

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

VUSIONGROUP SA,
Petitioner

v.

HANSHOW TECHNOLOGY CO., LTD.,
Patent Owner.

IPR2024-01045

U.S. Patent No. 11,087,272

**DECLARATION OF ANDREW WOLFE, PH.D.
IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 11,087,272**

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND AND QUALIFICATIONS	1
A.	Relevant Academic Experience	2
B.	Relevant Professional Experience.....	5
III.	MATERIALS REVIEWED	8
IV.	LEVEL OF ORDINARY SKILL IN THE ART	12
V.	RELEVANT LEGAL STANDARDS	13
VI.	SUMMARY OF OPINIONS.....	23
VII.	BRIEF DESCRIPTION OF THE TECHNOLOGY	23
A.	Background of the Technology.....	23
1.	<i>Machine Learning</i> by Tom Mitchell (“Mitchell”).....	23
2.	<i>A Statistical Model for General Contextual Object Recognition</i> by Peter Carbonetto, et al.	26
3.	<i>Hello! My name is... Buffy’ – Automatic Naming of Characters in TV Video</i> by Mark Everingham, et al. (“Everingham”).....	29
4.	<i>Recognizing Products: A Per-Exemplar Multi-Label Image Classification Approach</i> by Marian George, et al. (“George”).....	30
5.	<i>Product Counting Using Images with Application to Robot- based Retail Stock Assessment</i> by Nishant Kejriwal, et al. (“Kejriwal”)	31
6.	<i>Remote Retail Monitoring and Stock Assessment Using Mobile Robots</i> by Swagat Kumar, et al. (“Kumar”).....	32
7.	<i>Recognizing Groceries in situ Using in vitro Training Data</i> by Michele Merler, et al. (“Merler”).....	33
8.	<i>Ontological Supervision for Fine Grained Classification of Street View Storefronts</i> by Yair Movshovitz-Attias, et al. (“Movshovitz-Attias”)	35

9.	<i>Toward Retail Product Recognition on Grocery Shelves</i> by Gül Varol, et al. (“Varol”)	36
10.	<i>Searching the Web with Mobile Images for Location Recognition</i> by Tom Yeh, et al. (“Yeh”)	37
B.	Summary of Points on the Background of the Technology.....	38
VIII.	OVERVIEW OF THE ’272 PATENT	41
A.	Specification.....	41
B.	Prosecution History	45
C.	Neither the ’785 nor ’509 Provisionals Provide Written Description Support for the ’272 Patent Claims	46
D.	The ’272 Patent Is Entitled Only to a March 28, 2017, Priority Date Based on Its Originally-Filed Claims	49
IX.	OVERVIEW OF THE PRIOR ART REFERENCES	50
A.	Schwartz.....	50
1.	Analogous Art.....	50
2.	Schwartz is Entitled to the Earlier Effective Filing Date of its Provisional Application.....	50
3.	Overview	52
B.	Katircioglu.....	55
1.	Analogous Art.....	55
2.	Overview	55
C.	Avegliano ’430.....	61
1.	Analogous Art.....	61
2.	Overview	61
D.	Avegliano ’157.....	63
1.	Analogous Art.....	63
2.	Overview	64
E.	Rodriguez	67
1.	Analogous Art.....	67
2.	Overview	67

X.	CLAIM CONSTRUCTION	70
XI.	DETAILED EXPLANATION OF GROUNDS.....	71
A.	Ground 1: Schwartz in view of Katircioglu Renders Claims 1, 3-13, 15-21 Obvious	71
1.	Motivation to Combine Schwartz and Katircioglu.....	71
2.	Independent Claim 1	73
a.	Claim 1.0: “A method comprising:”	73
b.	Claim 1.1: “capturing one or more images of a portion of a shelf with a camera;”	74
c.	Claim 1.2: “detecting and reading content of a shelf label on the portion of the shelf within the one or more images;”	75
d.	Claim 1.3: “defining one or more product facing bounding boxes surrounding one or more corresponding facings of a product within the one or more images;”	82
e.	Claim 1.4: “associating the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;”	86
f.	Claim 1.5: “cross-correlating the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;”	92
g.	Claim 1.6: “building training data based on the association of the shelf label content with the cross- correlated product facing bounding boxes;”	94
h.	Claim 1.7: “using the training data to build a product classifier; and”	96
i.	Claim 1.8: “using the product classifier to classify an item as the product.”	98

3.	Independent Claim 10	100
4.	Independent Claim 18	102
5.	Claims 3, 15	104
a.	Claim 3: “The method of claim 1, further comprising using manual input to identify potential inventory designated in the one or more product facing bounding boxes.”	104
b.	Claim 15: “The method of claim 10, further comprising using manual input to identify potential inventory designated in the one or more bounding boxes.”	105
6.	Claim 4: “The method of claim 1, wherein using the training data to build a product classifier comprises integrating training to: (1) learn product image descriptors and (2) classify items into a neural network.”	107
7.	Claim 5: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises manually associating the shelf label content with each of the one or more product facing bounding boxes.”	108
8.	Claims 6, 17	110
a.	Claim 6: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”	110
b.	Claim 17: “The method of claim 10, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”	110
9.	Claim 7	111
a.	Claim 7.0: “The method of claim 1, wherein capturing one or more images of a portion of a shelf comprises	

capturing one or more images of a portion of a shelf in a first store; and further comprising:"	111
b. Claim 7.1.1: "capturing one or more further images of portion of another shelf in a second store;"	112
c. Claim 7.1.2: "reading content of another shelf label on the portion of the other shelf within the one or more further images;"	112
d. Claim 7.1.3: "defining one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and"	112
e. Claim 7.1.4: "associating the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;"	113
f. Claim 7.2: "wherein cross-correlating comprises cross-correlating the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the second store by matching the shelf label content with the other shelf label content; and"	113
g. Claim 7.3: "wherein building training data comprises building the training data based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store."	114
10. Claim 11	114
11. Claim 19	115
12. Claims 8, 12, 20	116
a. Claim 8: "The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images subsequent to capturing the one or more images."	116

b.	Claim 12: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further subsequent to capturing the one or more images.”	116
c.	Claim 20: “The system of claim 18, wherein instructions configured to capture one or more further images comprise instructions configured to capture the one or more further images subsequent to capturing the one or more images.”	116
13.	Claims 9, 13	117
a.	Claim 9: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”	117
b.	Claim 13: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”	117
14.	Claim 16: “The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.”	117
15.	Claim 21: “The system of claim 18, wherein instructions configured to capture one or more further images of another portion of a shelf comprise instructions configured to capture the one or more further images.”	119
B.	Even if the ’272 Patