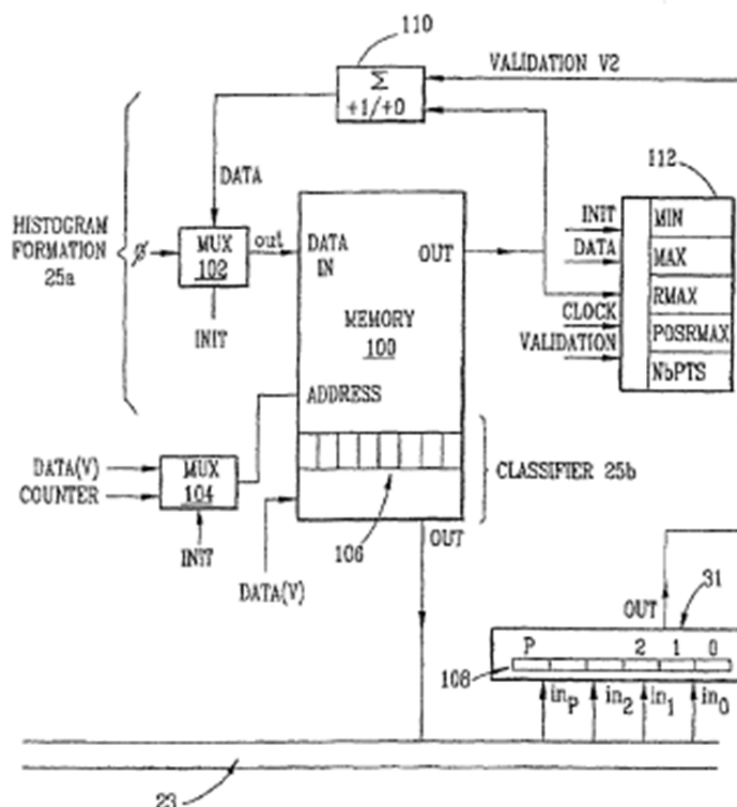
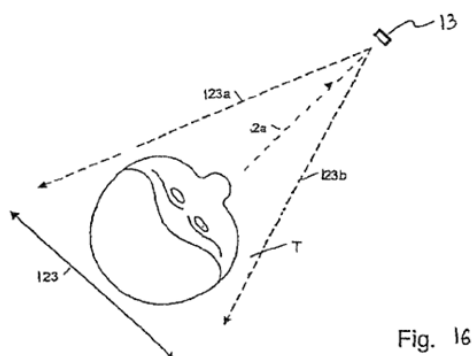


FIG.13

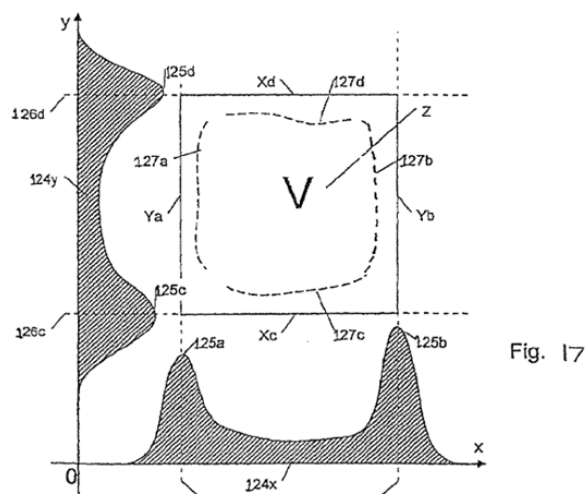
37. In Figure 13, velocity data for a pixel is input into a memory address and into classifier 25b. The classifier contains registers 106 that correspond to classes within the particular domain. Thus, for a classifier in a velocity histogram formation block, the classifier would have a register for each velocity class. Because the histogram will only be incremented for pixels satisfying the classification criteria that, when met, outputs a classification signal of “1,” the histogram in Figure 14a would be constructed using a classifier 25b that has each of the velocity-class registers set to “1.” In this example, each pixel that is input into classifier 25b would generate a classification signal of “1.” The histogram

would then be updated to include the input pixel, which it would do by incrementing the histogram bin corresponding to the appropriate velocity class. In other examples, a classifier may output a classification signal of “1” for only specific classes of a domain, rather than for all of the classes in a domain as in Figure 14a. Thus, for example, the classifier could choose pixels with only specific velocities for consideration in subsequent histograms. This feature may be used in conjunction with the output from other classification units to create histograms identifying only pixels meeting multiple classification criteria in a variety of domains.

38. The '001 Patent discloses that its teachings are applicable to a broad range of applications. For example, in one embodiment, the '001 Patent performs “automatic framing of a person . . . during a video conference.” Ex. 1001, '001 Patent at 22:10-12. In this application, histograms are constructed in the X- and Y-domains and count the number of pixels between successive frames where the differences in luminance are above certain threshold values. Ex. 1001, '001 Patent at 22:49-59. By this method, the '001 Patent teaches the system is able to determine the boundaries of the target based on peaks in the histograms generated. Ex. 1001, '001 Patent at 10:30-58. This application and result are shown in Figure 16 and 17, reproduced below:



Ex. 1001, Fig. 16



39. Other related applications disclosed by the '001 Patent include using a mounted spotlight or camera to automatically track a target, for example using a spotlight mounted on a helicopter to track a target on the ground or using automated stage lights to track a performer during a performance. Ex. 1001, '001 Patent at 23:40-44. The '001 Patent also teaches that “[t]he invention would similarly be applicable to weapons targeting systems.” Ex. 1001, '001 Patent at 23:44-45.

teaches that these key characteristics are computed by the condition:

For each pixel with a validation signal V2 of “1”:

(a) if the data value of the pixel < MIN (which is initially set to the maximum possible value of the histogram),

then write data value in MIN,

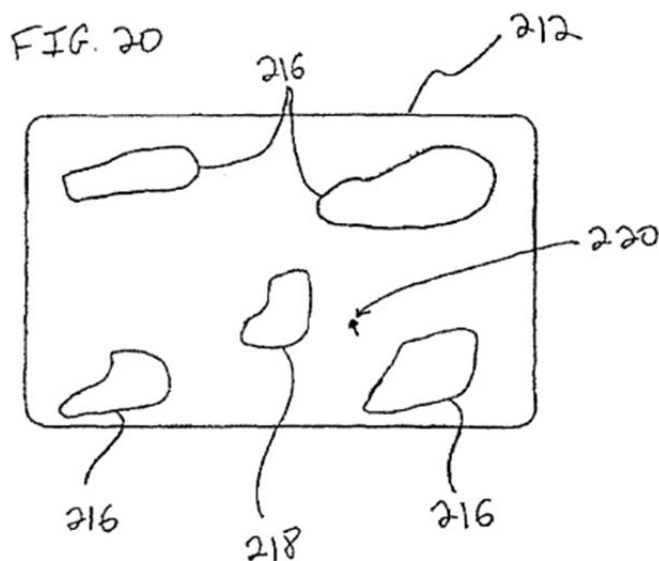
(b) if the data value of the pixel > MAX (which is initially set to the minimum possible value of the histogram), then

write data value in MAX

Id. at 19:56-62. Hence the '001 Patent defines the minimum of the X-projection histogram is the smallest X-coordinate of any pixel in the image region whose validation signal is “1.” Similarly the maximum is the largest X-coordinate of any pixel in the image region whose validation signal is “1.” The same holds true for the Y-projection histogram where the maximum and minimum are computed in the same way, but using the y coordinate axis.

41. After determining the X- and Y-minima and maxima of the histograms, the '001 Patent teaches that a center point of the target image may be found at the coordinates of $(X_{\text{MIN}} + X_{\text{MAX}})/2$ and $(Y_{\text{MIN}} + Y_{\text{MAX}})/2$. Ex. 1001, '001 Patent at 24:50-55. This is only one possible way of computing a center point of an object. Indeed, for an irregularly shaped object like those shown in Figure 20, below, there is not one clear center point. Rather, various center points such as

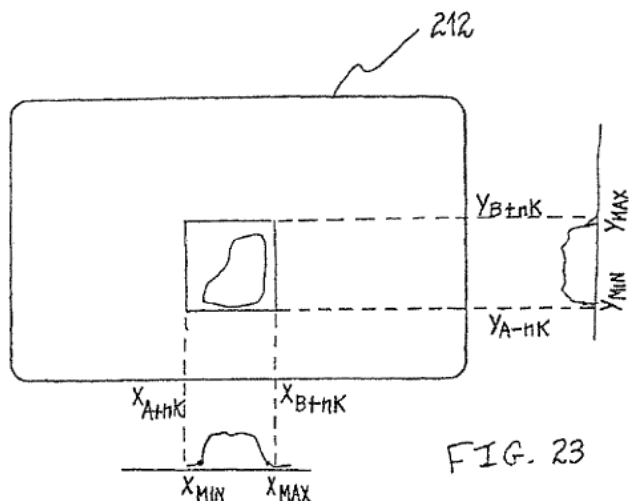
center of mass, center-of-area, and center of gravity might all give different coordinates, but would nevertheless accomplish the purpose of the '001 Patent.



42. After determining a center, the '001 Patent teaches that the system may determine the center position “at regular intervals, and preferably in each frame” and use this information “to modify the movement of camera 13 to center face V.” Ex. 1001, '001 Patent at 23:28-32. In other words, for each frame, the system would recalculate the histograms and use those points to again find the center coordinates.

43. In addition, the '001 Patent teaches that as part of the identification and tracking process, the imaging system may place a tracking box around the target that “may be displayed on monitor 212, or on another monitor as desired to visually track the target.” Ex. 1001, '001 Patent at 25:20-25. For example, Figure

23 shows an example of the tracking box in a frame:



V. SUMMARY OF OPINIONS

44. In preparing this declaration, I have reviewed at least the documents labeled Exhibits 1001-1010 and other materials referred to herein in connection with providing this declaration. In addition to these materials, I have relied on my education, experience, and my knowledge of practices and principles in the relevant field, e.g., image processing. My opinions have also been guided by my appreciation of how one of ordinary skill in the art would have understood the claims and specification of the '001 Patent around the time of the alleged invention, which I have been asked to assume is the earliest claimed priority date of July 22, 1996.

45. Based on my experience and expertise, it is my opinion that certain references teach or suggest all the features recited in Claims 1-4 of the '001 Patent, as explained in detail below. Specifically, it is my opinion that Claims 1-4 are

disclosed by Alton L. Gilbert et. al., *A Real-Time Video Tracking System*

(“Gilbert”) in combination with U.S. Patent No. 5,521,843 (“Hashima”). It is also my opinion that Claims 1-4 are disclosed by Hashima in combination with U.S. Patent No. 5,150,432 (“Ueno”). Finally, it is my opinion that Claims 1-4 are disclosed by Ueno in combination with Gilbert.

VI. LEVEL OF ORDINARY SKILL IN THE ART

46. Based on my review of the ’001 Patent specification, claims, file history, and prior art, I believe one of ordinary skill in the art around the time of the alleged invention of the ’001 Patent would have had either (1) a Master’s Degree in Electrical Engineering or Computer Science or the equivalent plus at least a year of experience in the field of image processing, image recognition, machine vision, or a related field or (2) a Bachelor’s Degree in Electrical Engineering or Computer Science or the equivalent plus at least three years of experience in the field of image processing, image recognition, machine vision, or a related field. Additional education could substitute for work experience and vice versa.

47. In determining the level of ordinary skill in the art, I was asked to consider, for example, the type of problems encountered in the art, 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.

48. My opinions concerning the '001 Patent claims are from the perspective of a person of ordinary skill in the art ("POSA"), as set forth above.

VII. CLAIM CONSTRUCTION

49. For my analysis, I have interpreted all claim terms according to their plain and ordinary meaning.

VIII. THE PRIOR ART TEACHES OR SUGGESTS EVERY STEP AND FEATURE OF THE CHALLENGED CLAIMS OF THE '001 PATENT

A. Overview Of The Prior Art References

1. **Alton L. Gilbert et al., A Real-Time Video Tracking System, PAMI-2 No. 1 IEEE Transactions on Pattern Analysis and Machine Intelligence 47 (Jan. 1980) ("Gilbert") (Ex. 1005)**

50. Gilbert arose out of efforts by researchers at U.S. Army White Sands Missile Range, New Mexico, in collaboration with New Mexico State University, Las Cruces, who developed a system that utilizes histograms to identify and track targets in flight such as rockets and aircraft. Gilbert was published in the January 1980 issue of IEEE Transactions on Pattern Analysis and Machine Intelligence, a peer-reviewed journal published by the Institute of Electrical and Electronics Engineers ("IEEE"). (Ex. 1010, Grenier Decl.)

51. Gilbert generally relates to "object identification and tracking applications of pattern recognition at video rates." Ex. 1005 at 47, Abstract. Gilbert discloses "a system for missile and aircraft identification and tracking . . .

applied in real time to identify and track objects.” Ex.1005, Gilbert at 47. It accomplishes this tracking by employing a series of processors, including an IO processor, a video processor, a projection processor, a tracker processor, and a control processor. Ex. 1005, Gilbert at 48. These elements are shown in Gilbert Figure 1:

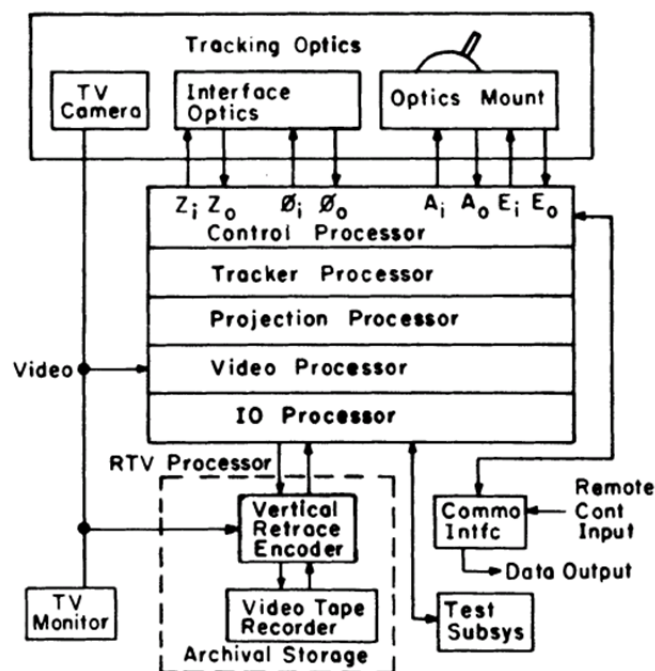


Fig. 1. RTV tracking system.

Ex. 1005, Gilbert at 48. Each of these processors performs a different role in identifying, locating, and ultimately tracking a target.

52. For example, the video processor receives a digitized video signal as input that comprises 60 frames/s. Although Gilbert uses the word “field” instead of “frame,” a POSA would have understood that a “frame” consists of two “fields” in the context in which they are used in Gilbert. (At the time, video signals were

interlaced such that a frame of a non-interlaced video consisted of two fields. The first field would be the odd numbered scanlines and the second field would be the even numbered scanlines recorded 1/60th of a second later than the first field.) . These frames each consist of a succession of pixels, arranged in a matrix of $n \times m$ pixels. Ex. 1005, Gilbert at 48. For each frame, the video processor creates a pixel intensity histogram across 256 gray levels. *Id.* These 256 gray levels represent the plurality of classes across which the histogram is formed. In other words, the Video Processor of Gilbert creates histograms using the luminance domain over classes of all luminance values. The video processor then uses these histograms and a probability estimate to determine whether a particular pixel belongs to the target, plume, or background region, thereby “separat[ing] the target from the background” and identifying the target. *Id.* This intensity histogram is entirely analogous to the velocity histogram in Figure 14a of the '001 Patent, as described at 20:54-59.

53. Although Gilbert specifically teaches creating histograms in the intensity (i.e., luminance) domain, it recognizes the possibility of using other characteristics derived from the relationship between pixels. For example, Gilbert recognizes the possibility that “texture, edge, and linearity measure” may also be used. *Id.*

54. After identifying the target, Gilbert teaches creating a binary picture

where the image is characterized by a series of 1s and 0s, where a “1” indicates the presence of the target and a “0” indicates its absence. *Id.* at 50. This binary selection process is entirely analogous to the histogram formation process described in the ’001 Patent at 18:16-19:24. In both cases a binary technique is used to determine which pixels meet certain criteria, and thus, which pixels will be used for further processing, including in the creation of projection histograms, as discussed below.

55. The projection processor is then used to identify the target location, orientation, and structure by analyzing the binary pixel matrices on a frame-by-frame basis. This binary picture is used by projection processor to create what the ’001 Patent and other prior art examples call “projection histograms” in the X- and Y-domains. Ex. 1005, Gilbert at 50. These projection histograms are analogous to the X- and Y-position histograms discussed in the ’001 Patent at 20:60-21:29. Figure 4 of Gilbert (annotated below) shows four different projection histograms formed using the target pixels:

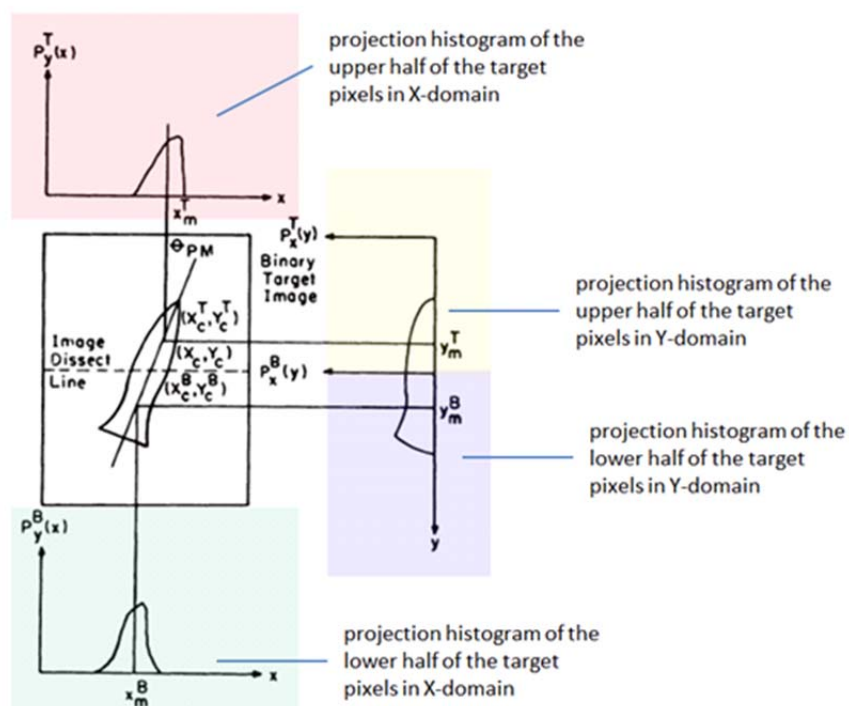


Fig. 4. Projection location technique.

In addition, Gilbert explains that a projection “gives the number of object points along parallel lines; hence it is a distribution of the target points for a given view angle.” Ex. 1005, Gilbert at 50. Thus, I have referred to these Figure 4 projections as projection histograms throughout this declaration.

56. The projection processor then computes a center position for the target by first dividing the image into a top and a bottom portion. *Id.* It then calculates a center position for both the top and the bottom portions of the target. *Id.* Each of these steps is accomplished using the X- and Y-projection histograms identified above. Finally, a line is drawn between the top-portion center-of-area and the bottom-portion center-of-area. *Id.* By calculating the mid-point of this line,

Gilbert determines a center point for the target image. *Id.* This process is described in more detail in the excerpt from Gilbert, below:

To precisely determine the target position and orientation, the target center-of-area points are computed for the top section (X_c^T , Y_c^T) and bottom section (X_c^B , Y_c^B) of the tracking parallelogram using the projections. Having these points, the target center-of-area (X_c , Y_c) and its orientation can be easily computed (Fig. 4):

$$X_c = \frac{X_c^T + X_c^B}{2}$$
$$Y_c = \frac{Y_c^T + Y_c^B}{2}$$
$$\phi = \tan^{-1} \frac{Y_c^T - Y_c^B}{X_c^T - X_c^B}.$$

The top and bottom target center-of-area points are used, rather than the target nose and tail points, since they are much easier to locate, and more importantly, they are less sensitive to noise perturbations.

Ex. 1005, Gilbert at 50. The results of this process are shown in Figure 4, reproduced below (with annotations):

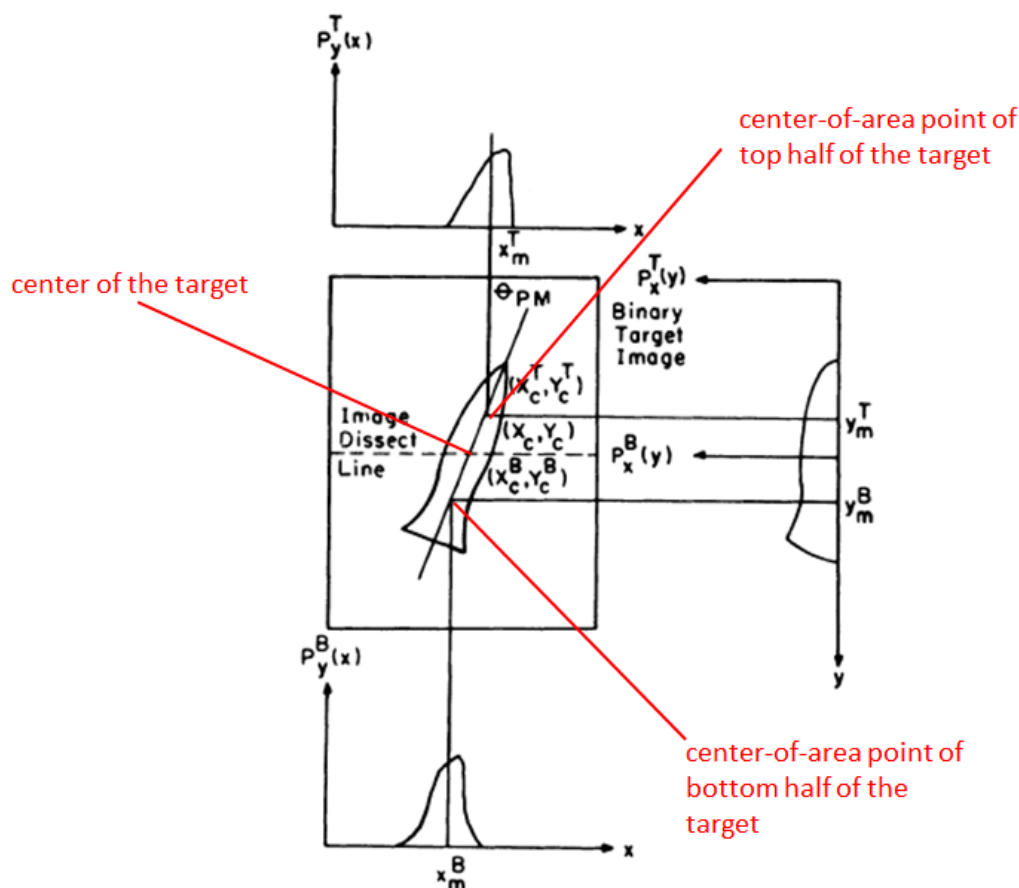


Fig. 4. Projection location technique.

57. Although Gilbert uses center-of-area calculations in the top and bottom halves of the target in order to calculate the center point because of considerations like noise perturbations and calculation of the orientation, it nonetheless teaches that actual maxima and minima such as the nose and tail points can also be used for the center calculation. Notably, the equations used to calculate X_C and Y_C would be exactly the same if extreme values were used rather than top and bottom center-of-area values as is actually done in Gilbert.

58. Gilbert further teaches that the histogram formation and target

identification process is performed on a frame-by-frame basis: “Outputs to Video Processor: 1) tracking window size, 2) tracking window shape, and 3) tracking window position. The outputs to the control processor are used to control the target location and size for the next frame.” *Id.* at 52.

59. The tracker processor analyzes information and controls the camera in order to keep the target centered and in the field of view. It analyzes information such as the target’s size, location, and orientation and outputs relevant information to both the control processor and the video processor. For example, the tracker processor outputs information such as “1) tracking window size, 2) tracking window shape, and 3) tracking window position.” *Id.* Using this information, the video processor is configured to draw a tracking box around the target image. This is shown in Figure 2, reproduced below:

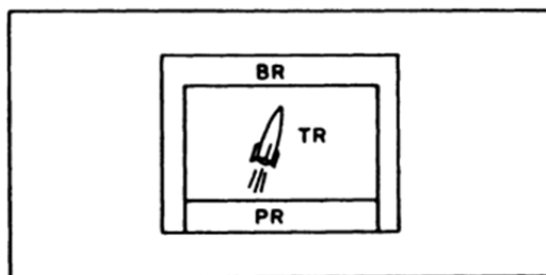
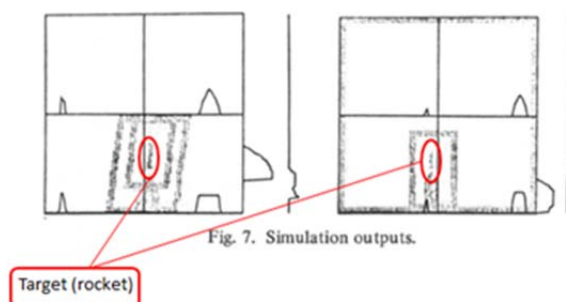


Fig. 2. Tracking window.

Through simulations, Gilbert demonstrated the effectiveness of coordination of the video processor and tracker processor by having the system place a tracking box

around the target in earlier video footage as shown in the annotated version of Figure 7 reproduced below:



60. Thus, it was well known to employ image processing systems to calculate histograms in multiple classes, in one of a plurality of domains on a frame-by-frame basis in order to track objects. Further, it was well known to determine a center point in a target as between X- and Y-minima and maxima of the target and to place a tracking box around the target.

2. U.S. Patent No. 5,521,843 (“Hashima”) (Ex. 1006)

61. Hashima discloses a “system for and method of recognizing and tracking a target mark . . . by processing an image of the target mark produced by a video camera.” Ex. 1006, Hashima at 1:6-12. As with the Gilbert, Hashima utilizes histograms to identify and track a target mark.

62. Hashima’s image processor reads input from the video camera, converts the image into a binary image, and constructs a histogram of the binary image:

FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. . . . The X-projected histogram represents the sum of pixels having the same X coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.

Ex. 1006, Hashima at 8:18-31.

63. After constructing the X- and Y- projection histograms, the image processor determines whether the image represents the target by counting the number of peaks and valleys in the projected histogram. Ex. 1006, Hashima at 9:8-9:13. If the number matches the number of peaks and valleys of the X- and Y-projected histograms, the system identifies the image as that of the target mark. *Id.* at 9:13-9:23. Hashima Figure 5 shows a flow chart diagramming this process, while Hashima Figure 6 demonstrates the X- and Y-histograms utilized:

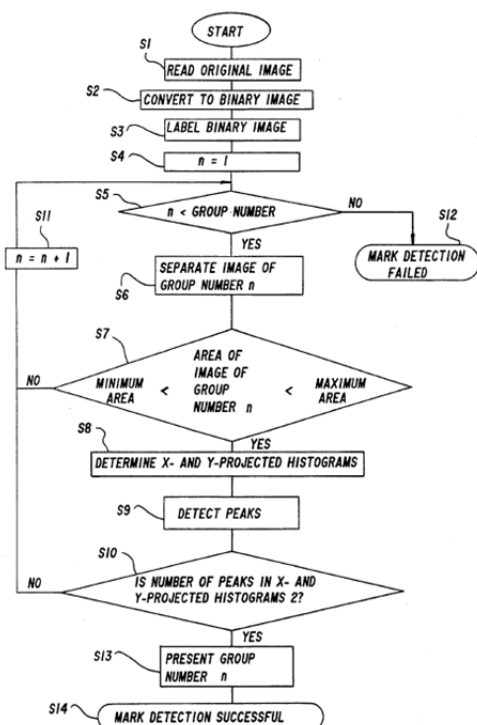


FIG. 5

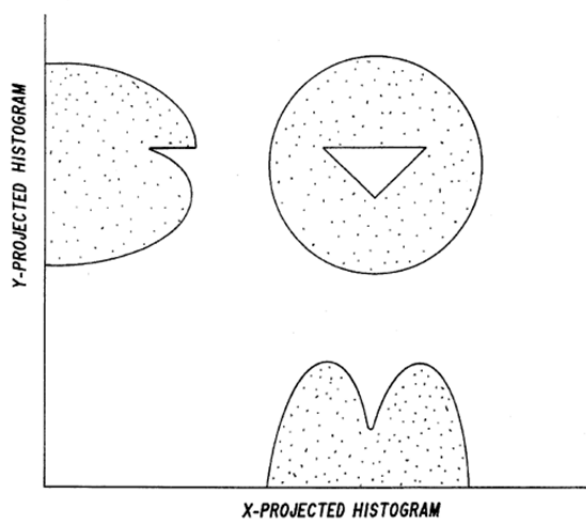


FIG. 6

64. This process of histogram generation and target identification is done on a frame-by-frame basis: “When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.” *Id.* at 24:23-27.

65. Hashima also explains that the X- and Y-minima and maxima are used to calculate the center of the target. *Id.* at 11:1-25. Specifically, Hashima teaches:

The image processor 40 . . . generat[es] a histogram 15 projected onto the X-axis and a histogram 16 projected onto the Y-axis, and then determines the X and Y coordinates of the central position Pm (mx, my) from the projected histograms 15, 16. Specifically, the X coordinate mx of the central position Pm (mx, my) can be determined using opposite end positions Xb1, Xb2 obtained from the X-projected histogram 15 according to the following equation (3):

$$mx=(Xb1+Xb2)/2 \quad (3).$$

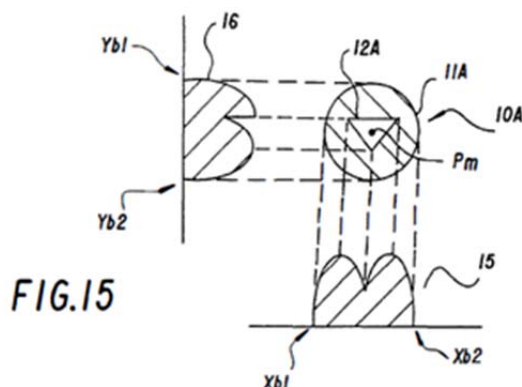
The Y coordinate my of the central position Pm (mx, my) can be determined using opposite end positions Yb1, Yb2 obtained from the Y-projected histogram 16 according to the following equation (4):

$$my=(Yb1+Yb2)/2 \quad (4).$$

Ex. 1006, Hashima at 11:6-25.

66. Figure 15 shows that the “opposite end positions” referred to in the quotation above are what the ’001 Patent calls the “X- and Y-minima and maxima.” In other words, the Xb1 and Yb2 are the smallest coordinates on the X- and Y-axis with a non-zero histogram value and Xb2 and Yb1 are similarly the

largest coordinates with a non-zero histogram value:



67. After the center of the target is determined, the center of the target is compared to the center of the image memory to find the shift of the target in the X- and Y-directions. The shift amount is used to move the robot arm (and the camera mounted thereon) toward the target, and is recalculated based on new images as the robot arm and the camera move toward the target, until the object is gripped. Fig. 27 shows a flow chart describing the process:

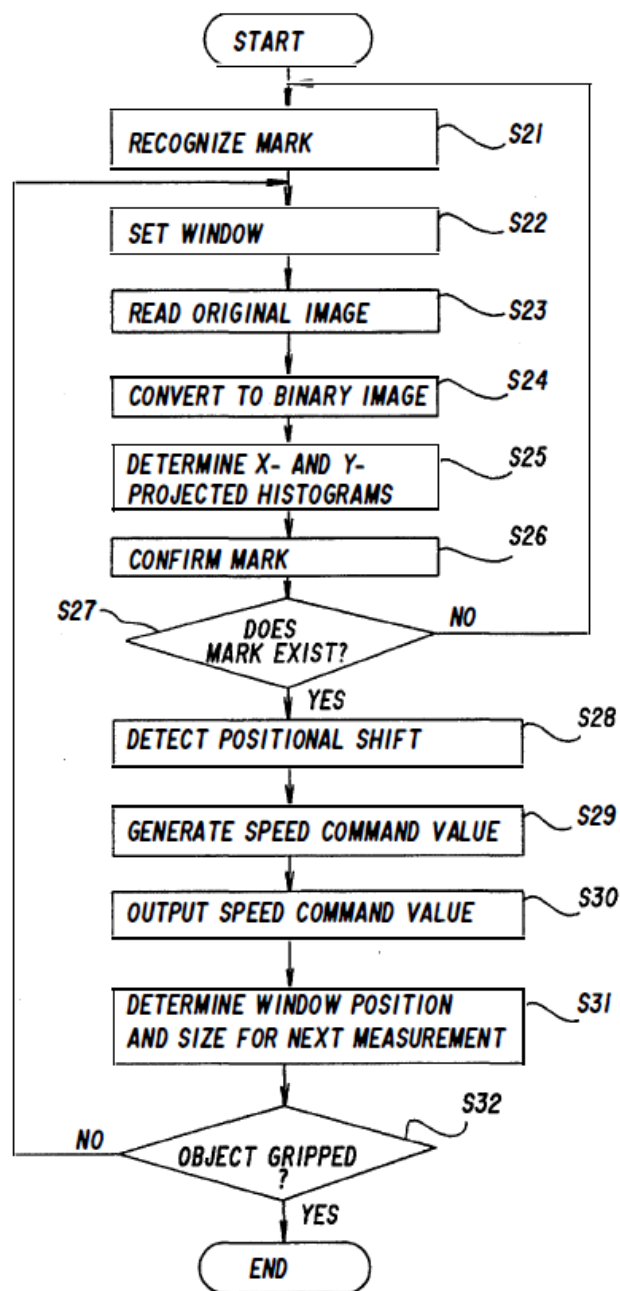


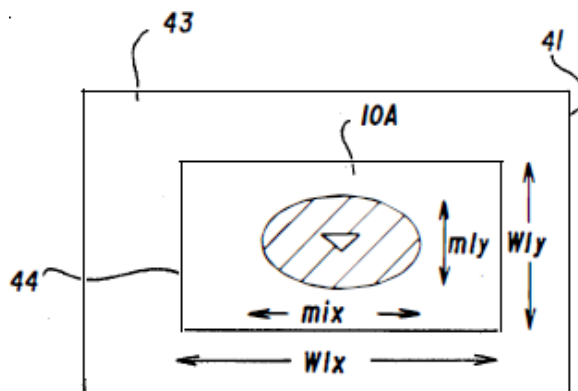
FIG.27

68. As the target is tracked, Hashima sets a rectangular window around the target—i.e., a tracking box. As the system receives each new frame from the video camera, the target and window locations are recalculated using the

histograms created from the processing of the new frame:

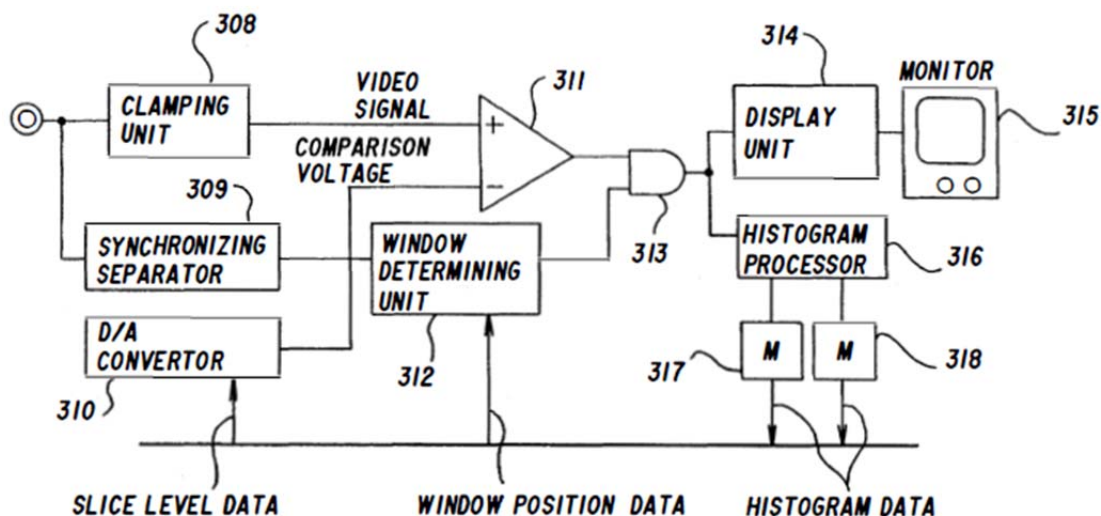
When the target mark 10 starts to be tracked, the window is established using the projected histogram information obtained when the target mark image is recognized. When the target mark 10 is subsequently tracked, the window [44] is established using new projected histogram information obtained upon each measurement made by the camera 20.

Ex. 1006, Hashima at 14:29-34. Figure 23 shows the target window 44 around the target mark 10A.



Hashima Fig. 23

69. A block diagram of the system in Hashima shows that after determining an appropriate window, the image is displayed on a monitor, 315. *See also*, Ex. 1007, Hashima at 25:34–38, Figure 52.



70. Like Gilbert, Hashima demonstrates that it was well known to employ image processing systems to calculate histograms in multiple classes, in one of a plurality of domains on a frame-by-frame basis in order to track objects. Further, it was well known to determine a center point in a target as between X- and Y-minima and maxima of the target and to place a tracking box around the target.

3. U.S. Patent No. 5,150,432 (“Ueno”) (Ex. 1007)

71. Ueno describes detecting the facial region in frames of video signals. In order to accomplish this, Ueno uses histograms in the X- and Y-domains to identify the top and bottom portions of the facial region (Y-minima and maxima), as well as the sides of the face (X-minima and maxima). Ex. 1007, Ueno at 7:17-45.

72. Ueno describes a video encoding system that includes “frame memory

101 for storing image signals for one frame” connected to “a facial region detecting circuit **102** for detecting human facial region of the frame data” *Id.* at 3:5-7. In Ueno, “video signals are sequentially input to the image input terminal **100** in frames.” *Id.* at 4:25-26. “The image signal is written in the frame memory **101** as frame data one frame by one frame. The frame data written in the frame memory **101** . . . is supplied to the facial region detecting circuit **102**. . . .” *Id.* at 4:33-37. The facial region detecting circuit of Ueno uses differences between two consecutive frames to create “interframe difference images.” Ueno then constructs histograms in X- and Y-domains using the interframe difference images.

73. FIG. 3 of Ueno (reproduced below) shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of “0” or “1” by the preset first threshold level. This process of binary conversion is analogous to the histogram formation process described in the ’001 Patent at 18:16-19:24. Then the number of pixels having the binary data value of “1” or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and projection histograms of the pixels (X-and Y-axis histograms) are formed. The face is detected based on these histograms.

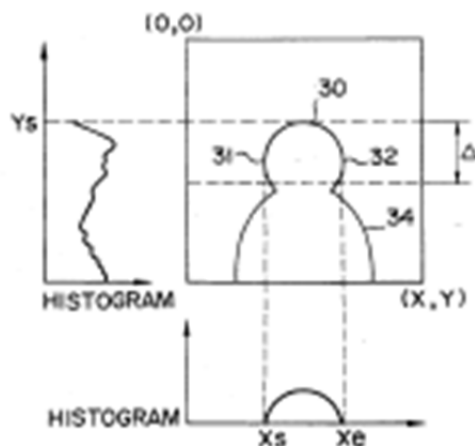


FIG. 3

74. Figure 6 shows block diagrams of the parallel histogram processors that create the X- and Y-histograms:

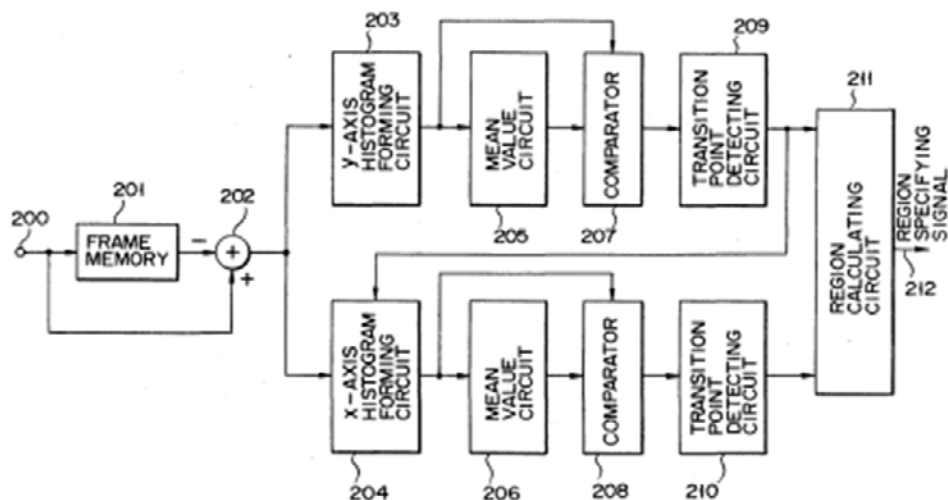
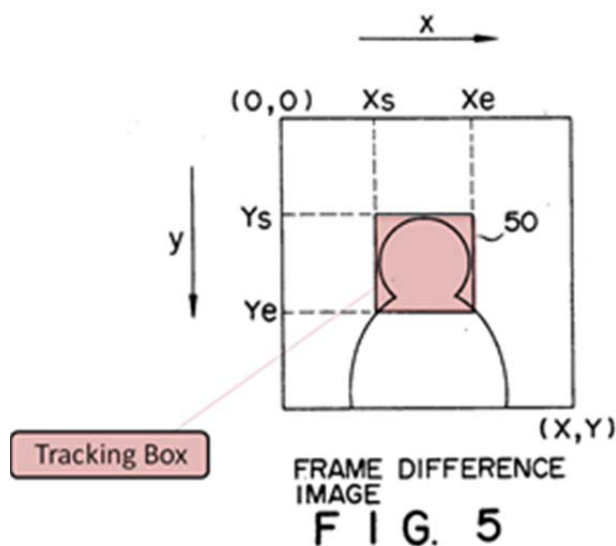


FIG. 6

75. The target, once identified, is tracked with a tracking box around the target “[b]y extracting the coordinates of the transition points of these histograms,” the face region is specified by a rectangle as shown in Figure 5 (annotated below):



76. Ueno specifically teaches that the tracking box is output to be displayed on the monitor by superimposing the box on the image:

In the self-monitoring mode, low-pass-filtered images are displayed on the monitoring display 710. After a region is specified by the region specifying apparatus 120, the region specifying signal is continuously output. Thus, said frame line showing the specified region is superimposed on the image and displayed on the monitoring display 710.

Ex. 1007, Ueno at 13:17-23.

77. Like Gilbert and Hashima, Ueno also demonstrates that it was well known to employ image processing systems to calculate histograms in multiple

classes, in one of a plurality of domains on a frame-by-frame basis in order to track objects. Further, it was well known to determine the X- and Y-minima and maxima of the target and to place a tracking box around the target.

B. Ground 1: Gilbert In View Of Hashima Teaches Or Suggests Every Step And Feature Of Claims 1-4

78. In my opinion, as shown in the charts below, Gilbert in view of Hashima and in view of the knowledge of a POSA, teaches or suggests every feature recited in Claims 1-4 of the '001 Patent.

1. Reasons To Combine Gilbert And Hashima

79. A POSA would have been motivated to combine Gilbert and Hashima because they are directed toward very similar systems that operate in a similar manner. As set forth above, both Gilbert and Hashima are “target recognition systems” directed to identifying and tracking a target using an image processing system with a video camera input (Ex. 1005, Gilbert at 47-48, Figure 1; Ex. 1006, Hashima at 1:6-15, 2:45-3:12, Figure 1), where the video input comprises a succession of frames, each frame comprising a succession of pixels (Ex. 1005, Gilbert at 47-48, 52; Ex. 1006, Hashima at 10:34-45, 24:13-25:33), in which the target comprises pixels in one or more of a plurality of classes and one or more of a plurality of domains (Ex. 1005, Gilbert at 48; Ex. 1006, Hashima at 8:18-30, 10:34-45). The Gilbert and Hashima systems perform this process on a frame-by-frame basis (Ex. 1005, Gilbert at 47-48, 52; Ex. 1006, Hashima at 24:13-25:23).

80. Additionally, both Gilbert and Hashima identify and track a moving target using histograms (Ex. 1005, Gilbert at 48-49; Ex. 1006, Hashima at 4:8-47), employ a camera that keeps the target in the field of view (Ex. 1005, Gilbert at 52; Ex. 1006, Hashima at 3:41-58), and recognize the importance of finding the center of a target (Ex. 1005, Gilbert at 50-51, Figure 4; Ex. 1006, Hashima at 11:1-30). Thus, they are not only in the same field of endeavor but also are directed to very similar systems and processes. A POSA would have recognized that Hashima's invention could improve a similar device, such as the system in Gilbert, in the same way, and thus would have been motivated to combine Gilbert and Hashima. A POSA would have also expected the combination to yield a predictable result because it would have involved applying known techniques to similar systems.

81. Such "target recognition systems" form a small, highly focused subject of the vast and general disciplines of image processing and computer vision. For example, one of the top conferences in these fields is the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), which was held in 1996 in San Francisco, publishing 137 papers in over 23 areas, with only one of these areas consisting of five papers on target recognition.

82. Indeed, there are many types of machine vision systems for tracking based on a variety of image analysis methods such as decomposition using Fourier, wavelet, or other bases, or feature detection yielding descriptors of edges, corners,

or other image features. Gilbert and Hashima both relate to the same, small, cross-section of these systems based on histograms, and specifically on projection histograms of classified and segmented image pixels (Ex. 1005, Gilbert at 48-49; Ex. 1006, Hashima at 4:8-47), employ a camera that keeps the target in the field of view (Ex. 1005, Gilbert at 52; Ex. 1006, Hashima at 3:41-58), and recognize the importance of finding the center of a target (Ex. 1005, Gilbert at 50-51, Figure 4; Ex. 1006, Hashima at 11:1-30). Thus, they are not only in the same field of endeavor but are directed to very similar systems and processes. A POSA would have been motivated to combine Gilbert and Hashima because the combination would have yielded a predictable result since it would have involved combining known techniques in a simple way to similar systems.

83. Moreover, the teachings in Gilbert and Hashima, taken as a whole, demonstrate additional reasons why a POSA would be motivated to combine the two references. For example, Gilbert recognizes that there are many possible domains that could be used for histogram plotting, but only explicitly teaches creating histograms in the intensity domain in a “video processor.” Ex. 1005, Gilbert at 48. Gilbert also discloses the construction of projections to determine the position of the target. These projections are entirely analogous to the projection histograms in the X- and Y-domain in Hashima (and entirely analogous to the position histograms in the ’001 Patent at 20:60-21:29).

84. Gilbert also uses X- and Y-projection histograms to identify the center of the target, and its orientation, after the pixels associated with the target have been separated from the plume and background. A POSA would have recognized that plotting histograms in other domains would result in certain advantages over the disclosure in Gilbert. For example, by having the capability of plotting histograms in additional domains, the system would have a higher likelihood of successfully recognizing a target in an image frame. This is because each additional domain provides another opportunity for correlation between that domain and the unique target being tracked. Additionally, by computing histograms across multiple domains simultaneously, a POSA would have been able to utilize the existing design of the histogram component of the “video processor” to increase system effectiveness.

85. Thus, a POSA would look to improve upon Gilbert by constructing histograms in other domains, and, in so doing, would look to references that teach creating histograms in other domains. Hashima is an example of a reference that teaches creating histograms in the X- and Y-domains to identify the target. Ex. 1006, Hashima at 4:8-47. Hashima showed that the projection described by Gilbert and performed by Gilbert’s “projection processor” were actually histograms using the X- and Y-coordinates of a class of image pixels. Thus, to identify the target, and improve on processing speed, a POSA would have been

motivated to combine Gilbert's image processing method with Hashima to replace Gilbert's projections with the X- and Y-domain histograms disclosed in Hashima. Moreover, there would have been a reasonable expectation of success that doing so would result in a more efficient target identification system.

86. A POSA would have also considered combining Gilbert with Hashima on the basis of Gilbert's inefficient center point calculation. Gilbert teaches to find a center of the target by first splitting the target image into a top and a bottom half. Next, a center-of-area is calculated for the top half and for the bottom half independently. Finally, Gilbert teaches finding the mid-point between the top-half center point and the bottom-half center point and discloses that this point is a center. Gilbert uses this center calculation to keep the target centered within the frame, making automatic adjustments to the direction of the camera to achieve this end. Gilbert teaches a complex method of determining a center that involves many computational steps. Gilbert teaches this center calculation to allow for a determination of the rocket's orientation in order to predict its trajectory. Gilbert also uses the center calculated using the two points to direct the lens toward the moving target such that the target remains in the camera's field of view. However, a POSA at the time of the '001 Patent would have recognized that a simpler center calculation for the entire target image would have been able to achieve all of the benefits of Gilbert while imposing a significantly reduced

computational burden. For example, by calculating the center point of the target for each of at least two frames, the orientation (and velocity) can be calculated more efficiently and robustly by simply using the time interval between the any two frames and calculating the distance travelled by the target from the change in location of the two center points from the first frame to the second.

87. Moreover, a POSA would have recognized that this modification would also have improved Gilbert so that it could track a wider variety of targets. This is because the Gilbert centering calculation—while providing useful orientation detail about elongated narrow objects like rockets—is unreliable for finding the orientation of objects with more general shapes or more general configurations. For example, the Gilbert method would fail to detect the orientation of a horizontal missile. Similarly, it is not adept to find the orientation of a more rounded object such as the triangle-disk marker used by Hashima and others. However, by adopting a simpler center point calculation, Gilbert would have had the flexibility to track objects with varying shapes in varying orientations.

88. As a result, it would have been desirable for a POSA to combine Gilbert with other references teaching simpler center calculations based on the end-points of a target in a projection histogram (i.e., X-minima and maxima and Y-minima and maxima), such as the calculation disclosed by Hashima. As discussed above, Hashima teaches the center coordinates “can be determined using the

opposite end positions” from the projected X- and Y-histograms, or in other words by using the min and the max. Upon finding the min and max, Hashima teaches a simple center calculation: $mx=(Xb1+Xb2)/2$, $my=(Yb1+Yb2)/2$.

89. A POSA would be motivated to make such a combination because it would reduce the number of calculations needed to calculate the center, resulting in faster, more efficient processing and improved tracking of a target. Indeed, rather than making three separate center calculations (one for the top, one for the bottom, and one for the overall target), a system that combined Gilbert and Hashima would only need to make a single calculation. These improvements could also reduce hardware expenses and use circuit die space more efficiently since fewer calculations could be computed in the same amount of time using less silicon. Moreover, a POSA would have reasonably expected that such a combination would have been successful in that it would have resulted in a more efficient computation system.

2. Claim 1

- a) **[1 pre]: A process of tracking a target, in an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels, the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains, the process performed by said system comprising, on a frame-by-frame basis**

90. The preamble of Claim 1 requires four different elements: (1) that the

process “track[] a target,” (2) that a specific input signal comprises a sequence of frames where each frame comprises multiple pixels, (3) that the pixels must be categorized according to their characteristics into at least one domain and one class selected from a plurality of domains and classes, and (4) that the method must repeat itself “on a frame-by-frame basis.” Both Gilbert and Hashima disclose each of these elements as described below, and as more fully laid out in the claim chart in **IX.B.6**.

91. Gilbert explicitly discloses the requirement for “[a] process of tracking a target.” As stated in the abstract and introduction of Gilbert, the entire point of developing Gilbert’s system was to “produce[] a system for missile and aircraft identification and tracking.” Ex. 1005, Gilbert at 47. Specifically, the system is capable of identifying the object, locating its center, and tracking it to keep it in the field of view of the camera, even as it is rapidly moving through space. *Id.* Hashima similarly discloses “[a] system for and a method of recognizing and tracking a target mark with a video camera.” Ex. 1006, Hashima at 1:7-9.

92. Gilbert also discloses the requirement for “an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels.” This limitation is met in its entirety simply because the input described is

a video signal acquired by a TV camera. *Id.* at 48. Any such video input inherently comprises a sequence of frames and each frame in a video input would also comprise many pixels. Nevertheless, Gilbert also explicitly describes both the frames and the pixels that comprise the input signal:

The scene in the FOV of the TV camera is digitized to form an $n \times m$ matrix representation

$$P = (p_{ij})_{n, m}$$

of the pixel intensities p_{ij} . As the TV camera scans the scene, the video signal is digitized at m equally spaced points across each horizontal scan. During each video field, there are n horizontal scans which generate an $n \times m$ discrete matrix representation at 60 fields/s. A resolution of $m = 512$ pixels per standard TV line results in a pixel rate of 96 ns per pixel.

Ex. 1005, Gilbert at 48. As with Gilbert, the fact that Hashima's input signal is a video signal is sufficient to meet this limitation. However, Hashima also explicitly teaches that "one frame of scanned image data from a camera is stored in a frame memory," and that "[a]n image . . . is read into an image memory . . . [which] is composed of 512 pixels in each of the X- and Y-directions." Thus, both Gilbert and Hashima disclose the input signal taught in the '001 Patent.

93. Gilbert meets the requirement for a “target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains.” The pixels described by Gilbert can also be categorized into classes and domains. This limitation is very similar to the step in [1 a] below, inasmuch as domains merely refer to pixel characteristics for use in a histogram and classes refer to the discrete divisions, i.e., bins, within the domain into which the pixels can be classified. *See* Ex. 1001, ’001 Patent at 4:2-6 (listing specific pixel characteristics for use as domains); 6:1-3 (the pixels are classified “within each domain to selected classes within the domain”). Gilbert specifically teaches that there are “many features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures.” Ex. 1005, Gilbert at 48. Gilbert focuses primarily on the pixel characteristic or domain of intensity, which it describes as being divisible into 256 gray levels or classes. *Id.* Likewise, the pixels in the input image in Hashima form a class within a domain. Although many of the same pixel characteristics must be present in Hashima as intrinsic pixel qualities, Hashima expressly teaches forming histograms in the X- and Y-domains: “The X-projected histogram represents the sum of pixels having the same X-coordinates, and the Y-projected histogram represents the sum of pixels having the same Y-coordinates.” Ex. 1006, Hashima at 8:27-31. In this case the class of pixels are pixels sharing the same value (e.g., 1) in the binary version of the image (using the pixels’ intensity),

and the domain of the pixels are pixel characteristics, including their intensity, and their X- or Y-coordinate depending on whether the X- or Y-projection histogram is computed. *See* Ex. 1001, '001 Patent at 4:2-6. Thus, both Gilbert and Hashima disclose pixels that can be categorized in a plurality of domains and a plurality of classes.

94. Gilbert also meets the limitation that “the process [be] performed by said system . . . on a frame-by-frame basis.” Gilbert teaches that the video input signal is comprised of 30 individual frames (60 fields) per second: “During each video field, there are n horizontal scans which generate an $n \times m$ discrete matrix representation at 60 fields/s.” Ex. 1005, Gilbert at 48. Although Gilbert uses the word “field” instead of “frame,” a POSA would have understood that a “frame” consists of two “fields” in the context in which they are used in Gilbert. (At the time, video signals were interlaced such that a frame of a non-interlaced video consisted of two fields. The first field would be the odd numbered scanlines and the second field would be the even numbered scanlines recorded 1/60th of a second later than the first field.) Gilbert further teaches that “[d]uring each field, the feature histograms are accumulated for the three regions of the tracking window.” *Id.* Like Gilbert, Hashima also discloses that its process is performed on a frame-by-frame basis. Specifically, it teaches that “it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera.” Ex. 1006, Hashima

at 24:24-25. Thus, both Gilbert and Hashima teach that the image tracking and identification process is performed on a frame-by-frame basis as claimed in the '001 Patent.

- b) **[1 a]: forming at least one histogram of the pixels in the one or more of a plurality of classes in the one or more of a plurality of domains, said at least one histogram referring to classes defining said target**

95. Both Gilbert and Hashima teach this step, as discussed below and as further shown in the claim chart at **IX.B.6**. In each case, at least one histogram is formed based on domains, namely characteristics of a collection of pixels. Examples of domains are seen in the '001 Patent: “The domains are preferably selected from the group consisting of i) luminance, ii) speed (V), iii) oriented direction (D1), iv) time constant (CO), v) hue, vi) saturation, and vii) first axis (x(m)), and viii) second axis (y(m)).” Ex. 1001, '001 Patent at 4:2-6. Gilbert similarly lists a number of pixel characteristics such as “texture, edge, and linearity measures” that could be used for histogram creation, and specifically discloses that “pixel intensity is used to describe the pixel feature chosen.” Ex. 1005, Gilbert at 48. Gilbert teaches that “pixel intensity is digitized and quantized into eight bits (256 gray levels) counted into one of six 256-level histogram memories.” These gray levels constitute the classes within the intensity domain. The target, or targets, are defined by the classes because Gilbert takes it as a basic assumption that “the target image has some video intensities not contained in the immediate

background.” *Id.* at 48. After dividing the image into background, plume and target regions, Gilbert teaches that “feature histograms are accumulated” for each of the three regions. *Id.*

96. Moreover, to the extent the Patent Owner argues that Gilbert’s intensity histogram does not include a plurality of “classes,” because the histogram contains every pixel (sorted into bins based on intensity level) rather than a subset of pixels that meet a threshold intensity characteristic, this argument should be rejected. Gilbert’s intensity histogram is entirely analogous to the velocity histogram in Figure 14a of the ’001 Patent, as described at 20:54-59.

97. Gilbert teaches the possibility of creating histograms in additional domains using pixel features other than intensity such as “texture, edge, and linearity measures.” Ex. 1005, Gilbert at 48.

98. Hashima explicitly teaches forming histograms using the X- and Y-axis domains: “X- and Y-projected histograms of the target mark image are determined. . . . The X-projected histogram represents the sum of pixels having the same X-coordinates, and the Y-projected histogram represents the sum of pixels having the same Y-coordinates.” Ex. 1006, Hashima at 24:14-21. X- and Y-axis are two of the domains expressly given as examples in the ’001 Patent. Ex. 1001, ’001 Patent at 4:2-6. When using the X- and Y-axis as the domains, the classes are the coordinates (or coordinate ranges) along those axes. These projection

histograms are analogous to the X- and Y- position histograms discussed in the '001 Patent at 20:60-21:29.

99. Gilbert similarly discloses that its projection processor forms X- and Y-projections using binary pictures generated by the video processor. As discussed above, these projections are in fact projection histograms like those used in Hashima. Moreover, these projection histograms are analogous to the X- and Y-position histograms discussed in the '001 Patent at 20:60-21:29. Thus, Gilbert discloses creating histograms in a plurality of domains, including the intensity, the X-, and the Y-domains.

100. Thus, both Gilbert and Hashima teach that histograms are formed in at least one of a plurality of domains with each pixel being categorized into one or more classes within said domain.

101. To the extent the Patent Owner argues that Gilbert's discussion of a plurality of "features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures" (*id.*) is insufficient to satisfy "a plurality of domains," Hashima explicitly discloses a plurality of domains by showing histograms on X-domain and Y-domain. In these histograms, each "class" within a domain counts the number of pixels meeting certain threshold criteria. For example, Gilbert's intensity histogram has 256 gray levels (i.e., 256 classes). The histogram is formed as each pixel in the frame is sorted into one of

the 256 classes that corresponds to that pixel's intensity level. In Hashima, the number of pixels having the value of "1" within a given column defined by an X-coordinate is counted into the class that corresponds to that X-coordinate value. To the extent the Patent Owner argues that Gilbert's discussion of a plurality of "features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measure" (Ex. 1005, Gilbert at 48), is insufficient to satisfy a "plurality of domains," Hashima explicitly discloses a plurality of domains by showing histograms on X-domain and Y-domain. *See* Ex. 1006, Hashima at 8:22-30.

c) [1 b]: identifying the target in said at least one histogram itself

102. Both Gilbert and Hashima disclose this step, as discussed below and shown in greater detail in the claim chart at **IX.B.6**. In both references, histograms are used to identify the target. Gilbert uses probability estimates based on the 256 level grayscale histograms to determine whether a particular pixel belongs to the target, plume, or background region. Ex. 1005, Gilbert at 48-50.

103. Similarly, in Hashima, peaks and valleys in the X- and Y-axis histograms are used to determine the location of the target: "Target marks of various configurations which are characterized by the extreme values of the X- and Y-projected histograms can be detected by the above process." Ex. 1006, Hashima at 10:20-24. In other words, the target mark is identified in Hashima directly from

the X- and Y-axis histograms. The '001 Patent describes an identical process for identifying a face from the histograms in Figure 17 and the accompanying description at 22:60-23:14.

- d) **[1 c]: wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target**

104. Both Gilbert and Hashima disclose the step of determining a center of the target, as described below and shown in more detail in the claim chart at **IX.B.6**. In Gilbert for example, a center is computed to “precisely determine the target position and orientation.” Ex. 1005, Gilbert at 50. There, the target image is divided into a top and bottom section and a center point is found for each. *Id.* Once finding a center point for the top and bottom portions, the “target center-of-area (X_C , Y_C) and its orientation can be easily computed” (*id.*) by connecting the top and bottom center points with each other and finding the midpoint of the resulting line. For a single body target, this point must necessarily be between the X- and Y-minima and maxima of the target. The centering calculation used in Gilbert is shown diagrammatically in Figure 4, below:

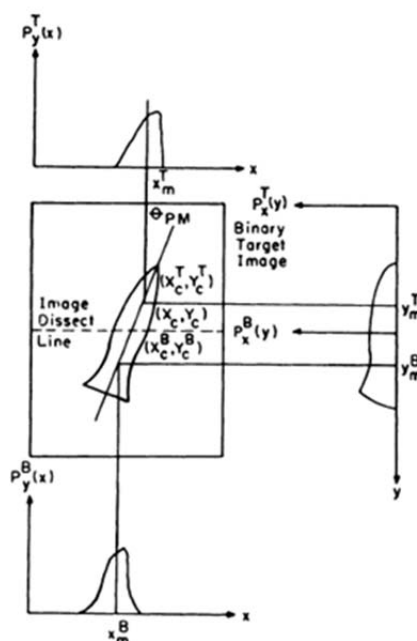


Fig. 4. Projection location technique.

Gilbert also suggests that it is possible for “the target nose and tail points” to be used in finding the center, and acknowledges that the “centroid of the target image” may be located solely on the basis of the vertical and horizontal projections without performing the complex center-of-area calculation discussed above. *Id.* at 50.

105. Hashima also discloses a method of finding a center point of the target image and explains how the X- and Y-minima and maxima are used to calculate the center of the target. Specifically, Hashima teaches:

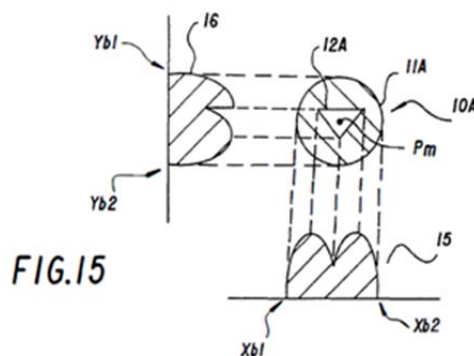
[T]he X coordinate m_x of the central position P_m (m_x , m_y) can be determined using opposite end positions X_{b1} , X_{b2} obtained from the X-projected histogram 15 according to the following equation (3):

$$m_x = (X_{b1} + X_{b2}) / 2 \quad (3).$$

The Y coordinate m_y of the central position P_m (m_x , m_y) can be determined using opposite end positions Y_{b1} , Y_{b2} obtained from the Y-projected histogram 16 according to the following equation (4):

$$m_y = (Y_{b1} + Y_{b2}) / 2 \quad (4).$$

Ex. 1006, Hashima at 11:13-24. Figure 15 shows that the “opposite end positions” referred to in the quotation above are the X- and Y-minima and maxima from the X- and Y-histograms and illustrates how they are used to compute the center of the target:



The equations used to find the center in Hashima are identical to those described for the same purpose in the '001 Patent at 24:50-55.

106. A POSA would not have understood Claim 1 to require a specific equation be used to calculate a center point. Indeed for irregularly shaped objects, like those tracked in the '001 Patent, the concept of a center point is ambiguous because there are actually various center points that could be calculated according to different methods. For example, a center could be the center of mass, or the mean of the coordinates of the border of the object, or the point with the minimal sum of distances to all points of the area measured along paths inside the area, or could have one of a number of other definitions. The only restriction that Claim 1 places on the center calculation is that it must be between the X- and Y-minima and maxima, or in other words, on the target itself. Using any of the centers described above, so long as they were on the target itself, would accomplish the aim of aiding in identification and tracking of the target.

107. To the extent the Patent Owner argues that Claim 1 requires that the X- and Y-minima and maxima be used in the calculation of the center point, it may be argued that Gilbert's disclosure, or at least the calculation performed in Gilbert, varies in some degree from this requirement. However, to the extent that Gilbert does not disclose this precise centering calculation, it is disclosed by Hashima, as discussed above.

3. Claim 2

- a) **The process according to Claim 1, wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target**

108. Claim 2 contains a limitation identical to the final limitation in Claim 1. Thus, it is my opinion that Gilbert and Hashima both disclose this limitation for the same reasons discussed above. *See* Section **IX.B.2.d**.

4. Claim 3

- a) **The process according to Claim 1, wherein identifying the target in said at least one histogram further comprises determining the center of the target at regular intervals**

109. Both Gilbert and Hashima teach that the target identification process, including the center determination, take place at regular intervals. For example, Gilbert discloses forming X- and Y-projection histograms on a frame-by-frame basis at a rate of 30 frames/second (60 fields/second). *See, e.g.*, Ex. 1005, Gilbert at 50 (“In the projection processor, these matrices are analyzed field-by-field at 60 fields/s using the projection based algorithm”). As discussed above, Gilbert uses the word “field” as a substitute for “frame.” *See also* Ex. 1005, Gilbert at 55 (“The result after six processing frames . . .”). Because part of the analyzing or processing referred to involves generating a histogram to identify the target, and finding the center of the target using the generated histogram, this process is

necessarily performed at regular intervals, namely once per frame or once every 1/60 second.

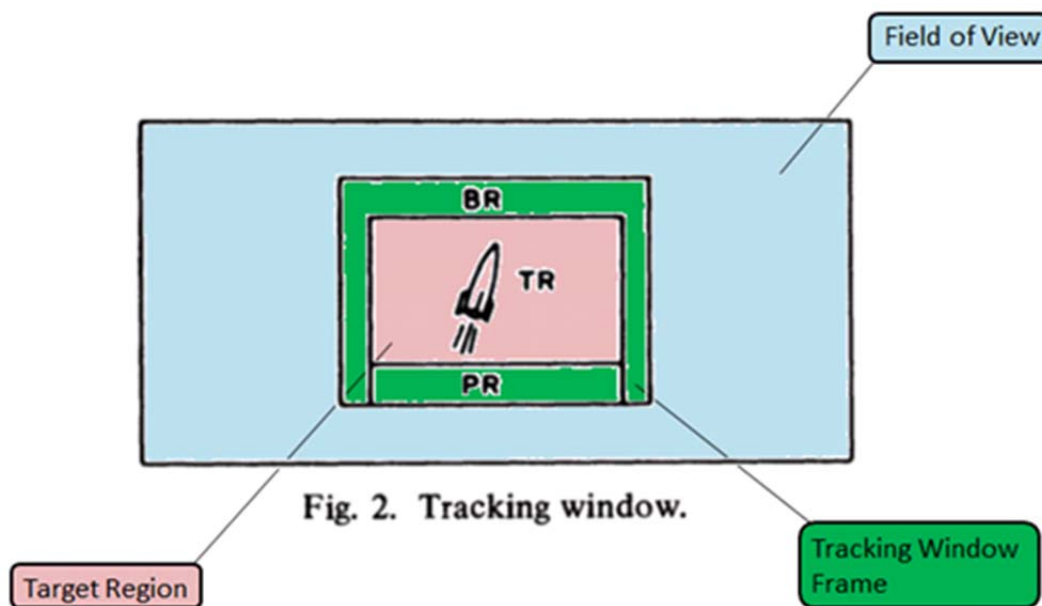
110. Hashima similarly discloses processing images on a frame-by-frame basis as they are supplied from the camera. For example, Hashima teaches that “When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of 25 image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.” *See, e.g.*, Ex. 1006, Hashima at 24:13-35. Hashima discloses embodiments where this process requires “at least 200 ms,” *id.*, and other embodiments where the “processing time corresponds to only one frame (33 ms).” *Id.* at 25:16-33. Thus, because determining the center of the target is part of the tracking process in both embodiments, it is done at the regular interval of 200 ms or on a frame-by-frame basis every 33 ms. *Id.* at 25:16-33.

111. The frame-by-frame determination of the center of the target in Gilbert and Hashima is identical to the description in the '001 Patent at 25:3-25 of determining the center position at regular intervals. To the extent Gilbert and Hashima do not explicitly disclose that the center of the target is determined at regular intervals, a POSA would have considered implementing this feature in order to achieve smooth, predictable processing.

5. Claim 4

a) The process according to Claim 1 further comprising drawing a tracking box around the target

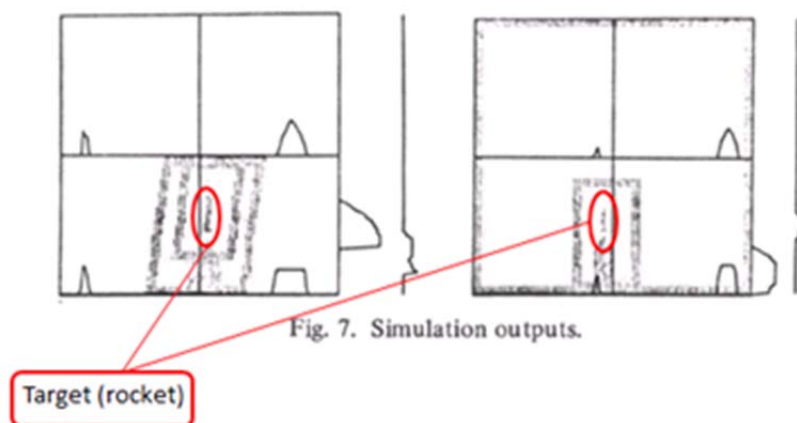
112. Gilbert discloses the step of drawing a tracking box around the target. For example, Gilbert teaches that “[a] tracking window is placed about the target image, . . . The tracking window frame is partitioned into a background region (BR) and a plume region (PR). The region inside the frame is called the target region (TR) as shown in Fig. 2.” Ex. 1005, Gilbert at 48. This tracking window is shown in Figure 2 from Gilbert as annotated below:



Gilbert also shows that in simulations (using pre-recorded video of a rocket in flight as an input signal), the system was successful in tracking target by drawing a tracking box. *See, e.g.*, Ex. 1005, Gilbert at 55 (“Fig. 7 contains two representative

simulation outputs selected from the first 100 fields of simulated digitized video.”).

This is shown in Figure 7 from Gilbert, as annotated below:



113. Thus, Gilbert discloses this limitation. To the extent the Patent Owner argues that Gilbert does not explicitly disclose drawing a tracking box around the target, a POSA would have considered implementing this feature in order to aid a user at a display to visually locate and follow a target.

114. Hashima also discloses a tracking box around the target: “each of the images grouped by the labeling process is separated by a window process.” Hashima at 8:55-57. This window is “established” by Hashima using the technique described at 13:61-14:34. Furthermore, the system disclosed in Fig. 53 described by Hashima at 24:44-25:23 includes a “window determining unit 312” that uses the “window position data” to restrict the video signal to only include the window contents through the “AND gate 313” and out to the “display unit 314” to the “monitor 315.” In this configuration, only the contents of the window would

be displayed on the monitor and the remainder of the screen would be black,
thereby drawing a (black) tracking box around the target currently under
consideration.

6. Detailed Application Of Gilbert And Hashima To Claims 1-4

115. The charts below provide a detailed mapping of Gilbert and Hashima to Claims 1-4 of the '001 Patent and show that Gilbert in view of Hashima teach or suggest every feature recited in Claims 1-4:

Claim 1	Exemplary Disclosures
[1.pre] A process of tracking a target in an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels, the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains, the process performed by said system comprising, on a frame-by-frame basis	Ex. 1005, Gilbert at 47-48: "The main optics is a high quality cinetheodolite used for obtaining extremely accurate (rms error - 3 arc-seconds) angular data on the position of an object in the FOV. It is positioned by the optical mount which responds to azimuthal and elevation drive commands, either manually or from an external source. The interface optics and imaging subsystem provides a capability to increase or decrease the imaged object size on the face of the silicon target vidicon through a 10:1 range, provides electronic rotation to establish a desired object orientation, performs an autofocus function, and uses a gated image intensifier to amplify the image and "freeze" the motion in the FOV. The camera output is statistically decomposed into background, foreground, target, and plume regions by the video processor, with this operation carried on at video rates for up to the full frame. The projection processor then analyzes the structure of the target regions to verify that the object selected as "target" meets the stored (adaptive) description of the object being tracked. The tracker processor determines a position in the FOV and a measured orientation of the target, and decides what level of confidence it has in the data and decision. The control processor then generates commands to orient the mount, control the interface optics, and provide real-

Claim 1	Exemplary Disclosures
	<p>time data output. An I/O processor allows the algorithms in the system to be changed, interfaces with a human operator for tests and operation, and provides data to and accepts data from the archival storage subsystem where the live video is combined with status and position data on a video tape.”</p> <p>Ex. 1005, Gilbert at 48: “The four tracking processors, in turn, separate the target image from the background, locate and describe the target image shape, establish an intelligent tracking strategy, and generate the camera pointing signals to form a fully automatic tracking system.”</p> <p>Ex. 1005, Gilbert at 48: “The video processor receives the digitized video, statistically analyzes the target and background intensity distributions, and decides whether a given pixel is background or target [8]. A real-time adaptive statistical clustering algorithm is used to separate the target image from the background scene at standard video rates. The scene in the FOV of the TV camera is digitized to form an $n \times m$ matrix representation</p> $P = (p_{ij})_{n, m}$ <p>of the pixel intensities P_{ij}. As the TV camera scans the scene, the video signal is digitized at m equally spaced points across each horizontal scan. During each video field, there are n horizontal scans which generate an $n \times m$ discrete matrix representation at 60 fields/s. A resolution of $m = 512$ pixels per standard TV line results in a pixel rate of 96 ns per pixel.”</p> <p>Ex. 1005, Gilbert at 48: “Every 96 ns, a pixel intensity is digitized and quantized into eight bits (256 gray levels), counted into one of six 256-level histogram memories, and then converted by a decision memory to a 2-bit code indicating its classification (target, plume, or background). There are many features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures. Throughout the following discussion of</p>

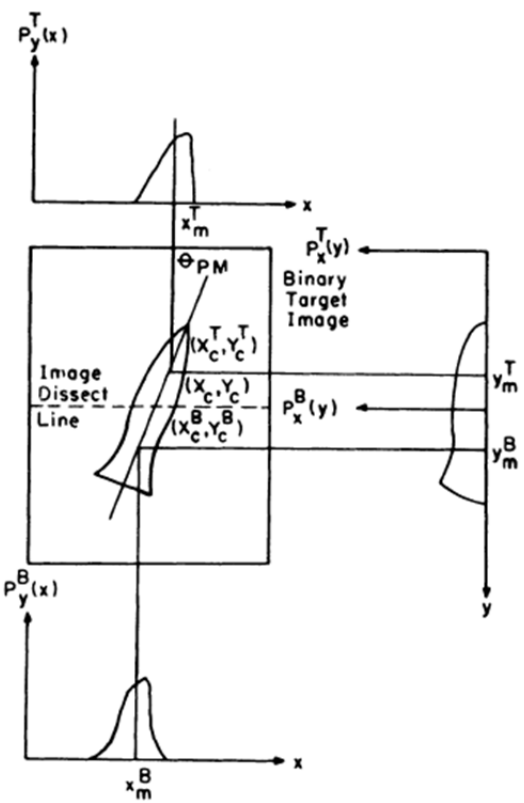
Claim 1	Exemplary Disclosures
	<p>the clustering algorithm, pixel intensity is used to describe the pixel features chosen.”</p> <p>Ex. 1005, Gilbert at 52: “The outputs to the control processor are used to control the target location and size for the next frame.”</p> <p>Ex. 1006, Hashima at 1:6-16: “The present invention relates to a system for and a method of recognizing and tracking a target mark using a video camera, and more particularly to a system for and a method of recognizing and tracking a target mark for detecting the position and attitude of the target mark by processing an image of the target mark produced by a video camera, detecting a shift of the position of the target mark from a predetermined position, and controlling the position and attitude of a processing mechanism based on the detected shift.”</p> <p>Ex. 1006, Hashima at 8:18-30: “FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. As shown in FIG. 6, the extreme values (maximum or minimum values which may be called "peaks" or "valleys") of each of the X- and Y-projected histograms comprise two peaks and one valley. The number of these extreme values remain unchanged even when the target mark rotates. The X-projected histogram represents the sum of pixels having the same X coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.”</p> <p>Ex. 1006, Hashima at 10:34-45: “FIG. 14 is illustrative of the manner in which shifts in X- and Y-directions of a target mark are determined. In FIG. 14, the target mark 10 is imaged as being shifted from the center of the screen of the camera 20. An image (target mark image) 10A of the</p>

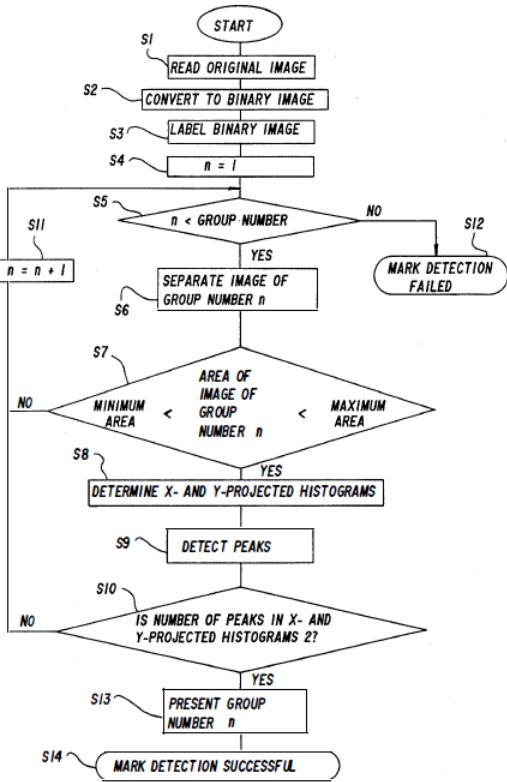
Claim 1	Exemplary Disclosures
	<p>target mark 10 which is imaged by the camera 20 is read into an image memory 41 of the image processor 40, and recognized in an image coordinate system 42. The target mark image 10A is composed of a black circular image 11A and a white triangular image 12A. The image memory 41 is composed of 512 pixels in each of X and Y directions. The coordinate position of each point in the image memory is indicated according to the unit of pixels.”</p> <p>Ex. 1006, Hashima at 14:29-34: “When the target mark 10 starts to be tracked, the window is established using the projected histogram information obtained when the target mark image is recognized. When the target mark 10 is subsequently tracked, the window is established using new projected histogram information obtained upon each measurement made by the camera 20.”</p> <p>Ex. 1006, Hashima at 24:13-27: “The above embodiment employs a general-purpose image processor as shown in FIG. 51(A). According to commands given over a bus, one frame of scanned image data from a camera is stored in a frame memory 301, and a window is established and stored in a frame memory 302. Then, the image data is converted into binary image data which is stored in a frame memory 303. Finally, projected histograms are generated from the binary image data and stored in a histogram memory 304. The data stored in the histogram memory 304 is read by a CPU for detecting positional shifts.</p> <p>When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.”</p> <p>Ex. 1006, Hashima at 25:16-23: “The above window comparison, storage, conversion to the binary signal, and generation of projected histograms are carried out during</p>

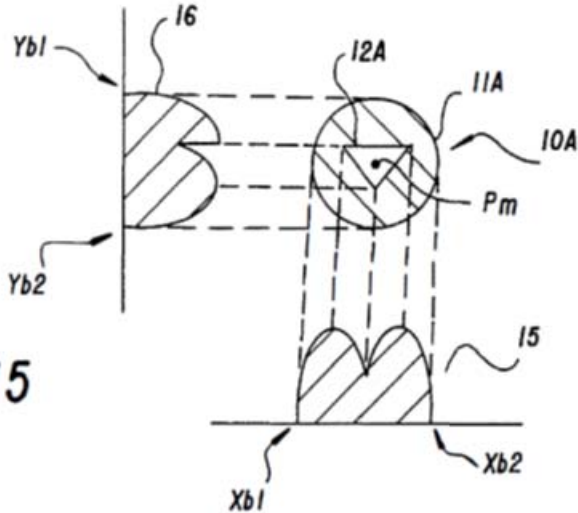
Claim 1	Exemplary Disclosures
	<p>horizontal scans, and the histogram values are stored in the memories 317, 317 during vertical blanking times. Therefore, the window comparison, storage, conversion to the binary signal, and generation of projected histograms, and the storage of the histogram values are completed until a next frame of image data starts to be scanned.</p>
<p>[1.a] forming at least one histogram of the pixels in the one or more of a plurality of classes in the one or more of a plurality of domains, said at least one histogram referring to classes defining said target,</p>	<p>Ex. 1005, Gilbert at 48: “Every 96 ns, a pixel intensity is digitized and quantized into eight bits (256 gray levels), counted into one of six 256-level histogram memories, and then converted by a decision memory to a 2-bit code indicating its classification (target, plume, or background). There are many features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures. Throughout the following discussion of the clustering algorithm, pixel intensity is used to describe the pixel features chosen. . . .</p> <p>The tracking window frame is partitioned into a background region (BR) and a plume region (PR). The region inside the frame is called the target region (TR) as shown in Fig. 2. During each field, the feature histograms are accumulated for the three regions of each tracking window.</p> <p>The feature histogram of a region R is an integer-value, integer argument function $h^R(x)$. The domain of $h^R(x)$ is $[0, d]$, where d corresponds to the dynamic range of the analog-to-digital converter, and the range of $h^R(x)$ is $[0, r]$, where r is the number of pixels contained in the region R; thus, there are $r + 1$ possible values of $h^R(x)$. Since the domain $h^R(x)$ is a subset of the integers, it is convenient to define $h^R(x)$ as a one-dimensional array of integers</p> <p>$h(0), h(1), h(2), \dots, h(d)$.</p> <p>Letting x_i denote the ith element in the domain of x (e.g., $x_{25} = 24$), and $x(j)$ denote the jth sample in the region R (taken in any order), $h^R(x)$ may be generated by the sum</p>

Claim 1	Exemplary Disclosures
	$h^R(x_i) = \sum_{j=1}^r \delta_{x_i, x(j)} \quad ,,$ <p>Ex. 1005, Gilbert at 49: “As each pixel in the region is processed, one (and only one) element of H is incremented as</p> $h[x(j)] \leftarrow h[x(j)] + 1.$ <p>When the entire- region has been scanned, h contains the distributions of pixels over intensity and is referred to as the feature histogram of the region R.”</p> <p>Ex. 1006, Hashima at 8:18-30: “FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. As shown in FIG. 6, the extreme values (maximum or minimum values which may be called "peaks" or "valleys") of each of the X- and Y-projected histograms comprise two peaks and one valley. The number of these extreme values remain unchanged even when the target mark rotates. The X-projected histogram represents the sum of pixels having the same X coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.”</p> <p>Ex. 1006, Hashima at 24:13-27: “The above embodiment employs a general-purpose image processor as shown in FIG. 51(A). According to commands given over a bus, one frame of scanned image data from a camera is stored in a frame memory 301, and a window is established and stored in a frame memory 302. Then, the image data is converted into binary image data which is stored in a frame memory 303. Finally, projected histograms are generated from the binary image data and stored in a histogram memory 304. The data stored in the histogram memory 304 is read by a</p>

Claim 1	Exemplary Disclosures
	<p>CPU for detecting positional shifts.</p> <p>When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.”</p>
<p>[1.b] and identifying the target in said at least one histogram itself,</p>	<p>Ex. 1005, Gilbert at 50: “The video processor described above separates the target image from the background and generates a binary picture, where target presence is represented by a "1" and target absence by a "0." The target location, orientation, and structure are characterized by the pattern of 1 entries in the binary picture matrix, and the target activity is characterized by a sequence of picture matrices. In the projection processor, these matrices are analyzed field-by-field at 60 fields/s using projection-based classification algorithms to extract the structural and activity parameters needed to identify and track the target. . . .</p> <p>In general, a projection integrates the intensity levels of a picture along parallel lines through the pattern, generating a function called the projection. . . .”</p> <p>Ex. 1005, Gilbert at 50: “To precisely determine the target position and orientation, the target center-of-area points are computed for the top section (X_c^T, Y_c^T) and bottom section (X_c^B, Y_c^B) of the tracking parallelogram using the projections. Having these points, the target center-of-area (X_c, Y_c) and its orientation can be easily computed (Fig. 4):</p> $X_c = \frac{X_c^T + X_c^B}{2}$ $Y_c = \frac{Y_c^T + Y_c^B}{2}$ $\phi = \tan^{-1} \frac{Y_c^T - Y_c^B}{X_c^T - X_c^B}.$

Claim 1	Exemplary Disclosures
	<p>Ex. 1005, Gilbert at Figure 4:</p>  <p>Fig. 4. Projection location technique.</p> <p>Ex. 1006, Hashima at 8:18-30: “FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. As shown in FIG. 6, the extreme values (maximum or minimum values which may be called "peaks" or "valleys") of each of the X- and Y-projected histograms comprise two peaks and one valley. The number of these extreme values remain unchanged even when the target mark rotates. The X-projected histogram represents the sum of pixels having the same X coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.”</p> <p>Ex. 1006, Hashima at 10:18-26: “Various modifications</p>

Claim 1	Exemplary Disclosures
	<p>are possible in the detection (extraction) of a target mark. The target mark to be detected is not limited to the those in the above embodiments. Target marks of various configurations which are characterized by the extreme values of X- and Y-projected histograms can be detected by the above process. The feature extraction using extreme values may be based on either a combination of maximum and minimum values or only one of maximum and minimum values.”</p> <p>Ex. 1006, Hashima at Figure 5:</p>  <p style="text-align: center;">FIG.5</p>
<p>[1.c] wherein identifying the target in said at least one histogram further comprises determining a center of the target to be</p>	<p>Ex. 1006, Hashima at 11:1–25: “A process of determining the central position P_m (m_x, m_y) of the target mark image 10A will be described below. FIG. 15 is illustrative of the manner in which the central position P_m of the target mark is determined. In FIG. 15, the central position P_m (m_x, m_y) of the target mark 10A corresponds to the center of a black circle image 11A. The image processor 40 integrates pixel</p>

Claim 1	Exemplary Disclosures
<p>between X and Y minima and maxima of the target.</p>	<p>values "0", "1" of the target mark image 10A that has been converted into a binary image, thereby generating a histogram 15 projected onto the X-axis and a histogram 16 projected onto the Y-axis, and then determines the X and Y coordinates of the central position P_m (mx, my) from the projected histograms 15, 16. Specifically, the X coordinate mx of the central position P_m (mx, my) can be determined using opposite end positions X_{b1}, X_{b2} obtained from the X-projected histogram 15 according to the following equation (3):</p> $mx=(X_{b1} +X_{b2})/2 \quad (3)$ <p>The Y coordinate my of the central position P_m (mx, my) can be determined using opposite end positions Y_{b1}, Y_{b2} obtained from the Y-projected histogram 16 according to the following equation (4):</p> $my=(Y_{b1}+Y_{b2})/2 \quad (4)''$ <p>Ex. 1006, Hashima at Figure 15:</p>  <p>FIG.15</p> <p>Ex. 1005, Gilbert at 50: “To precisely determine the target position and orientation, the target center-of-area points are computed for the top section (X_c^T, Y_c^T) and bottom section (X_c^B, Y_c^B) of the tracking parallelogram using the</p>

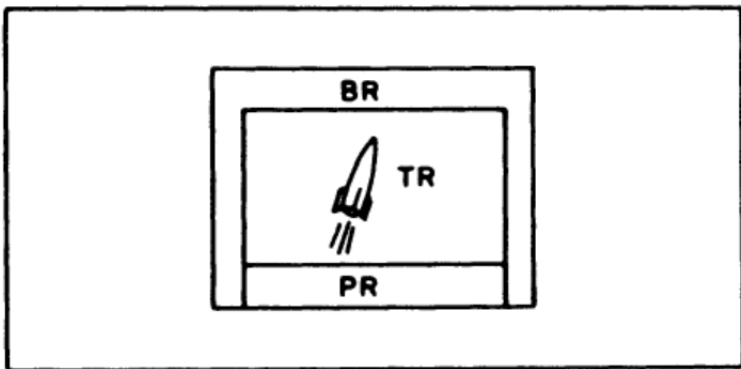
<p>Claim 1</p>	<p>Exemplary Disclosures</p>
	<p>projections. Having these points, the target center-of-area (X_c, Y_c) and its orientation can be easily computed (Fig. 4):</p> $X_c = \frac{X_c^T + X_c^B}{2}$ $Y_c = \frac{Y_c^T + Y_c^B}{2}$ $\phi = \tan^{-1} \frac{Y_c^T - Y_c^B}{X_c^T - X_c^B}.$ <p>The top and bottom target center-of-area points are used, rather than the target nose and tail points, since they are much easier to locate, and more importantly, they are less sensitive to noise perturbations.”</p> <p>Ex. 1005, Gilbert at Figure 4:</p> <p>Fig. 4. Projection location technique.</p>

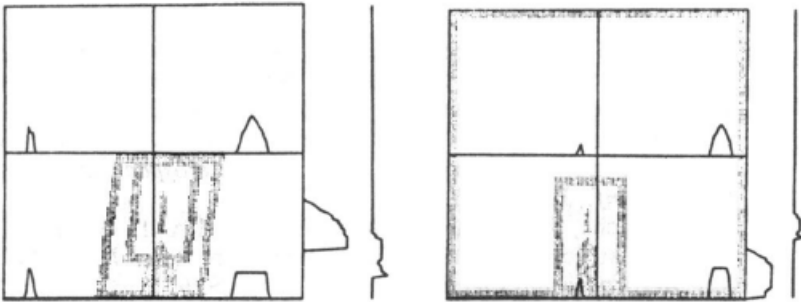
Claim 2	Exemplary Disclosures
<p>[2] The process according to claim 1, wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target.</p>	<p><i>See, e.g.</i>, disclosures in Claim [1.c].</p>

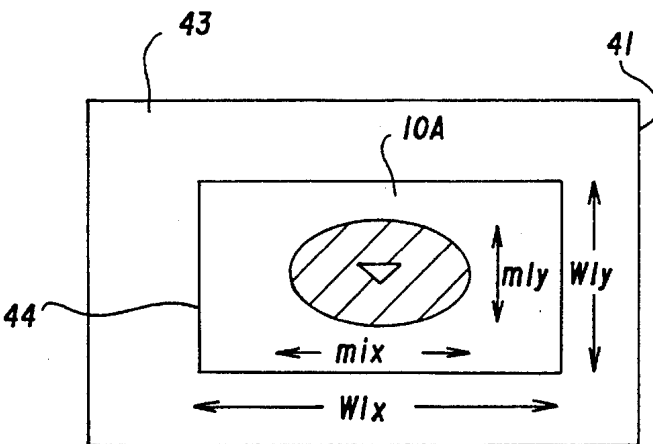
Claim 3	Exemplary Disclosures
<p>[3] The process according to claim 1, wherein identifying the target in said at least one histogram further comprises determining the center of the target at regular intervals.</p>	<p>Gilbert at 47: “The camera output is statistically decomposed into background, foreground, target, and plume regions by the video processor, with this operation carried on at video rates for up to the full frame. The projection processor then analyzes the structure of the target regions to verify that the object selected as "target" meets the stored (adaptive) description of the object being tracked.”</p> <p>Gilbert at 50: “The video processor described above separates the target image from the background and generates a binary picture, where target presence is represented by a "1" and target absence by a "0." The target location, orientation, and structure are characterized by the pattern of 1 entries in the binary picture matrix, and the target activity is characterized by a sequence of picture matrices. In the projection processor, these matrices are analyzed field-by-field at 60 fields/s using projection-based classification algorithms to extract the structural and activity parameters needed to identify and track the target.”</p> <p>Gilbert at 55: “The result after six processing frames, shown in Fig. 8(b), verifies the effectiveness of the processing algorithms used in the RTV tracker.”</p>

Claim 3	Exemplary Disclosures
	<p>Hashima at 24:13-31: “The above embodiment employs a general-purpose image processor as shown in FIG. 51(A). According to commands 15 given over a bus, one frame of scanned image data from a camera is stored in a frame memory 301, and a window is established and stored in a frame memory 302. Then, the image data is converted into binary image data which is stored in a frame memory 303. Finally, projected histograms 20 are generated from the binary image data and stored in a histogram memory 304. The data stored in the histogram memory 304 is read by a CPU for detecting positional shifts. When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of 25 image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated. The time required to write and read the data and the time required to access the bus total about 50 ms. Therefore, the general purpose image processor requires at least 200 ms to obtain projected histogram data of one frame of image data.”</p> <p>Hashima at 25:16-33: “The above window comparison, storage, conversion to the binary signal, and generation of projected histograms are carried out during horizontal scans, and the histogram values are stored in the memories 317, 317 during vertical blanking times. Therefore, the window comparison, storage, conversion to the binary signal, and generation of projected histograms, and the storage of the histogram values are completed until a next frame of image data starts to be scanned.</p> <p>As described above, the general-purpose image processor repeats accessing frame memories which are not necessary for the tracking algorithm and requires a processing time of at least 200 ms to determine projected histograms once. The image processor according to this embodiment requires not frame memory, and hence does not access any frame memory, with the result that its processing time</p>

Claim 3	Exemplary Disclosures
	corresponds only one frame (33 ms). The image processor can thus achieve a high-speed processing and is capable of tracking a target mark that moves at a high speed.”

Claim 4	Exemplary Disclosures
[4] The process according to claim 1 further comprising drawing a tracking box around the target.	<p>Gilbert at 48: “The basic assumption of the clustering algorithm is that the target image has some video intensities not contained in the immediate background. A tracking window is placed about the target image, as shown in Fig. 2, to sample the background intensities immediately adjacent to the target image. . . .</p> <p>The tracking window frame is partitioned into a background region (BR) and a plume region (PR). The region inside the frame is called the target region (TR) as shown in Fig. 2. During each field, the feature histograms are accumulated for the three regions of each tracking window.”</p> <p>Gilbert at Figure 2:</p>  <p style="text-align: center;">Fig. 2. Tracking window.</p> <p>Gilbert at 55: “The simulation displays the target and plume points as they are identified by the video processor. These points are superimposed on a cross hair which locates the boresight of the tracking optics. In addition to the digitized images, the simulation displays the target and plume tracking windows and the projections accumulated for each video field. These outputs provide a direct</p>

Claim 4	Exemplary Disclosures
	<p>measure of the performance of the four processors as well as an overall measure of the tracking accuracy of the system. . . .</p> <p>Fig. 7 contains two representative simulation outputs selected from the first 100 fields of simulated digitized video.”</p> <p>Gilbert at Figure 7:</p>  <p style="text-align: center;">Fig. 7. Simulation outputs.</p> <p>Hashima 13:64-14:34: “FIG. 23 showing the manner in which a window is established. In FIG. 23, since the camera 20 and the target mark 10 are not positioned perpendicularly to each other, the target mark image 10A is seen flat. The target mark image 10A is read into the image memory 41 of the image processor 40, and recognized, and a window 44 is established around the recognized target mark image 10A.</p> <p>The position of the center of gravity of the target mark image 10A which is obtained by the projected histogram processing when the target mark image 10A is recognized is regarded as the center of the window 44. Sizes $W1x$, $W1y$ of the window 44 are obtained by multiplying vertical and horizontal sizes $m1x$, $m1y$ of the target mark image 10A which are obtained by the projected histogram processing when the target mark image 10A is recognized, by a predetermined ratio, e.g., 2. In order to keep the region of the window 44 reliably in the image memory 41, a white</p>

Claim 4	Exemplary Disclosures
	<p>region that is slightly greater than the product of the vertical and horizontal sizes of the target mark 10 and the predetermined ratio is provided in advance around the black circle 11 of the target mark 10 on the object 1. In this manner, a white region 43 can always be kept in the image memory 41. Even if the distance between the target mark 10 and the camera 20 varies resulting in a change in the size of the target mark image 10A as read into the image memory 41, the target mark image 10A does not exceed the white region 43, making it possible to establish the window 44 around the target mark image 10A.</p> <p>By keeping the white region 43 around the target mark image 10A and setting the window 44 in the white region 43, it is possible to remove a black region such as noise other than the target mark image 10A.</p> <p>When the target mark 10 starts to be tracked, the window is established using the projected histogram information obtained when the target mark image is recognized. When the target mark 10 is subsequently tracked, the window is established using new projected histogram information obtained upon each measurement made by the camera 20.”</p> <p>Hashima at Figure 23:</p> <p style="text-align: center;">FIG.23</p> 

Claim 4	Exemplary Disclosures
	<i>See also</i> , Hashima at 8:55-57, 13:61-14:34, 24:44-25:23, Figure 53.

C. Ground 2: Hashima And Ueno Teaches Or Suggests Every Step And Feature Of Claims 1-4

116. In my opinion, as shown in the charts below, Hashima in view of Ueno and in view of the knowledge of a POSA, teaches or suggests every feature recited in Claims 1-4 of the '001 Patent.

1. Reasons To Combine Hashima And Ueno

117. A POSA would have been motivated to combine Hashima and Ueno because they are directed toward very similar systems that operate in a similar manner. As set forth above, both Hashima and Ueno are “target recognition systems” directed to identifying and tracking a target using an image processing system with a video camera input (Ex. 1006, Hashima at 1:6-15, 2:45-3:12, Figure 1; Ex. 1007, Ueno at 3:1-8, 4:25-41), where the video input comprises a succession of frames, each frame comprising a succession of pixels (Ex. 1006, Hashima at 10:34-45, 24:13-25:33; Ex. 1007, Ueno at 1:44-49, 4:25-41), in which the target comprises pixels in one or more of a plurality of classes and one or more of a plurality of domains (Ex. 1006, Hashima at 8:18-30, 10:34-45; Ex. 1007, Ueno at 7:8-16), and performing the process on a frame-by-frame basis (Ex. 1006, Hashima at 24:13-25:23; Ex. 1007, Ueno at 4:25-41, 5:31-40).

118. Both Hashima and Ueno identify and track a moving target using histograms (Ex. 1006, Hashima at 4:8-47; Ex. 1007, Ueno at 7:7-8:22). As mentioned above, there are many types of machine vision systems for tracking based on a variety of image analysis methods such as decomposition using Fourier, wavelet or other bases, or feature detection yielding descriptors of edges, corners or other image features. Hashima and Ueno both relate to the same, small, cross-section of these systems based on histograms, and specifically on projection histograms of classified and segmented image pixels.

119. While Ueno uses the application of facial detection, it does so by designing a general target recognition system that can be trained using faces but can also be trained to identify other targets. Ueno Fig. 6 is described as a “facial region detecting circuit” but uses projection histograms for the detection of “a specific region.” One of the top conferences in computer vision at the time was the IEEE Conference on Computer Vision and Pattern Recognition. In 1996, around the ‘001 priority date, there was a CVPR session devoted to “Facial Detection and Recognition,” but all of these papers focused on the detection of faces through features (corners, edges, and cusps at the boundaries of the eyes, lips, nose and face) or through direct comparisons of images rather than projection histograms as described by both Hashima and Ueno. Hence Ueno would have stood out as a unique approach when compared to mainstream facial detectors, and also had the

distinction of being a general target recognition system that was demonstrated on faces, and so would have served as an obvious choice to combine with Hashima to broaden the targets that histogram projection could detect to range from pre-designed and more rigidly varying fiducial images of, e.g., a disk surrounding a suspended triangle, to also include faces and other more naturally varying targets.

120. Thus, they are not only in the same field of endeavor of target recognition, but also are directed to very similar systems and processes for implementing projection histograms. A POSA would have recognized that Ueno's invention could improve a similar device, such as the system in Hashima, in the same way, and thus would have been motivated to combine Hashima and Ueno. A POSA would have also expected the combination to yield a predictable result because it would have involved applying known techniques to similar systems.

121. In particular, although Hashima describes a system that identifies and tracks a target and projects the tracked object on a display for viewing by a user, it does not teach any means by which the system highlights the target for easier visualization by a user. However, being able to visualize a target more easily on a screen provides numerous advantages. For example, a tracking box enables the user to visually confirm that the correct object has been identified and is being tracked by the image processing system, in the context of the contents of the entire video frame. Additionally, the user may use a mouse or keyboard input to specify

a target mark to be tracked. A tracking box allows the user to more easily pick out the target object in a situation where noise or other objects in the field of view otherwise make it more difficult to visually identify the target. For these reasons, a POSA would be motivated to combine Hashima with Ueno, which teaches a tracking box and allows a user to select a target using an input to the image processing system. Moreover, there would have been a reasonable expectation of success that in doing so a more user friendly system would have resulted.

2. Claim 1

- a) **[1 pre]: A process of tracking a target," in an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels, the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains, the process performed by said system comprising, on a frame-by-frame basis**

122. The preamble in Claim 1 requires four different elements: (1) that the process “track[] a target,” (2) that a specific input signal comprises a sequence of frames where each frame comprises multiple pixels, (3) that the pixels must be able to be categorized according to their characteristics into at least one domain and one class, selected from a plurality of domains and classes, and (4) that the method must repeat itself “on a frame-by-frame basis.” Both Hashima and Ueno disclose each of these steps as described below, and as more fully laid out in the claim chart in **IX.C.6**.

123. Hashima discloses “[a] process of tracking a target.” As stated in the abstract and summary of the invention, Hashima is directed toward “[a] system for and a method of recognizing and tracking a target mark with a video camera.” Ex. 1006, Hashima at 1:7-9. Although Ueno does not specifically call out the its tracking functionality, it does teach a terminal “for inputting movement image signals” that is connected to a frame memory. Ex. 1007, Ueno at 3:1-3. Further it teaches that the terminal is connected to a “facial region detecting circuit . . . for detecting the human facial region of the frame data.” *Id.* at 3:6-8. Taken together, these and similar disclosures clearly teach a system and method for identifying and tracking a face across multiple frames of a video signal.

124. Hashima also discloses “an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels.” This limitation is met in its entirety simply because the input described is a video signal acquired by a TV camera. Ex. 1006, Hashima at 1:7-15. Any such video input inherently comprises a sequence of frames and each frame in a video input would also comprise many pixels. Nevertheless, Hashima also expressly teaches that “one frame of scanned image data from a camera is stored in a frame memory,” and that “[a]n image . . . is read into an image memory . . . [which] is composed of 512 pixels in each of the X- and Y-directions.” *Id.* at 24:14-17; 10:37-44.

125. Similarly, Ueno teaches that video signals are sequentially input to the image input terminal . . . in frames.” Ex. 1007, Ueno at 4:25-26. Moreover, Ueno discloses that each frame is comprised of many pixels, as a POSA would expect in any video signal. *See id.* at 4:39-41, 7:8-16. Thus, both Hashima and Ueno disclose the input signal required in Claim 1.

126. Hashima meets the limitation “the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains.” The pixels described by Hashima can be categorized into classes and domains. This limitation is very similar to the step in [1 a] below, inasmuch as domains merely refer to pixel characteristics for use in a histogram and classes refer to the discrete divisions, i.e., bins, within the domain into which the pixels can be classified. *See* Ex. 1001, ’001 Patent at 4:2-6 (listing specific pixel characteristics for use as domains); 6:1-3 (the pixels are classified “within each domain to selected classes within the domain”). Hashima specifically teaches that the pixels in the input image in Hashima can be categorized into classes and domains. Although many of the same pixel characteristics as those listed in the ’001 Patent must be present in Hashima as intrinsic pixel qualities, Hashima expressly teaches forming histograms in the X- and Y-domains: “The X-projected histogram represents the sum of pixels having the same X-coordinates, and the Y-projected histogram represents the sum of pixels having the same Y-coordinates.” Ex. 1006, Hashima at 8:27-31. For these

X- and Y-histograms, the “classes” are the coordinates themselves or the coordinate ranges into which the pixels are quantized along the X- and Y-axis. The ’001 Patent specifically teaches that the X- and Y-axis are among the preferred possible domains. *See* Ex. 1001, ’001 Patent at 4:2-6.

127. Like Hashima, Ueno teaches constructing histograms in the X- and Y-domains. *See* Ex. 1007, Ueno at 7:8-16. Moreover, Ueno also teaches that the video input signal is split into a number of different constituent parts. Specifically, Ueno teaches that: “The video signal is, for example, obtained by encoding analog image signals obtained through picking-up of a human with a TV camera (not illustrated) into luminance signal Y- and two types of color signals (or color-difference signals).” *Id.* at 4:26-30. These image signals correspond closely to three of the described domains in the ’001 Patent, specifically hue and saturation (two color signals) and luminance (the graylevel signal). Thus, both Hashima and Ueno disclose pixels with a plurality of domains and a plurality of classes.

128. Hashima also meets the limitation that “the process [be] performed by said system . . . on a frame-by-frame basis.” Specifically, Hashima teaches that “it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera.” Ex. 1006, Hashima at 24:24-25. Ueno also clearly teaches that “[t]he image signal is written in the frame memory as frame data one frame by one frame. The frame data written in the frame memory . . . can also be read out every

frame.” Moreover, Ueno specifically teaches creating histograms based on the differences between consecutive frames. *Id.* at 7:8-16. In order to accomplish creating interframe difference images, Ueno necessarily processes the image signal on a frame-by-frame basis. Thus, both Hashima and Ueno teach that the image tracking and identification process is performed on a frame-by-frame basis as claimed in the ’001 Patent.

- b) **[1 a]: forming at least one histogram of the pixels in the one or more of a plurality of classes in the one or more of a plurality of domains, said at least one histogram referring to classes defining said target**

129. Both Hashima and Ueno disclose this step, as discussed below and as further shown in the claim chart at **IX.C.6**. In each case, at least one histogram is formed based on pixel characteristics (domains) such as intensity or coordinates X and Y. Examples of domains are seen in the ’001 Patent: “The domains are preferably selected from the group consisting of i) luminance, ii) speed (V), iii) oriented direction (D1), iv) time constant (CO), v) hue, vi) saturation, and vii) first axis (x(m)), and viii) second axis (y(m)).” Ex. 1001, ’001 Patent at 4:2-6. Hashima expressly teaches forming histograms using the X- and Y-axis domains: “X- and Y-projected histograms of the target mark image are determined. . . . The X-projected histogram represents the sum of pixels having the same X-coordinates, and the Y-projected histogram represents the sum of pixels having the same Y-coordinates.” Ex. 1006, Hashima at 8:27-30. The X- and Y-coordinates of pixels

serve as two of the domains expressly given as examples in the '001 Patent. Ex. 1001, '001 Patent at 4:2-6. In these histograms, each "class" within a domain counts the number of pixels meeting certain threshold criteria. For example, in both Hashima and Ueno, the number of pixels having a value of "1" within a given column defined by an X-coordinate is counted into the class that corresponds to the X-coordinate value.

130. In Ueno, a similar process occurs, but the histograms are generated on interframe images and specifically identify pixels that are different in the two frames. Like in Hashima, histograms are generated in the domains of the X- and Y-axis:

FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.

Ex. 1007, Ueno at 7:7-16. In Ueno, as in Hashima, target pixels are categorized according to classes within the chosen domain, for example position along a relevant axis for X- and Y-axis histograms. Thus, both Hashima and Ueno teach that histograms are formed in at least one of a plurality of domains with each pixel being categorized into one or more classes within said domain.

131. Moreover, these X- and Y- projection histograms in Hashima and Ueno are analogous to the X- and Y- position histograms discussed in the '001 Patent at 20:60-21:29.

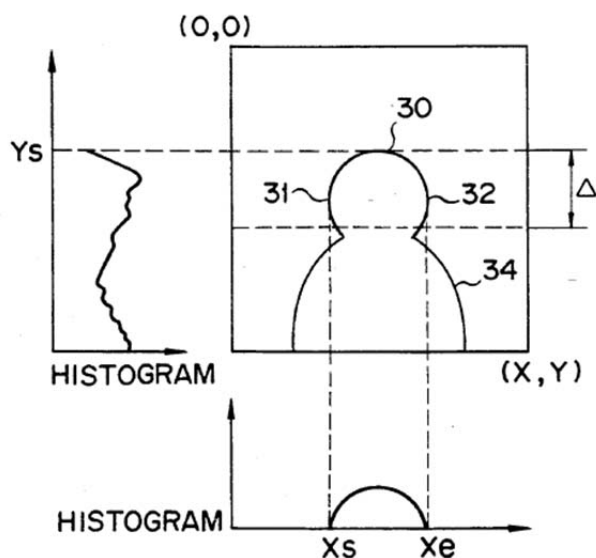
c) [1 b]: identifying the target in said at least one histogram itself

132. Both Hashima and Ueno disclose this step, as discussed below and shown in greater detail in the claim chart at **IX.C.6**. In each case, histograms are used to identify the target. For example, in Hashima, peaks and valleys in the X- and Y-axis histograms are used to determine the location of the target: “Target marks of various configurations which are characterized by the extreme values of the X- and Y-projected histograms can be detected by the above process.” Ex. 1006, Hashima at 10:20-24. In other words, the target mark is identified in Hashima directly from the X- and Y-axis histograms.

133. Similarly, in Ueno, interframe differences are formed into X- and Y- histograms and used to identify a human face:

FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of “0” or “1” by the preset first threshold level. Then the number of pixels having the binary data value of “1” or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.

Ex. 1007, Ueno at 7:7-16. Figure 3 shows how the target is identified directly from the X- and Y-histograms:



F I G. 3

134. These methods of identifying a face using histograms are nearly identical to the way that the '001 Patent describes identifying a face from the histograms in Figure 17 and the accompanying description at 22:60-23:14. Moreover, the binary selection process identified in both Hashima and Ueno are entirely analogous to the histogram formation process discussed in the '001 Patent at 18:16-19:24.

- d) [1 c]: wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target**

135. Hashima discloses the step of determining a center of the target, as described below and shown in more detail in the claim chart at **IX.C.6**. Hashima discloses a method of finding a center point of the target image and explains how

the X- and Y-minima and maxima are used to calculate the center of the target.

Specifically, Hashima teaches:

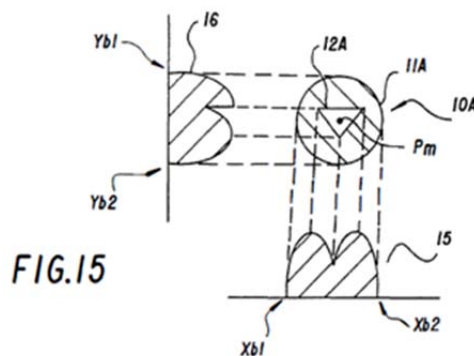
[T]he X coordinate m_x of the central position P_m (m_x , m_y) can be determined using opposite end positions X_{b1} , X_{b2} obtained from the X-projected histogram 15 according to the following equation (3):

$$m_x = (X_{b1} + X_{b2}) / 2 \quad (3).$$

The Y coordinate m_y of the central position P_m (m_x , m_y) can be determined using opposite end positions Y_{b1} , Y_{b2} obtained from the Y-projected histogram 16 according to the following equation (4):

$$m_y = (Y_{b1} + Y_{b2}) / 2 \quad (4).$$

Ex. 1006, Hashima at 11:13-24. Figure 15 shows that the “opposite end positions” referred to in the quotation above are the X- and Y-minima and maxima from the X- and Y-histograms and illustrates how they are used to compute the center of the target:



The equations used to find the center in Hashima are identical to those described for the same purpose in the '001 Patent at 24:50-55. In addition, although Ueno specifically discloses using the X-axis histogram to locate the left and right end of the target (i.e., X- minima and maxima), and using the Y-axis histogram to locate the top and bottom of the target (i.e., Y-minima and maxima), it does not explicitly disclose finding the center point of the target from these maxima and minima. Of course calculating a center point is simply a mathematical identity based on determining the X- and Y-minima and maxima, so this limitation is inherently disclosed by Ueno or at a minimum would have been obvious in view of the knowledge of a POSA.

3. Claim 2

- a) **The process according to Claim 1, wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target**

136. Claim 2 contains a limitation identical to the final limitation in Claim 1. Thus, it is my opinion that Hashima and Ueno both disclose this

limitation for the same reasons discussed above. *See* Section **IX.C.2.d**.

4. Claim 3

- a) **The process according to Claim 1, wherein identifying the target in said at least one histogram further comprises determining the center of the target at regular intervals**

137. Both Hashima and Ueno teach that the target identification process, including the center determination, take place at regular intervals. For example, Hashima teaches that “When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of 25 image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.” *See, e.g.*, Ex. 1006, Hashima at 24:13-35. Hashima discloses embodiments where this process requires “at least 200 ms,” *id.*, and other embodiments where the “processing time corresponds to only one frame (33 ms).” *Id.* at 25:16-33. Thus, because determining the center of the target is part of the tracking process in both embodiments, it is done at the regular interval of 200 ms or on a frame-by-frame basis every 33 ms. This frame-by-frame determination of the center of the target in Hashima is identical to the description in the ’001 Patent at 25:3-25 of determining the center position at regular intervals.

138. Ueno similarly discloses processing the camera input one frame at a time. *See e.g.*, Ex. 1007, Ueno at 1:45-46 (“[A] specific region in image signals

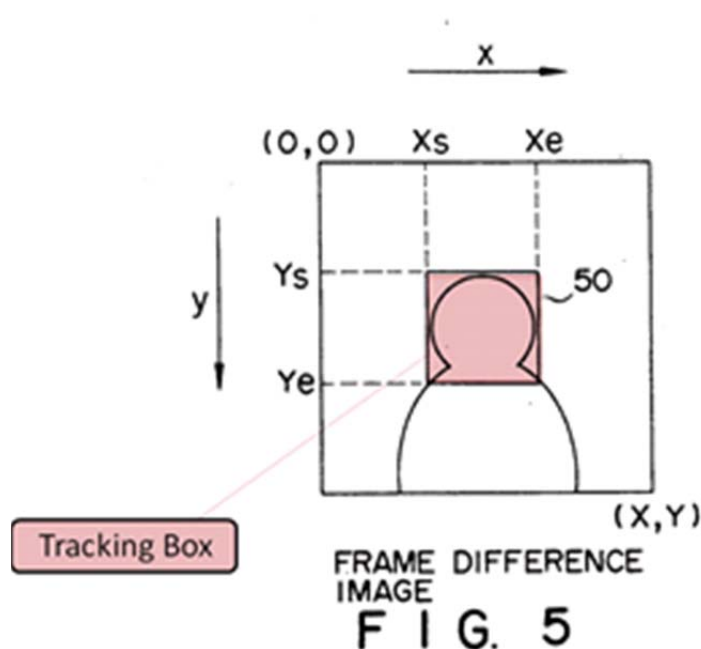
sequentially input for every frame is detected which corresponds to a movable image”); 4:25-41 (“Video signals are sequentially input to the image input terminal 100 in frames. . . . The image signal is written in the frame memory 101 as frame data one frame by one frame. . . . The data can also be read out every frame.”). Because determining the X- and Y-minima and maxima is part of target location and identification process, it necessarily takes place on the regular interval of each frame.

139. To the extent the Patent Owner argues that Ueno and Hashima do not explicitly disclose that the center of the target is determined at regular intervals, a POSA would have considered implementing this feature in order to achieve smooth, predictable processing.

5. Claim 4

a) The process according to Claim 1 further comprising drawing a tracking box around the target

140. Hashima discloses this step as discussed in above paragraph 114, in connection with Ground 1, Claim 4. Ueno also discloses the step of drawing a tracking box around the target. *See, e.g.*, Ex. 1007, Ueno at 7:43-45 (“The facial region is specified by a rectangle 50 designated by coordinates Xs, Xe, Ys, and Ye as shown in FIG. 5.”). This tracking box is shown in Ex. 1007, Ueno at Figure 5, annotated below (the solid black line around the box is in the original figure):



141. Ueno also clarifies that the tracking box is displayed on a monitor or display. For example, Ueno teaches that the region of interest is displayed with a frame that is superimposed on the original image and displayed on the monitor:

In the self-monitoring mode, low-pass-filtered images are displayed on the monitoring display 710. After a region is specified by the region specifying apparatus 120, the region specifying signal is continuously output. Thus, said frame line showing the specified region is superimposed on the image and displayed on the monitoring display 710.

Ex. 1007, Ueno at 13:17-23. *See also* Ex. 1007, Ueno at 12:6-25; Figure 18.

142. Thus, Ueno discloses this limitation. To the extent Ueno does not explicitly disclose drawing a tracking box around the target, a POSA would have considered implementing this feature in order to aid a user at a display to visually locate and follow a target.

6. Detailed Application Of Hashima And Ueno To Claims 1-4

143. The charts below provide a detailed mapping of Hashima and Ueno to Claims 1-4 of the '001 Patent and show that Hashima in view of Ueno teach or suggest every feature recited in Claims 1-4:

Claim 1	Exemplary Disclosures
[1.pre] A process of tracking a target in an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels, the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains, the process performed by said system comprising, on a frame-by-frame basis	<p>Hashima at 1:6-16: “The present invention relates to a system for and a method of recognizing and tracking a target mark using a video camera, and more particularly to a system for and a method of recognizing and tracking a target mark for detecting the position and attitude of the target mark by processing an image of the target mark produced by a video camera, detecting a shift of the position of the target mark from a predetermined position, and controlling the position and attitude of a processing mechanism based on the detected shift.”</p> <p>Hashima at 8:18-30: “FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. As shown in FIG. 6, the extreme values (maximum or minimum values which may be called "peaks" or "valleys") of each of the X- and Y-projected histograms comprise two peaks and one valley. The number of these extreme values remain unchanged even when the target mark rotates. The X-projected histogram represents the sum of pixels having the same X</p>

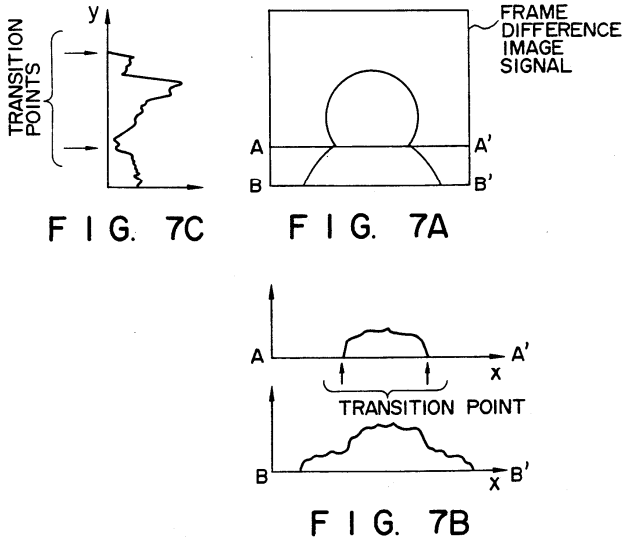
Claim 1	Exemplary Disclosures
	<p>coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.”</p> <p>Hashima at 10:34-45: “FIG. 14 is illustrative of the manner in which shifts in X- and Y-directions of a target mark are determined. In FIG. 14, the target mark 10 is imaged as being shifted from the center of the screen of the camera 20. An image (target mark image) 10A of the target mark 10 which is imaged by the camera 20 is read into an image memory 41 of the image processor 40, and recognized in an image coordinate system 42. The target mark image 10A is composed of a black circular image 11A and a white triangular image 12A. The image memory 41 is composed of 512 pixels in each of X and Y directions. The coordinate position of each point in the image memory is indicated according to the unit of pixels.”</p> <p>Hashima at 14:29-34: “When the target mark 10 starts to be tracked, the window is established using the projected histogram information obtained when the target mark image is recognized. When the target mark 10 is subsequently tracked, the window is established using new projected histogram information obtained upon each measurement made by the camera 20.”</p> <p>Hashima at 24:13-27: “The above embodiment employs a general-purpose image processor as shown in FIG. 51(A). According to commands given over a bus, one frame of scanned image data from a camera is stored in a frame memory 301, and a window is established and stored in a frame memory 302. Then, the image data is converted into binary image data which is stored in a frame memory 303. Finally, projected histograms are generated from the binary image data and stored in a histogram memory 304. The data stored in the histogram memory 304 is read by a CPU for detecting positional shifts.</p> <p>When the above general-purpose image processor is</p>

Claim 1	Exemplary Disclosures
	<p>employed, it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.”</p> <p>Hashima at 25:16-23: “The above window comparison, storage, conversion to the binary signal, and generation of projected histograms are carried out during horizontal scans, and the histogram values are stored in the memories 317, 317 during vertical blanking times. Therefore, the window comparison, storage, conversion to the binary signal, and generation of projected histograms, and the storage of the histogram values are completed until a next frame of image data starts to be scanned.”</p> <p>Ueno at 1:44-49: “According to the present invention, a specific region in image signals sequentially input for every frame is detected which corresponds to a movable image having the specific region and the region specifying signal to separate the specific region from other regions is output.”</p> <p>Ueno at 3:1-8: “In the embodiment shown in FIG. 1, a terminal 100 for inputting movement image signals is connected to a frame memory 101 for storing image signals for one frame (this is called frame data). The readout terminal of the frame memory 101 is connected to the input terminal of a facial region detecting circuit 102 for detecting the human facial region of the frame data and that of a low-pass filter 103 for filtering the frame data.”</p> <p>Ueno at 4:25-41 : “Video signals are sequentially input to the image input terminal 100 in frames. The video signal is, for example, obtained by encoding analog image signals obtained through picking-up of a human with a TV camera (not illustrated) into luminance signal Y and two types of color signals (or color-difference signals) CB and CR and converting these encoded signals into digital data called</p>

Claim 1	Exemplary Disclosures
	<p>CIF or QCIF by an A-D converter and format converter (not illustrated). The image signal is written in the frame memory 101 as frame data one frame by one frame. The frame data written in the frame memory 101 is read out every a plurality of blocks and supplied to the facial region detecting circuit 102 and low-pass filter 103. The data can also be read out every frame. The data readout unit is larger than the block size (8 pixels×8 pixels) which is the conversion unit (encoding unit) in the DCT circuit 107.”</p> <p>Ueno at 5:31-40: “The locally-decoded data obtained through inverse DCT is stored in the frame memory built in the motion compensated interframe prediction circuit 113 and referenced so that the next frame data will be encoded. That is, the motion compensated interframe prediction circuit 113 compares the locally-decoded data of the previous frame stored in the built-in frame memory with the current frame data sent from the low-pass filter 103 to generate motion vector information.”</p> <p>Ueno at 7:8-16: “Information for the interframe difference image is converted into binary data of "O" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.”</p>
<p>[1.a] forming at least one histogram of the pixels in the one or more of a plurality of classes in the one or more of a plurality of domains, said at least one histogram referring to classes defining said target,</p>	<p>Hashima at 8:18-30: “FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. As shown in FIG. 6, the extreme values (maximum or minimum values which may be called "peaks" or "valleys") of each of the X- and Y-projected histograms comprise two peaks and one valley. The number of these extreme values remain unchanged even</p>

Claim 1	Exemplary Disclosures
	<p>when the target mark rotates. The X-projected histogram represents the sum of pixels having the same X coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.”</p> <p>Hashima at 24:13-27: “The above embodiment employs a general-purpose image processor as shown in FIG. 51(A). According to commands given over a bus, one frame of scanned image data from a camera is stored in a frame memory 301, and a window is established and stored in a frame memory 302. Then, the image data is converted into binary image data which is stored in a frame memory 303. Finally, projected histograms are generated from the binary image data and stored in a histogram memory 304. The data stored in the histogram memory 304 is read by a CPU for detecting positional shifts.</p> <p>When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated.”</p> <p>Ueno at 7:7-16: “FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.”</p> <p>Ueno at Figure 3:</p>

Claim 1	Exemplary Disclosures
	<div data-bbox="604 273 1136 772" data-label="Figure"> </div> <p data-bbox="792 825 971 867" style="text-align: center;">FIG. 3</p> <p data-bbox="548 898 1429 1234">Ueno at 8:5-12: FIG. 6 shows a facial region detecting circuit using the principle in FIG. 3. In this facial region detecting circuit, the x-axis histogram shown in FIG. 7B and the histogram shown in FIG. 7C are formed by the frame difference image shown in FIG. 7A. By extracting the coordinates of the transition points of these histograms, the face region is specified by a rectangle as shown in FIG. 5."</p> <p data-bbox="548 1283 824 1325">Ueno at Figure 6:</p> <div data-bbox="560 1344 1323 1753" data-label="Diagram"> </div> <p data-bbox="933 1774 1047 1806" style="text-align: center;">FIG. 6</p>

Claim 1	Exemplary Disclosures
	<p>Ueno at Figure 7:</p>  <p>FIG. 7C FIG. 7A</p> <p>FIG. 7B</p>
<p>[1.b] and identifying the target in said at least one histogram itself,</p>	<p>Hashima at 8:18-30: “FIG. 5 shows a sequence for detecting the target mark. In this embodiment, it is assumed that a target mark image with a triangle located within a circle is detected from a projected image. X- and Y-projected histograms of the target mark image are determined. As shown in FIG. 6, the extreme values (maximum or minimum values which may be called "peaks" or "valleys") of each of the X- and Y-projected histograms comprise two peaks and one valley. The number of these extreme values remain unchanged even when the target mark rotates. The X-projected histogram represents the sum of pixels having the same X coordinates, and the Y-projected histogram represents the sum of pixels having the same Y coordinates.”</p> <p>Hashima at 10:18-26: “Various modifications are possible in the detection (extraction) of a target mark. The target mark to be detected is not limited to the those in the above embodiments. Target marks of various configurations which are characterized by the extreme values of X- and Y-projected histograms can be detected by the above process. The feature extraction using extreme values may be based on either a combination of maximum and minimum values or only one of maximum and minimum values.”</p>

Claim 1

Exemplary Disclosures

Hashima at Figure 5:

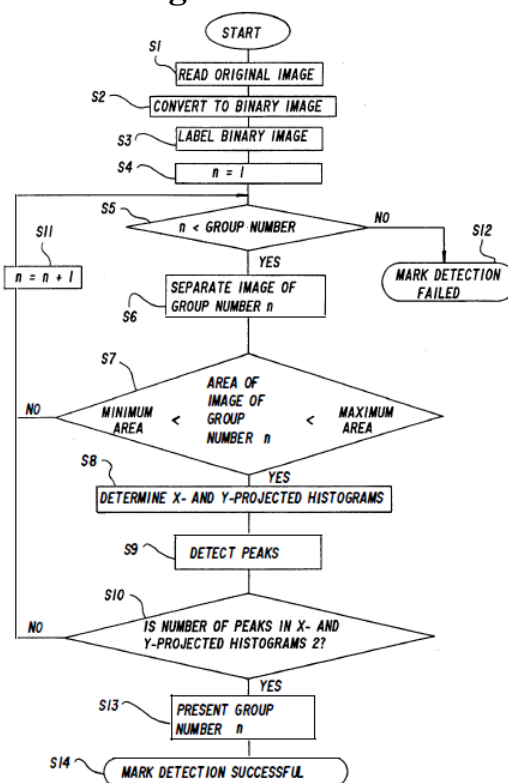


FIG.5

Ueno at 7:8-25: “FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.

Face detection starts with the top of a head 30. The top of the head 30 is detected by searching the y-axis histogram from the top and selecting the point Ys which first exceeds the present second threshold value.

Claim 1

Exemplary Disclosures

After the top of the head 30 is detected, the left end 31 and right end 32 of the head 30 are detected by searching the x-axis histogram from the left and right and selecting the points X_s and X_e which first exceed the second threshold value.”

Ueno at Figure 3:

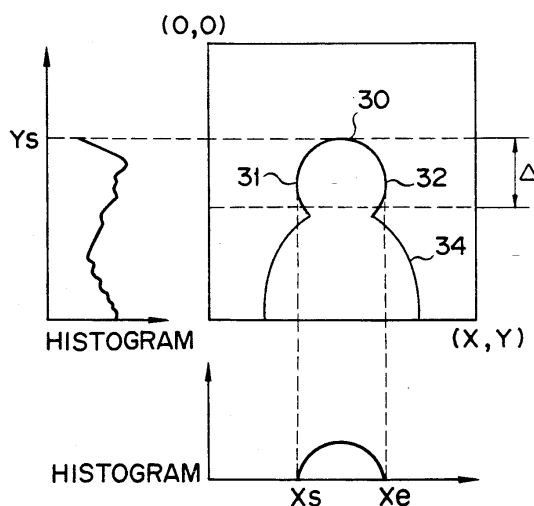
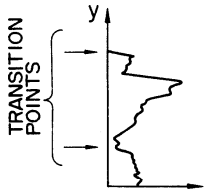
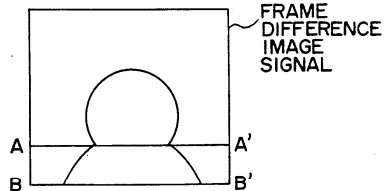
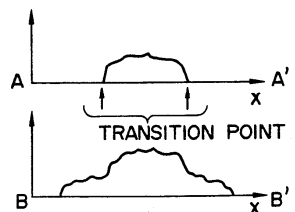
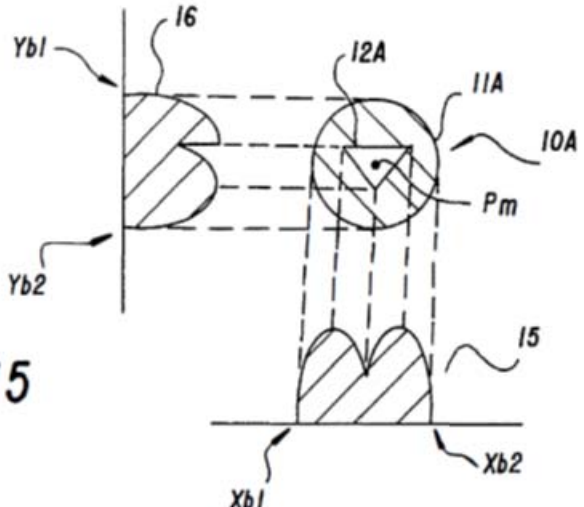


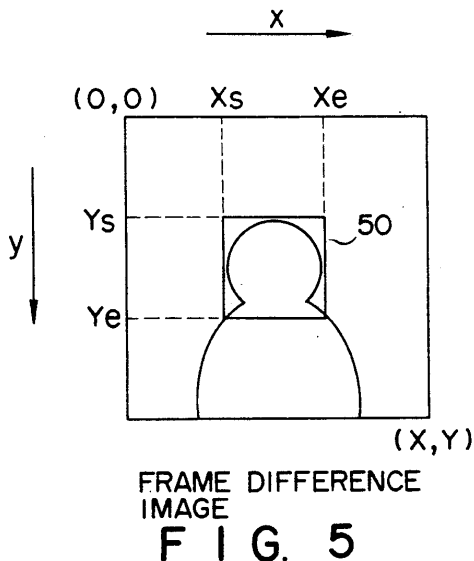
FIG. 3

Ueno at 8:5-12: FIG. 6 shows a facial region detecting circuit using the principle in FIG. 3. In this facial region detecting circuit, the x-axis histogram shown in FIG. 7B and the histogram shown in FIG. 7C are formed by the frame difference image shown in FIG. 7A. By extracting the coordinates of the transition points of these histograms, the face region is specified by a rectangle as shown in FIG. 5.

Ueno at Figure 7:

Claim 1	Exemplary Disclosures
	   <p style="text-align: center;">F I G. 7C F I G. 7A</p> <p style="text-align: center;">F I G. 7B</p>
<p>[1.c] wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target.</p>	<p>Hashima at 11:1–25: “A process of determining the central position Pm (mx, my) of the target mark image 10A will be described below. FIG. 15 is illustrative of the manner in which the central position Pm of the target mark is determined. In FIG. 15, the central position Pm (mx, my) of the target mark 10A corresponds to the center of a black circle image 11A. The image processor 40 integrates pixel values "0", "1" of the target mark image 10A that has been converted into a binary image, thereby generating a histogram 15 projected onto the X-axis and a histogram 16 projected onto the Y-axis, and then determines the X and Y coordinates of the central position Pm (mx, my) from the projected histograms 15, 16. Specifically, the X coordinate mx of the central position Pm (mx, my) can be determined using opposite end positions Xb1, Xb2 obtained from the X-projected histogram 15 according to the following equation (3):</p> $mx=(Xb1 +Xb2)/2 \quad (3)$ <p>The Y coordinate my of the central position Pm (mx, my) can be determined using opposite end positions Yb1, Yb2 obtained from the Y-projected histogram 16 according to the following equation (4):</p>

Claim 1	Exemplary Disclosures
	<p>my=(Yb1+Yb2)/2 (4)”</p> <p>Hashima at Figure 15:</p>  <p>Ueno at 7:7-45: “Facial detection is executed according to the histograms.</p> <p>Face detection starts with the top of a head 30. The top of the head 30 is detected by searching the y-axis histogram from the top and selecting the point Ys which first exceeds the present second threshold value.</p> <p>After the top of the head 30 is detected, the left end 31 and right end 32 of the head 30 are detected by searching the x-axis histogram from the left and right and selecting the points Xs and Xe which first exceed the second threshold value. . . .</p> <p>Then, the bottom end of the facial region is decided. . . . The facial region is specified by a rectangle 50 designated by coordinates Xs, Xe, Ys, and Ye as shown in FIG. 5.”</p> <p>Ueno at Figure 5:</p>

Claim 1	Exemplary Disclosures
	 <p style="text-align: center;">FRAME DIFFERENCE IMAGE FIG. 5</p>

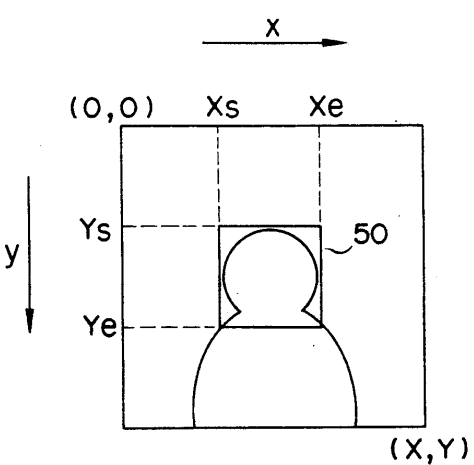
Claim 2	Exemplary Disclosures
[2] The process according to claim 1, wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target.	<i>See, e.g., disclosures in Claim [1.c].</i>


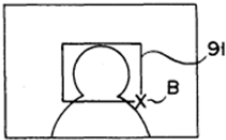
Claim 3	Exemplary Disclosures
[3] The process according to claim 1, wherein identifying the target in said at least one histogram further comprises	Hashima at 24:13-31: “The above embodiment employs a general-purpose image processor as shown in FIG. 51(A). According to commands 15 given over a bus, one frame of scanned image data from a camera is stored in a frame memory 301, and a window is established and stored in a frame memory 302. Then, the image data is converted into

Claim 3	Exemplary Disclosures
<p>determining the center of the target at regular intervals.</p>	<p>binary image data which is stored in a frame memory 303. Finally, projected histograms 20 are generated from the binary image data and stored in a histogram memory 304. The data stored in the histogram memory 304 is read by a CPU for detecting positional shifts. When the above general-purpose image processor is employed, it is necessary to write and rea[d] one frame of 25 image data each time it is supplied from the camera, a window is established, the data is converted into binary image data, and projected histograms are generated. The time required to write and read the data and the time required to access the bus total about 50 ms. Therefore, the general purpose image processor requires at least 200 ms to obtain projected histogram data of one frame of image data.”</p> <p>Hashima at 25:16-33: “The above window comparison, storage, conversion to the binary signal, and generation of projected histograms are carried out during horizontal scans, and the histogram values are stored in the memories 317, 317 during vertical blanking times. Therefore, the window comparison, storage, conversion to the binary signal, and generation of projected histograms, and the storage of the histogram values are completed until a next frame of image data starts to be scanned.</p> <p>As described above, the general-purpose image processor repeats accessing frame memories which are not necessary for the tracking algorithm and requires a processing time of at least 200 ms to determine projected histograms once. The image processor according to this embodiment requires not frame memory, and hence does not access any frame memory, with the result that its processing time corresponds only one frame (33 ms). The image processor can thus achieve a high-speed processing and is capable of tracking a target mark that moves at a high speed.”</p> <p>Ueno at 1:44-49: “According to the present invention, a specific region in image signals sequentially input for every</p>

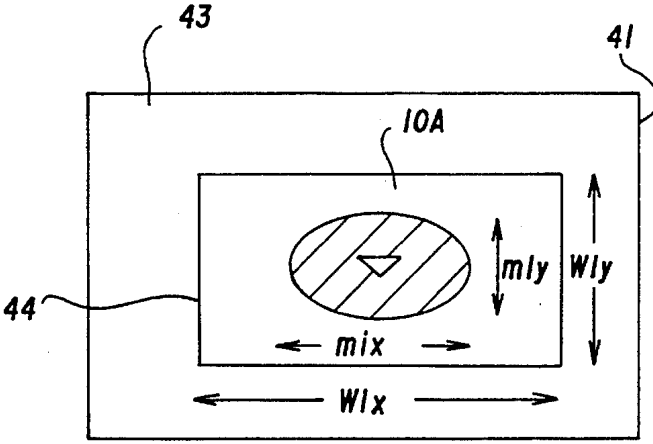
Claim 3	Exemplary Disclosures
	<p>frame is detected which corresponds to a movable image having the specific region and the region specifying signal to separate the specific region from other regions is output.”</p> <p>Ueno at 4:25-39: “Video signals are sequentially input to the image input terminal 100 in frames. The video signal is, for example, obtained by encoding analog image signals obtained through picking-up of a human with a TV camera (not illustrated) into luminance signal Y and two types of color signals (or color-difference signals) CB and CR and converting these encoded signals into digital data called CIF or QCIF by an A-D converter and format converter (not illustrated). The image signal is written in the frame memory 101 as frame data one frame by one frame. The frame data written in the frame memory 101 is read out every a plurality of blocks and supplied to the facial region detecting circuit 102 and low-pass filter 103. The data can also be read out every frame.”</p> <p>Ueno at 5:31-40: “The locally-decoded data obtained through inverse DCT is stored in the frame memory built in the motion compensated interframe prediction circuit 113 and referenced so that the next frame data will be encoded. That is, the motion compensated interframe prediction circuit 113 compares the locally-decoded data of the previous frame stored in the built-in frame memory with the current frame data sent from the low-pass filter 103 to generate motion vector information.”</p>

Claim 4	Exemplary Disclosures
<p>[4] The process according to claim 1 further comprising drawing a tracking box around the target.</p>	<p>Ueno at 7:7-45: “Facial detection is executed according to the histograms.</p> <p>Face detection starts with the top of a head 30. The top of the head 30 is detected by searching the y-axis histogram from the top and selecting the point Ys which first exceeds</p>

Claim 4	Exemplary Disclosures
	<p>the present second threshold value.</p> <p>After the top of the head 30 is detected, the left end 31 and right end 32 of the head 30 are detected by searching the x-axis histogram from the left and right and selecting the points X_s and X_e which first exceed the second threshold value. . . .</p> <p>Then, the bottom end of the facial region is decided. . . . The facial region is specified by a rectangle 50 designated by coordinates X_s, X_e, Y_s, and Y_e as shown in FIG. 5.”</p> <p>Ueno at Figure 5:</p>  <p>FRAME DIFFERENCE IMAGE FIG. 5</p> <p>Ueno at 12:6-25: “The region specifying apparatus 120 comprises a coordinate input unit 801 and a region calculating unit 802 as shown in FIG. 17. The coordinate input apparatus 801 is an apparatus in which coordinates can directly be input in numerical values through a keyboard and uses a pointing device such as a mouse and a touch panel. Especially, a mouse or touch panel is very convenient for users because they can input coordinates accurately corresponding to the positions on the screen while viewing the image on the display screen. The region calculating circuit 802 is configured similarly to the region</p>

Claim 4	Exemplary Disclosures
	<p>calculating circuit 211 in FIG. 6, which outputs the region specifying signal 803 for discriminating the region specified by a plurality of coordinates sent from the coordinated input apparatus from other regions. To specify a region, two points A and B, for example, are specified by a mouse or touch panel as shown in FIG. 18. Thus, a rectangle 91 with corners A and B is specified.”</p> <p>Ueno at Figures 17-18:</p> <p style="text-align: center;">F I G. 17</p> <div style="text-align: center;">  <p style="margin-top: 10px;">↓</p>  </div> <p style="text-align: center;">F I G. 18</p> <p>Ueno at 13:17-23: “In the self-monitoring mode, low-pass-filtered images are displayed on the monitoring display 710. After a region is specified by the region specifying apparatus 120, the region specifying signal is continuously output. Thus, said frame line showing the specified region is superimposed on the image and displayed on the monitoring display 710.”</p> <p>Hashima 13:64-14:34: “FIG. 23 showing the manner in which a window is established. In FIG. 23, since the camera 20 and the target mark 10 are not positioned perpendicularly to each other, the target mark image 10A is seen flat. The target mark image 10A is read into the image memory 41 of the image processor 40, and recognized, and a window 44 is established around the recognized target mark image 10A.</p>

Claim 4	Exemplary Disclosures
	<p>The position of the center of gravity of the target mark image 10A which is obtained by the projected histogram processing when the target mark image 10A is recognized is regarded as the center of the window 44. Sizes $W1x$, $W1y$ of the window 44 are obtained by multiplying vertical and horizontal sizes $m1x$, $m1y$ of the target mark image 10A which are obtained by the projected histogram processing when the target mark image 10A is recognized, by a predetermined ratio, e.g., 2. In order to keep the region of the window 44 reliably in the image memory 41, a white region that is slightly greater than the product of the vertical and horizontal sizes of the target mark 10 and the predetermined ratio is provided in advance around the black circle 11 of the target mark 10 on the object 1. In this manner, a white region 43 can always be kept in the image memory 41. Even if the distance between the target mark 10 and the camera 20 varies resulting in a change in the size of the target mark image 10A as read into the image memory 41, the target mark image 10A does not exceed the white region 43, making it possible to establish the window 44 around the target mark image 10A.</p> <p>By keeping the white region 43 around the target mark image 10A and setting the window 44 in the white region 43, it is possible to remove a black region such as noise other than the target mark image 10A.</p> <p>When the target mark 10 starts to be tracked, the window is established using the projected histogram information obtained when the target mark image is recognized. When the target mark 10 is subsequently tracked, the window is established using new projected histogram information obtained upon each measurement made by the camera 20.”</p> <p>Hashima at Figure 23:</p>

Claim 4	Exemplary Disclosures
	<p style="text-align: center;">FIG.23</p>  <p>See also, Hashima at 8:55-57, 13:61-14:34, 24:44-25:23, Figure 53.</p>

D. Ground 3: Ueno In View Of Gilbert Teaches Or Suggests Every Step And Feature Of Claims 1-4

144. In my opinion, as shown in the charts below, Ueno in view of Gilbert and in view of the knowledge of a POSA, teaches or suggests every feature recited in Claims 1-4 of the '001 Patent, as construed by Petitioner.

1. Reasons To Combine Ueno And Gilbert

145. A POSA would have been motivated to combine Ueno and Gilbert because they are directed toward very similar systems that operate in a similar manner. Indeed, although there are many types of machine vision and tracking systems, Ueno and Gilbert both relate to the same, small, cross-section of these systems. As set forth above, both Ueno and Gilbert are directed to identifying and

tracking a target using an image processing system with a video camera input (Ex. 1007, Ueno at 3:1-8, 4:25-41; Ex. 1005, Gilbert at 47-48, Figure 1), where the video input comprises a succession of frames, each frame comprising a succession of pixels (Ex. 1007, Ueno at 1:44-49, 4:25-41; Ex. 1005, Gilbert at 47-48, 52), in which the target comprises pixels in one or more of a plurality of classes and one or more of a plurality of domains (Ex. 1007, Ueno at 7:8-16; Ex. 1005, Gilbert at 48). The Ueno and Gilbert systems also perform this process on a frame-by-frame basis (Ex. 1007, Ueno at 4:25-41, 5:31-40; Ex. 1005, Gilbert at 47-48, 52).

146. Moreover, the teachings in Gilbert and Ueno, taken as a whole, demonstrate additional reasons why a POSA would be motivated to combine the two references. For example, Gilbert recognizes that there are many possible domains that could be used for histogram plotting, but only explicitly teaches creating histograms in the intensity domain in a “video processor.” Ex. 1005, Gilbert at 48. Gilbert also discloses the construction of projections to determine the position of the target. However, these projections are entirely analogous to the “x- and y-axis histograms” of Ueno (and entirely analogous to the position histograms in the ’001 Patent at 20:60-21:29). Ueno showed that the projection described by Gilbert and performed by Gilbert’s “projection processor” were actually histograms using the X- and Y-coordinates of a class of image pixels. Thus, to improve on processing efficiency by using sharing components and/or

their designs, a POSA would have been motivated to combine Gilbert's image processing method with Ueno to utilize Gilbert's video processor for both the luminance histograms disclosed in Gilbert and the x- and y-axis histograms disclosed in Ueno. Moreover, there would have been a reasonable expectation of success that doing so would result in a more efficient target identification system.

147. Additionally, both Ueno and Gilbert identify and track a moving target using histograms (Ex. 1007, Ueno at 7:7-8:22; Ex. 1005, Gilbert at 48-49). Thus, they are not only in the same field of endeavor but also are directed to very similar systems and processes. A POSA would have recognized that Gilbert's invention could be applied to a similar device, such as the system in Ueno, in the same way, and thus would have been motivated to combine Ueno and Gilbert. A POSA would have also expected the combination to yield a predictable result because it would have involved applying known techniques to similar systems.

148. In particular, although Ueno describes a system that identifies and tracks a target and draws a tracking box around the target, it does not disclose a means of finding a center of the target. A POSA would have been motivated to improve Ueno's system to add a step of calculating the target's center, because this center calculation enables the addition of other functionality. For example a POSA would then have been able to add functionality to allow the camera to adjust its field of view to follow the target, a capability not present in Ueno's camera, which

is stationary. *Id.* Adding the center calculation would have allowed the Ueno system more flexibility to track targets physically moving around a room during a video call and would have allowed the Ueno system to be useful in other related applications. Thus, a POSA would have been motivated to look for prior art that teaches a tracking system where the system calculates the target center and where the camera automatically adjusts to keep the target centered in the field of view. As described above, Gilbert is an example of such a system and teaches that “[t]he boresight correction signals are used to control the azimuth and elevation pointing angles of the telescope. . . . keeping the target visible within the FOV.” Ex. 1005, Gilbert at 52. Moreover, there would have been a reasonable expectation of success that in doing so, a more accurate target identification system would result. Adding center calculation functionality would also allow Ueno to determine the orientation of the target face, which might aid in identifying facial features such as the eyes, nose, or mouth.

149. Although Gilbert recognizes that the center of the target can be found using “the target nose and tail points,” i.e., the X- and Y-maxima and minima, Gilbert considers this method to be prone to noise perturbations and thus not suitable for missile tracking systems. Gilbert thus employs a center calculation using the center-of-area of two halves of the target. A POSA would have recognized that Gilbert elected a center calculation using the center-of-area of two

halves of the target based on considerations that do not apply to a system like that described in Ueno. For example, Gilbert teaches the difficulty in finding the exact nose and tail points of a rocket and a significant risk that noise in the tracking system would make using the nose and tail points unreliable. However, in Ueno, the targets are human faces moving at slow speeds in contexts where noise would be at a minimum and not interfere with finding the extreme values. Thus, in order to benefit from faster, more efficient processing, a POSA would have been motivated to apply the simpler center calculation using the nose and tail points described in Gilbert rather than the more complex center-of-area calculation that Gilbert opted to use. This would have been an especially reasonable combination given that Ueno already calculates the X- and Y-minima and maxima, making the use of the extreme values to find the center a simple calculation involving only one additional step.

150. Thus, a POSA would have been motivated to combine Ueno's image processing system with that described by Gilbert to enhance the video camera to add center calculation capabilities that enable the camera to automatically track a target by physically adjusting the optical axis of the camera as disclosed in Gilbert, and would have expected such a combination to have a reasonable expectation of success.

2. Claim 1

- a) **[1 pre]: A process of tracking a target in an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels, the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains, the process performed by said system comprising, on a frame-by-frame basis**

151. The preamble in Claim 1 requires four different elements: (1) that the process “track[] a target,” (2) that a specific input signal comprise a sequence of frames where each frame comprises multiple pixels, (3) that the pixels must be able to be categorized according to their characteristics into at least one domain and one class, selected from a plurality of domains and classes, and (4) that the method must repeat itself “on a frame-by-frame basis.” Both Ueno and Gilbert disclose each of these steps as described below, and as more fully laid out in the claim chart in **IX.D.6**.

152. Ueno and Gilbert both disclose “[a] process of tracking a target.” Although Ueno does not specifically call out its tracking functionality, it does teach a terminal “for inputting movement image signals” that is connected to a frame memory. Ex. 1007, Ueno at 3:1-3. Further it teaches that the terminal is connected to a “facial region detecting circuit . . . for detecting the human facial region of the frame data.” *Id.* at 3:6-8. Taken together, these and similar disclosures clearly teach a system and method for identifying and tracking a face

across multiple frames of a video signal.

153. Gilbert disclosure of this limitation is more specific. As stated in the abstract and introduction of Gilbert, the entire point of developing the described system was to “produce[] a system for missile and aircraft identification and tracking.” Ex. 1005, Gilbert at 47. Specifically, the system is capable of identifying the object, locating its center, and tracking it to keep it in the field of view of the camera, even as it is rapidly moving through space. *Id.*

154. Ueno also discloses “an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels.” This limitation is met in its entirety simply because the input described is a video signal acquired by a video camera. Ex. 1007, Ueno at 4:25-26. Any such video input necessarily comprises a sequence of frames and each frame in a video input would also comprise many pixels. Nevertheless Ueno also teaches that “video signals are sequentially input to the image input terminal . . . in frames.” *Id.* Moreover, Ueno discloses that each frame is comprised of many pixels, as a POSA would expect in any video signal. *See id.* at 4:39-41, 7:8-16.

155. Gilbert also explicitly describes that both frames and pixels comprise the input signal:

The scene in the FOV of the TV camera is digitized to form an $n \times m$ matrix representation

$$P = (p_{ij})_{n, m}$$

of the pixel intensities p_{ij} . As the TV camera scans the scene, the video signal is digitized at m equally spaced points across each horizontal scan. During each video field, there are n horizontal scans which generate an $n \times m$ discrete matrix representation at 60 fields/s. A resolution of $m = 512$ pixels per standard TV line results in a pixel rate of 96 ns per pixel.

Ex. 1005, Gilbert at 48. Thus, both Ueno and Gilbert disclose the input signal taught in the '001 Patent.

156. Ueno meets the limitation “the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains.” This limitation is very similar to the step in [1 a] below, inasmuch as domains merely refer to pixel characteristics for use in a histogram and classes refer to the discrete divisions within the domain, i.e., bins, into which the pixels can be classified. *See* Ex. 1001, '001 Patent at 4:2-6 (listing specific pixel characteristics for use as domains); 6:1-3 (the pixels are classified “within each domain to selected classes within the domain”). Ueno teaches constructing histograms in the X- and Y-

domains. *See* Ex. 1007, Ueno at 7:8-16. Moreover, Ueno also teaches that the video input signal is split into a number of different constituent parts. Specifically, Ueno teaches that: “The video signal is, for example, obtained by encoding analog image signals obtained through picking-up of a human with a TV camera (not illustrated) into luminance signal Y- and two types of color signals (or color-difference signals).” *Id.* at 4:26-30. These image signals correspond closely to three of the described domains in the ’001 Patent, specifically hue and saturation (two color signals) and luminance.

157. The pixels described by Gilbert can also be categorized into classes and domains. Gilbert specifically teaches that there are “many features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures.” Ex. 1005, Gilbert at 48. Gilbert focuses primarily on the pixel characteristic or domain of intensity, which it describes as being divisible into 256 gray levels or classes. *Id.*

158. Ueno also meets the limitation that “the process [be] performed by said system . . . on a frame-by-frame basis.” Ueno clearly teaches that “[t]he image signal is written in the frame memory as frame data one frame by one frame. The frame data written in the frame memory . . . can also be read out every frame.” Moreover, Ueno specifically teaches creating histograms based on the differences between consecutive frames. *Id.* at 7:8-16. In order to accomplish creating

interframe difference images, Ueno necessarily processes the image signal on a frame-by-frame basis.

159. Gilbert similarly teaches that the video input signal is comprised of 60 individual frames per second: “During each video field, there are n horizontal scans which generate an $n \times m$ discrete matrix representation at 60 fields/s.” Ex. 1005, Gilbert at 48. Although Gilbert uses the word “field” instead of “frame,” a POSA would have understood these terms to be interchangeable. Indeed, it appears likely that Gilbert frequently uses the word field to avoid confusion with the tracking window frame, which is sometimes used in close proximity to “field.” *See id.* Gilbert further teaches that the target identification and tracking process occurs during each field (or frame). For example, Gilbert teaches that “[d]uring each field, the feature histograms are accumulated for the three regions of the tracking window.” *Id.* Thus, both Ueno and Gilbert teach that the image tracking and identification process is performed on a frame-by-frame basis as claimed in the ’001 Patent.

- b) **[1 a]: forming at least one histogram of the pixels in the one or more of a plurality of classes in the one or more of a plurality of domains, said at least one histogram referring to classes defining said target**

160. Both Ueno and Gilbert teach this limitation, as discussed below and as further shown in the claim chart at **IX.D.6**. In each case, at least one histogram is formed based on pixel characteristics (domains). Examples of domains are seen in

the '001 Patent: "The domains are preferably selected from the group consisting of i) luminance, ii) speed (V), iii) oriented direction (D1), iv) time constant (CO), v) hue, vi) saturation, and vii) first axis (x(m)), and viii) second axis (y(m))." Ex. 1001, '001 Patent at 4:2-6. In Ueno, histograms are generated on interframe images and specifically identify pixels that are different in the two frames. The disclosed histograms are generated in the domains of the X- and Y-axis:

FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.

Ex. 1007, Ueno at 7:7-16. Thus, in Ueno, the number of pixels having the value of "1" within a given column defined by an X-coordinate position in the interframe difference image is counted into the class that corresponds to that X-coordinate

value. These histograms in Ueno are analogous to the X- and Y- position histograms discussed in the '001 Patent at 20:60-21:29.

161. Gilbert similarly forms a histogram of the intensity domain, and teaches other features such as “texture, edge, and linearity measures” that could be used for histogram creation. Ex. 1005, Gilbert at 48. Gilbert teaches that “pixel intensity is digitized and quantized into eight bits (256 gray levels) counted into one of six 256-level histogram memories.” These gray levels constitute the classes within the intensity domain. The histogram is formed as each pixel in the frame is sorted into one of the 256 classes that corresponds to that pixel’s intensity level. The target, or targets, are defined by the classes because Gilbert takes it as a basic assumption that “the target image has some video intensities not contained in the immediate background.” *Id.* at 48. After dividing the image into background, plume and target regions, Gilbert teaches that “feature histograms are accumulated” for each of the three regions. *Id.* Moreover, to extent the Patent Owner argues that Gilbert’s intensity histogram does not include a plurality of “classes,” because the histogram contains every pixel (sorted into bins based on intensity level) rather than a subset of pixels that meet a threshold intensity characteristic, this argument should be rejected. Gilbert’s intensity histogram is entirely analogous to the velocity histogram in Figure 14a of the '001 Patent, as described at 20:54-59.

162. Moreover, Gilbert also discloses a projection processor that forms X- and Y-projection histograms using binary pictures generated by the video processor. Ex. 1005, Gilbert at Fig. 4. These projection histograms are analogous to the X- and Y- position histograms discussed in the '001 Patent at 20:60-21:29. Ex. 1002. Thus, Gilbert discloses creating histograms in a plurality of domains, including the intensity, the X-, and the Y-domains.

163. In these histograms, each “class” within a domain counts the number of pixels meeting certain threshold criteria. For example, Gilbert’s intensity histogram has 256 grayscale levels (i.e., 256 classes). Ex. 1005, Gilbert at 48. The histogram is formed as each pixel in the frame is sorted into one of the 256 classes for which that pixel’s intensity level fits. *Id.* In Ueno, the number of pixels having the value “1” (which indicates that a pixel meets a certain intensity threshold) within a given column defined by an X- coordinate is counted into the class that corresponds to that X-coordinate value. Similarly, in the X- and Y-projection histograms of Gilbert Figure 4, the number of pixels having the value “1” in the binary picture (which indicates that a pixel belongs to the target) within a given column defined by an X- or Y-coordinate is counted into the class that corresponds to that X- or Y-coordinate value. These binary selection processes in Ueno and Gilbert are entirely analogous to the histogram formation process discussed in the '001 Patent at 18:16-19:24.

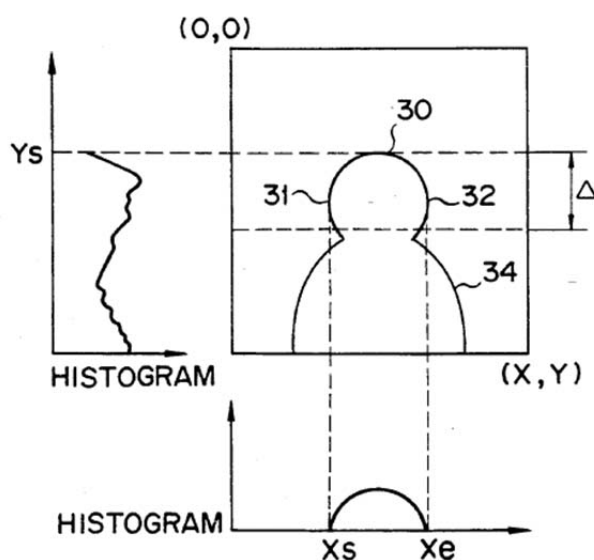
164. Thus, both Ueno and Gilbert teach that histograms are formed in at least one of a plurality of domains with each pixel being categorized into one or more classes within said domain.

c) **[1 b]: identifying the target in said at least one histogram itself**

165. Both Ueno and Gilbert disclose this step, as discussed below and shown in greater detail in the claim chart at **IX.D.6**. In each case, histograms are used to identify the target. For example, in Ueno, interframe differences are formed into X- and Y-histograms and used to identify a human face:

FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.

Ex. 1007, Ueno at 7:7-16. Figure 3 shows how the target is identified directly from the X- and Y-histograms:



F I G. 3

166. Similarly, in Gilbert, an intensity histogram is used to separate the target from the background and the plume. Ex. 1005, Gilbert at 48-50. This result is further used by the video processor and projection processor to locate and track the target. *Id.* Gilbert also teaches that the feature histograms “provide an estimate of the probability of feature *x* occurring in the background, plume, and target regions.” *Id.* at 49.

- d) [1 c]: wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between *X* and *Y* minima and maxima of the target**

167. Gilbert disclose the step of determining a center of the target, as described below and shown in more detail in the claim chart at **IX.D.6**. In Gilbert, a center is computed to “precisely determine the target position and orientation.”

Ex. 1005, Gilbert at 50. There, the target image is divided into a top and bottom section and a center point is found for each. *Id.* Once finding a center point for the top and bottom portions, the “target center –of-area (X_C, Y_C) and its orientation can be easily computed” *id.*, by connected the top and bottom center points with each other and finding the midpoint of the resulting line. For a single body target, this point must necessarily be between the X- and Y-minima and maxima of the target. The centering calculation used in Gilbert is shown diagrammatically in Figure 4, below:

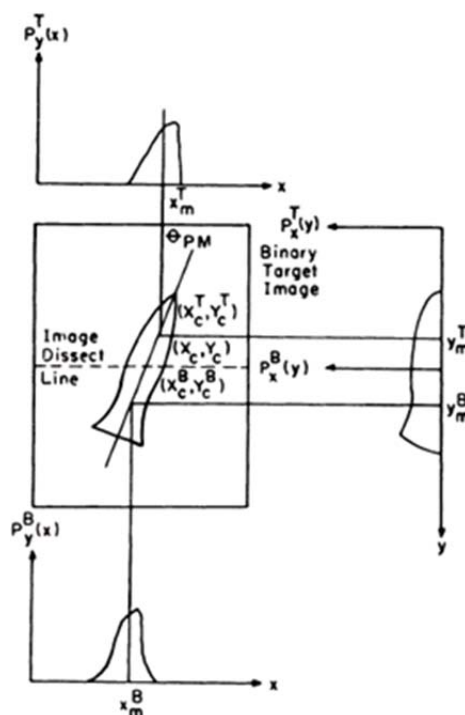


Fig. 4. Projection location technique.

Gilbert also teaches that it is possible for “the target nose and tail points” to be used in finding the center, and acknowledges that the “centroid of the target

image” may be located solely on the basis of the vertical and horizontal projections without performing the complex center-of-area calculation discussed above. *Id.* at 50.

168. A POSA would not have understood Claim 1 to require a specific equation be used to calculate a center point. Indeed for irregularly shaped objects, like those tracked in the '001 Patent, the concept of a center point is ambiguous at best because there are actually various center points that could be calculated according to different methods. For example, a center could be the center of mass, or the mean of the coordinates of the border of the object, or the point with the minimal sum of distances to all points of the area measured along paths inside the area, or could have one of a number of other definitions. Claim 1 of the '001 Patent only puts as a restriction that the center must be between the X- and Y-minima and maxima, or in other words, on the target itself. Using any of the centers described above, so long as they were on the target itself, would accomplish the aim of aiding in identification and tracking of the target.

169. To the extent the Patent Owner argues that Claim 1 requires that the X- and Y-minima and maxima be used in the calculation of the center point, it may be argued that Gilbert's disclosure, or at least the express calculation done in Gilbert, varies in some degree from this requirement. However, to the extent that Gilbert does not disclose this precise centering calculation, the standard centering

calculation disclosed in the '001 Patent would have been well known to a POSA.

A POSA would have desired to implement this standard centering calculation as the most cost effective and simple way to calculate a center point.

3. Claim 2

- a) **The process according to Claim 1, wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target**

170. Claim 2 contains a limitation identical to the final limitation in Claim 1. Thus, it is my opinion that Ueno and Gilbert both disclose this limitation for the same reasons discussed above. *See* Section **IX.D.2.d**.

4. Claim 3

- a) **The process according to Claim 1, wherein identifying the target in said at least one histogram further comprises determining the center of the target at regular intervals**

171. Both Ueno and Gilbert teach that the target identification process, including the center determination, take place at regular intervals. For example, Ueno discloses processing the camera input one frame at a time. *See e.g.*, Ex. 1007, Ueno at 1:45-46 (“[A] specific region in image signals sequentially input for every frame is detected which corresponds to a movable image”); 4:25-41 (“Video signals are sequentially input to the image input terminal 100 in frames. . . . The image signal is written in the frame memory 101 as frame data one frame by

one frame. . . . The data can also be read out every frame.”). Because determining the X- and Y-minima and maxima is part of target location and identification process, it necessarily takes place on the regular interval of each frame.

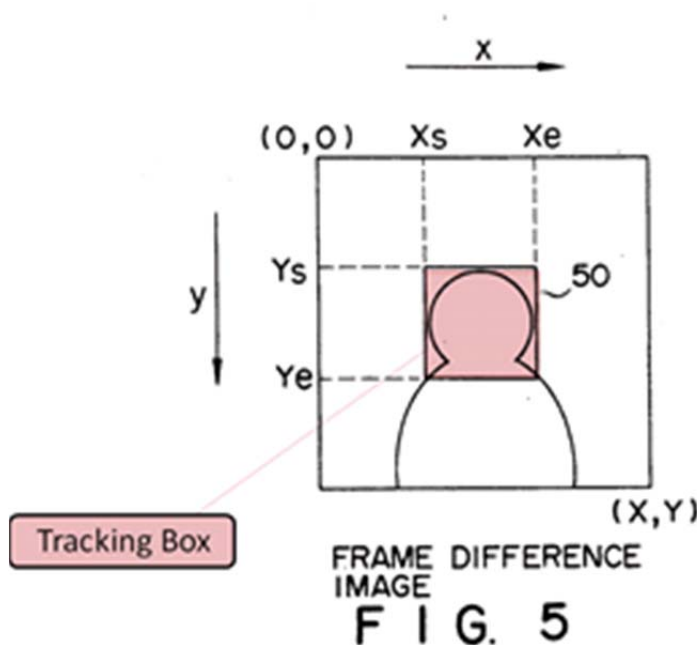
172. Similarly, Gilbert discloses analyzing the binary matrices on a frame-by-frame basis at a rate of 30 frames/second. *See, e.g.*, Ex. 1005, Gilbert at 50 (“In the projection processor, these matrices are analyzed field-by-field at 60 fields/s [i.e. 30 frames/s] using the projection based algorithm”). *See also* Ex. 1005, Gilbert at 55 (“The result after six processing frames . . .”). Because part of the analyzing or processing referred to involves generating a histogram to identify the target, and further finding the center of the target using the generated histogram, this process is necessarily performed at regular intervals, namely once per frame or once every 1/60 second. This frame-by-frame determination of the center of the target in Gilbert is identical to the description in the ’001 Patent at 25:3-25of determining the center position at regular intervals.

173. To the extent the Patent Owner argues that Ueno and Gilbert do not explicitly disclose that the center of the target is determined at regular intervals, a POSA would have implemented this feature in order to achieve smooth, predictable processing.

5. Claim 4

a) The process according to Claim 1 further comprising drawing a tracking box around the target

174. Ueno discloses the step of drawing a tracking box around the target. *See, e.g.*, Ex. 1007, Ueno at 7:43-45 (“The facial region is specified by a rectangle 50 designated by coordinates X_s , X_e , Y_s , and Y_e as shown in FIG. 5.”). This tracking box is shown in Ex. 1007, Ueno at Figure 5, annotated below (the solid black line around the box is in the original figure):

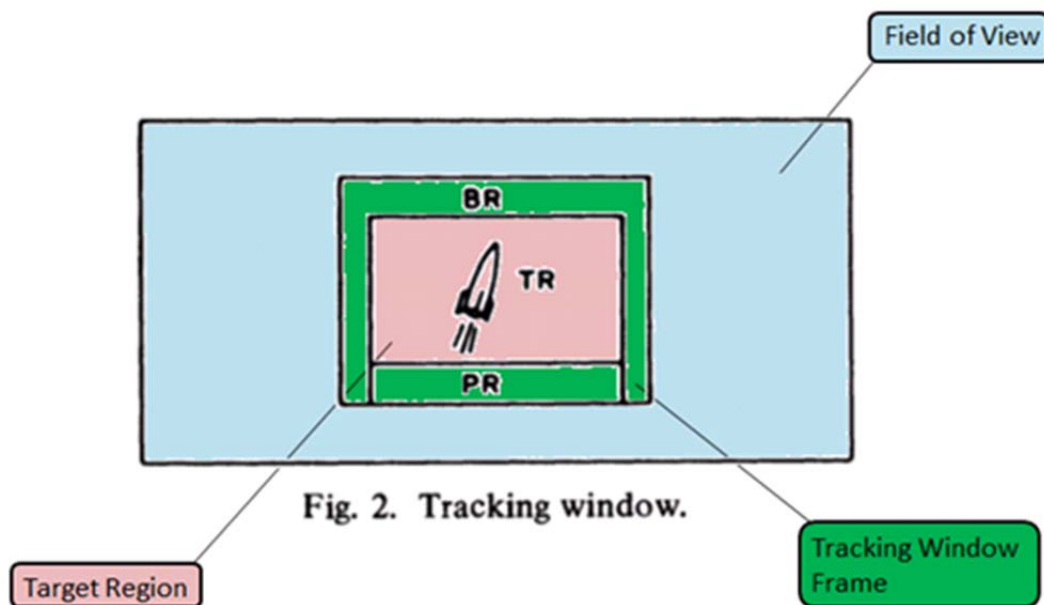


175. Ueno also clarifies that the tracking box is displayed on a monitor or display. For example, Ueno teaches that the region of interest is displayed with a frame that is superimposed on the original image and displayed on the monitor:

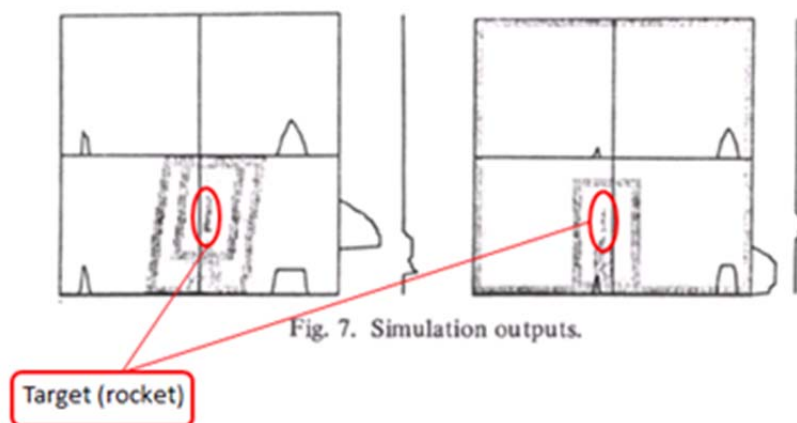
In the self-monitoring mode, low-pass-filtered images are displayed on the monitoring display 710. After a region is specified by the region specifying apparatus 120, the region specifying signal is continuously output. Thus, said frame line showing the specified region is superimposed on the image and displayed on the monitoring display 710.

Ex. 1007, Ueno at 13:17-23. *See also* Ex. 1007, Ueno at 12:6-25; Figure 18.

176. Gilbert also discloses the step of drawing a tracking box around the target. For example, Gilbert teaches that “[a] tracking window is placed about the target image, . . . The tracking window frame is partitioned into a background region (BR) and a plume region (PR). The region inside the frame is called the target region (TR) as shown in Fig. 2.” Ex. 1005, Gilbert at 48. This tracking window is shown in Figure 2 from Gilbert as annotated below:



Gilbert also shows that in simulations (using pre-recorded video of a rocket in flight as an input signal), the system was successful in tracking target by drawing a tracking box. *See, e.g.,* Ex. 1005, Gilbert at 55 (“Fig. 7 contains two representative simulation outputs selected from the first 100 fields of simulated digitized video.”). This is shown in Figure 7 from Gilbert, as annotated below:



177. Thus, Ueno and Gilbert discloses this limitation. To the extent the

Patent Owner argues that one of Ueno and Gilbert do not explicitly disclose drawing a tracking box around the target, a POSA would have desired to implement this feature in order to aid a user at a display to visually locate and follow a target.

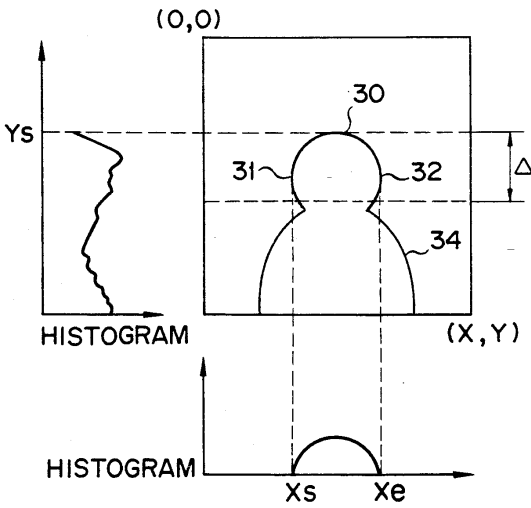
6. Detailed Application Of Ueno And Gilbert To Claims 1-4

178. The charts below provide a detailed mapping of Gilbert and Ueno to Claims 1-4 of the '001 Patent and show that Gilbert in view of Ueno teach or suggest every feature recited in Claims 1-4:

Claim 1	Exemplary Disclosures
[1.pre] A process of tracking a target in an input signal implemented using a system comprising an image processing system, the input signal comprising a succession of frames, each frame comprising a succession of pixels, the target comprising pixels in one or more of a plurality of classes in one or more of a plurality of domains, the process performed by said system comprising, on a frame-by-frame basis	<p>Ueno at 1:44-49: “According to the present invention, a specific region in image signals sequentially input for every frame is detected which corresponds to a movable image having the specific region and the region specifying signal to separate the specific region from other regions is output.”</p> <p>Ueno at 3:1-8: “In the embodiment shown in FIG. 1, a terminal 100 for inputting movement image signals is connected to a frame memory 101 for storing image signals for one frame (this is called frame data). The readout terminal of the frame memory 101 is connected to the input terminal of a facial region detecting circuit 102 for detecting the human facial region of the frame data and that of a low-pass filter 103 for filtering the frame data.”</p> <p>Ueno at 4:25-41 : “Video signals are sequentially input to the image input terminal 100 in frames. The video signal is, for example, obtained by encoding analog image signals obtained through picking-up of a human with a TV camera (not illustrated) into luminance signal Y and two types of color signals (or color-difference signals) CB and CR and converting these encoded signals into digital data called</p>

Claim 1	Exemplary Disclosures
	<p>CIF or QCIF by an A-D converter and format converter (not illustrated). The image signal is written in the frame memory 101 as frame data one frame by one frame. The frame data written in the frame memory 101 is read out every a plurality of blocks and supplied to the facial region detecting circuit 102 and low-pass filter 103. The data can also be read out every frame. The data readout unit is larger than the block size (8 pixels×8 pixels) which is the conversion unit (encoding unit) in the DCT circuit 107.”</p> <p>Ueno at 7:8-16: “Information for the interframe difference image is converted into binary data of "O" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.”</p> <p>Gilbert at 47-48: “The main optics is a high quality cinetheodolite used for obtaining extremely accurate (rms error - 3 arc-seconds) angular data on the position of an object in the FOV. It is positioned by the optical mount which responds to azimuthal and elevation drive commands, either manually or from an external source. The interface optics and imaging subsystem provides a capability to increase or decrease the imaged object size on the face of the silicon target vidicon through a 10:1 range, provides electronic rotation to establish a desired object orientation, performs an autofocus function, and uses a gated image intensifier to amplify the image and "freeze" the motion in the FOV. The camera output is statistically decomposed into background, foreground, target, and plume regions by the video processor, with this operation carried on at video rates for up to the full frame. The projection processor then analyzes the structure of the target regions to verify that the object selected as "target" meets the stored (adaptive) description of the object being</p>

Claim 1	Exemplary Disclosures
	<p>tracked. The tracker processor determines a position in the FOV and a measured orientation of the target, and decides what level of confidence it has in the data and decision. The control processor then generates commands to orient the mount, control the interface optics, and provide real-time data output. An I/O processor allows the algorithms in the system to be changed, interfaces with a human operator for tests and operation, and provides data to and accepts data from the archival storage subsystem where the live video is combined with status and position data on a video tape.”</p> <p>Gilbert at 48: “The four tracking processors, in turn, separate the target image from the background, locate and describe the target image shape, establish an intelligent tracking strategy, and generate the camera pointing signals to form a fully automatic tracking system.”</p> <p>Gilbert at 48: “The video processor receives the digitized video, statistically analyzes the target and background intensity distributions, and decides whether a given pixel is background or target [8]. A real-time adaptive statistical clustering algorithm is used to separate the target image from the background scene at standard video rates. The scene in the FOV of the TV camera is digitized to form an $n \times m$ matrix representation</p> $P = (p_{ij})_{n, m}$ <p>of the pixel intensities P_{ij}. As the TV camera scans the scene, the video signal is digitized at m equally spaced points across each horizontal scan. During each video field, there are n horizontal scans which generate an $n \times m$ discrete matrix representation at 60 fields/s. A resolution of $m = 512$ pixels per standard TV line results in a pixel rate of 96 ns per pixel.”</p> <p>Gilbert at 48: “Every 96 ns, a pixel intensity is digitized and quantized into eight bits (256 gray levels), counted into one of six 256-level histogram memories, and then</p>

Claim 1	Exemplary Disclosures
	<p>converted by a decision memory to a 2-bit code indicating its classification (target, plume, or background). There are many features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures. Throughout the following discussion of the clustering algorithm, pixel intensity is used to describe the pixel features chosen.”</p> <p>Gilbert at 52: “The outputs to the control processor are used to control the target location and size for the next frame.”</p>
<p>[1.a] forming at least one histogram of the pixels in the one or more of a plurality of classes in the one or more of a plurality of domains, said at least one histogram referring to classes defining said target,</p>	<p>Ueno at 7:7-16: “FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.”</p> <p>Ueno at Figure 3:</p>  <p style="text-align: center;">FIG. 3</p>

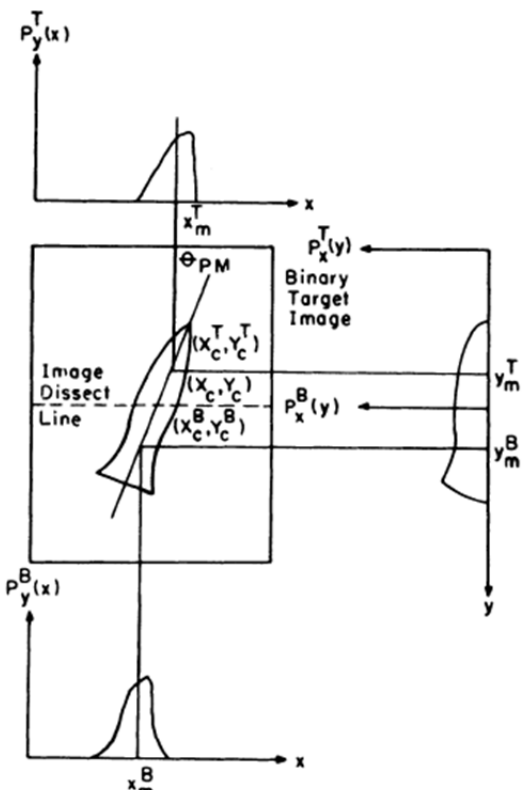
Claim 1	Exemplary Disclosures
	<p>Ueno at 8:5-12: FIG. 6 shows a facial region detecting circuit using the principle in FIG. 3. In this facial region detecting circuit, the x-axis histogram shown in FIG. 7B and the histogram shown in FIG. 7C are formed by the frame difference image shown in FIG. 7A. By extracting the coordinates of the transition points of these histograms, the face region is specified by a rectangle as shown in FIG. 5.</p> <p>Ueno at Figure 6:</p> <p style="text-align: center;">F I G. 6</p> <p>Ueno at Figure 7:</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>F I G. 7C</p> </div> <div style="text-align: center;"> <p>F I G. 7A</p> </div> </div> <div style="text-align: center; margin-top: 20px;"> <p>F I G. 7B</p> </div> <p>Gilbert at 48: “Every 96 ns, a pixel intensity is digitized</p>

Claim 1	Exemplary Disclosures
	<p>and quantized into eight bits (256 gray levels), counted into one of six 256-level histogram memories, and then converted by a decision memory to a 2-bit code indicating its classification (target, plume, or background). There are many features that can be functionally derived from relationships between pixels, e.g., texture, edge, and linearity measures. Throughout the following discussion of the clustering algorithm, pixel intensity is used to describe the pixel features chosen. . . .</p> <p>The tracking window frame is partitioned into a background region (BR) and a plume region (PR). The region inside the frame is called the target region (TR) as shown in Fig. 2. During each field, the feature histograms are accumulated for the three regions of each tracking window.</p> <p>The feature histogram of a region R is an integer-value, integer argument function $h^R(x)$. The domain of $h^R(x)$ is $[0, d]$, where d corresponds to the dynamic range of the analog-to-digital converter, and the range of $h^R(x)$ is $[0, r]$, where r is the number of pixels contained in the region R; thus, there are $r + 1$ possible values of $h^R(x)$. Since the domain $h^R(x)$ is a subset of the integers, it is convenient to define $h^R(x)$ as a one-dimensional array of integers</p> $h(0), h(1), h(2), \dots, h(d).$ <p>Letting x_i denote the ith element in the domain of x (e.g., $x_{25} = 24$), and $x(j)$ denote the jth sample in the region R (taken in any order), $h^R(x)$ may be generated by the sum</p> $h^R(x_i) = \sum_{j=1}^r \delta x_i, x(j) \quad ,$ <p>Gilbert at 49: “As each pixel in the region is processed, one (and only one) element of H is incremented as</p> $h[x(j)] \leftarrow h[x(j)] + 1.$

Claim 1	Exemplary Disclosures
	<p>When the entire- region has been scanned, <i>h</i> contains the distributions of pixels over intensity and is referred to as the feature histogram of the region <i>R</i>.”</p>
<p>[1.b] and identifying the target in said at least one histogram itself,</p>	<p>Ueno at 7:8-25: “FIG. 3 shows an interframe difference image input to the facial region detecting circuit 102. Information for the interframe difference image is converted into binary data of "0" or "1" by the preset first threshold level. Then the number of pixels having the binary data value of "1" or the value equal to or more than the first threshold value is counted in the vertical and horizontal directions of the screen and histograms of the pixels (x-and y-axis histograms) are formed. Facial detection is executed according to the histograms.</p> <p>Face detection starts with the top of a head 30. The top of the head 30 is detected by searching the y-axis histogram from the top and selecting the point <i>Ys</i> which first exceeds the present second threshold value.</p> <p>After the top of the head 30 is detected, the left end 31 and right end 32 of the head 30 are detected by searching the x-axis histogram from the left and right and selecting the points <i>Xs</i> and <i>Xe</i> which first exceed the second threshold value.”</p> <p>Ueno at Figure 3:</p>

<p>Claim 1</p>	<p>Exemplary Disclosures</p>
	<div data-bbox="607 277 1138 779"> </div> <div data-bbox="792 821 969 863"> <p>FIG. 3</p> </div> <div data-bbox="548 896 1430 1230"> <p>Ueno at 8:5-12: FIG. 6 shows a facial region detecting circuit using the principle in FIG. 3. In this facial region detecting circuit, the x-axis histogram shown in FIG. 7B and the histogram shown in FIG. 7C are formed by the frame difference image shown in FIG. 7A. By extracting the coordinates of the transition points of these histograms, the face region is specified by a rectangle as shown in FIG. 5.</p> </div> <div data-bbox="548 1281 824 1323"> <p>Ueno at Figure 7:</p> </div> <div data-bbox="597 1331 1216 1575"> <div data-bbox="597 1331 799 1575"> <p>FIG. 7C</p> </div> <div data-bbox="837 1331 1216 1575"> <p>FIG. 7A</p> </div> </div> <div data-bbox="837 1608 1128 1866"> <p>FIG. 7B</p> </div>

Claim 1	Exemplary Disclosures
	<p>Gilbert at 50: “The video processor described above separates the target image from the background and generates a binary picture, where target presence is represented by a "1" and target absence by a "0." The target location, orientation, and structure are characterized by the pattern of 1 entries in the binary picture matrix, and the target activity is characterized by a sequence of picture matrices. In the projection processor, these matrices are analyzed field-by-field at 60 fields/s using projection-based classification algorithms to extract the structural and activity parameters needed to identify and track the target. . .</p> <p>In general, a projection integrates the intensity levels of a picture along parallel lines through the pattern, generating a function called the projection. . . .”</p> <p>Gilbert at 50: “To precisely determine the target position and orientation, the target center-of-area points are computed for the top section (X_c^T, Y_c^T) and bottom section (X_c^B, Y_c^B) of the tracking parallelogram using the projections. Having these points, the target center-of-area (X_c, Y_c) and its orientation can be easily computed (Fig. 4):</p> $X_c = \frac{X_c^T + X_c^B}{2}$ $Y_c = \frac{Y_c^T + Y_c^B}{2}$ $\phi = \tan^{-1} \frac{Y_c^T - Y_c^B}{X_c^T - X_c^B}.$ <p>Gilbert at Figure 4:</p>

Claim 1	Exemplary Disclosures
	 <p data-bbox="677 1064 1125 1098">Fig. 4. Projection location technique.</p>
<p data-bbox="178 1134 534 1549">[1.c] wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target.</p>	<p data-bbox="547 1134 1430 1383">Gilbert at 50: “To precisely determine the target position and orientation, the target center-of-area points are computed for the top section (X_c^T, Y_c^T) and bottom section (X_c^B, Y_c^B) of the tracking parallelogram using the projections. Having these points, the target center-of-area (X_c, Y_c) and its orientation can be easily computed (Fig. 4):</p> $X_c = \frac{X_c^T + X_c^B}{2}$ $Y_c = \frac{Y_c^T + Y_c^B}{2}$ $\phi = \tan^{-1} \frac{Y_c^T - Y_c^B}{X_c^T - X_c^B}.$ <p data-bbox="547 1701 1430 1852">The top and bottom target center-of-area points are used, rather than the target nose and tail points, since they are much easier to locate, and more importantly, they are less sensitive to noise perturbations.”</p>

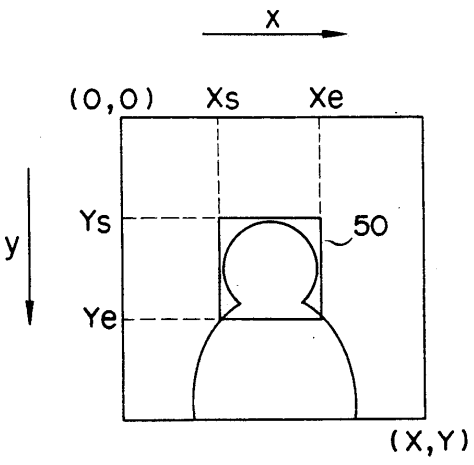
<p>Claim 1</p>	<p>Exemplary Disclosures</p>
	<p>Gilbert at Figure 4:</p> <p>Fig. 4. Projection location technique.</p>


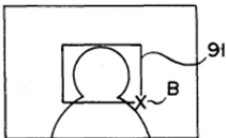
Claim 2	Exemplary Disclosures
[2] The process according to claim 1, wherein identifying the target in said at least one histogram further comprises determining a center of the target to be between X and Y minima and maxima of the target.	<i>See, e.g.</i> , disclosures in Claim [1.c].

Claim 3	Exemplary Disclosures
<p>[3] The process according to claim 1, wherein identifying the target in said at least one histogram further comprises determining the center of the target at regular intervals.</p>	<p>Ueno at 1:44-49: “According to the present invention, a specific region in image signals sequentially input for every frame is detected which corresponds to a movable image having the specific region and the region specifying signal to separate the specific region from other regions is output.”</p> <p>Ueno at 4:25-39: “Video signals are sequentially input to the image input terminal 100 in frames. The video signal is, for example, obtained by encoding analog image signals obtained through picking-up of a human with a TV camera (not illustrated) into luminance signal Y and two types of color signals (or color-difference signals) CB and CR and converting these encoded signals into digital data called CIF or QCIF by an A-D converter and format converter (not illustrated). The image signal is written in the frame memory 101 as frame data one frame by one frame. The frame data written in the frame memory 101 is read out every a plurality of blocks and supplied to the facial region detecting circuit 102 and low-pass filter 103. The data can also be read out every frame.”</p> <p>Ueno at 5:31-40: “The locally-decoded data obtained through inverse DCT is stored in the frame memory built in the motion compensated interframe prediction circuit 113 and referenced so that the next frame data will be encoded. That is, the motion compensated interframe prediction circuit 113 compares the locally-decoded data of the previous frame stored in the built-in frame memory with the current frame data sent from the low-pass filter 103 to generate motion vector information.”</p> <p>Gilbert at 47: “The camera output is statistically decomposed into background, foreground, target, and plume regions by the video processor, with this operation carried on at video rates for up to the full frame. The projection processor then analyzes the structure of the target regions to verify that the object selected as "target"</p>

Claim 3	Exemplary Disclosures
	<p>meets the stored (adaptive) description of the object being tracked.”</p> <p>Gilbert at 50: “The video processor described above separates the target image from the background and generates a binary picture, where target presence is represented by a "1" and target absence by a "0." The target location, orientation, and structure are characterized by the pattern of 1 entries in the binary picture matrix, and the target activity is characterized by a sequence of picture matrices. In the projection processor, these matrices are analyzed field-by-field at 60 fields/s using projection-based classification algorithms to extract the structural and activity parameters needed to identify and track the target.”</p> <p>Gilbert at 55: “The result after six processing frames, shown in Fig. 8(b), verifies the effectiveness of the processing algorithms used in the RTV tracker.”</p>

Claim 4	Exemplary Disclosures
<p>[4] The process according to claim 1 further comprising drawing a tracking box around the target.</p>	<p>Ueno at 7:7-45: “Facial detection is executed according to the histograms.</p> <p>Face detection starts with the top of a head 30. The top of the head 30 is detected by searching the y-axis histogram from the top and selecting the point Ys which first exceeds the present second threshold value.</p> <p>After the top of the head 30 is detected, the left end 31 and right end 32 of the head 30 are detected by searching the x-axis histogram from the left and right and selecting the points Xs and Xe which first exceed the second threshold value. . . .</p> <p>Then, the bottom end of the facial region is decided. . . . The facial region is specified by a rectangle 50 designated by coordinates Xs, Xe, Ys, and Ye as shown in FIG. 5.”</p>

Claim 4	Exemplary Disclosures
	<p>Ueno at Figure 5:</p>  <p style="text-align: center;">FRAME DIFFERENCE IMAGE FIG. 5</p> <p>Ueno at 12:6-25: “The region specifying apparatus 120 comprises a coordinate input unit 801 and a region calculating unit 802 as shown in FIG. 17. The coordinate input apparatus 801 is an apparatus in which coordinates can directly be input in numerical values through a keyboard and uses a pointing device such as a mouse and a touch panel. Especially, a mouse or touch panel is very convenient for users because they can input coordinates accurately corresponding to the positions on the screen while viewing the image on the display screen. The region calculating circuit 802 is configured similarly to the region calculating circuit 211 in FIG. 6, which outputs the region specifying signal 803 for discriminating the region specified by a plurality of coordinates sent from the coordinated input apparatus from other regions. To specify a region, two points A and B, for example, are specified by a mouse or touch panel as shown in FIG. 18. Thus, a rectangle 91 with corners A and B is specified.”</p> <p>Ueno at Figures 17-18:</p>

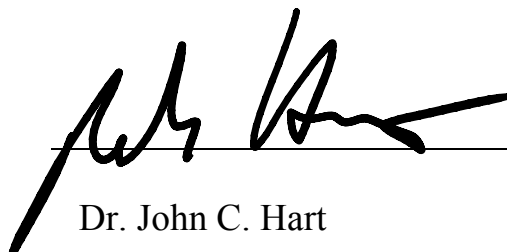
Claim 4	Exemplary Disclosures
	<p data-bbox="641 264 781 296">F I G. 17</p>  <p data-bbox="727 499 756 537">↓</p>  <p data-bbox="670 701 812 732">F I G. 18</p> <p data-bbox="548 793 1438 1087">Ueno at 13:17-23: “In the self-monitoring mode, low-pass-filtered images are displayed on the monitoring display 710. After a region is specified by the region specifying apparatus 120, the region specifying signal is continuously output. Thus, said frame line showing the specified region is superimposed on the image and displayed on the monitoring display 710.”</p> <p data-bbox="548 1134 1438 1386">Gilbert at 48: “The basic assumption of the clustering algorithm is that the target image has some video intensities not contained in the immediate background. A tracking window is placed about the target image, as shown in Fig. 2, to sample the background intensities immediately adjacent to the target image. . . .</p> <p data-bbox="548 1432 1438 1684">The tracking window frame is partitioned into a background region (BR) and a plume region (PR). The region inside the frame is called the target region (TR) as shown in Fig. 2. During each field, the feature histograms are accumulated for the three regions of each tracking window.”</p> <p data-bbox="548 1730 860 1764">Gilbert at Figure 2:</p>

Claim 4	Exemplary Disclosures
	<div data-bbox="576 268 1317 632" data-label="Image">A diagram of a tracking window. It consists of a large outer rectangle. Inside it is a smaller rectangle. Inside that is another rectangle. The top of the innermost rectangle is labeled 'BR'. The bottom is labeled 'PR'. In the center of the innermost rectangle is a stylized drawing of a rocket or missile, labeled 'TR' to its right.</div> <p data-bbox="745 657 1146 695">Fig. 2. Tracking window.</p> <p data-bbox="548 711 1422 1136">Gilbert at 55: “The simulation displays the target and plume points as they are identified by the video processor. These points are superimposed on a cross hair which locates the boresight of the tracking optics. In addition to the digitized images, the simulation displays the target and plume tracking windows and the projections accumulated for each video field. These outputs provide a direct measure of the performance of the four processors as well as an overall measure of the tracking accuracy of the system. . . .</p> <p data-bbox="548 1182 1365 1304">Fig. 7 contains two representative simulation outputs selected from the first 100 fields of simulated digitized video.”</p> <p data-bbox="548 1350 860 1388">Gilbert at Figure 7:</p> <div data-bbox="558 1444 1349 1745" data-label="Figure">Two side-by-side simulation outputs. Each is a square frame divided into four quadrants by a horizontal and vertical line. In the bottom-left quadrant of each frame, there is a dark, irregular shape representing a target or plume. The shapes are slightly different in the two frames, showing a sequence of frames.</div> <p data-bbox="805 1759 1122 1791">Fig. 7. Simulation outputs.</p>

IX. CONCLUSION

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

Date: November 29, 2016



Dr. John C. Hart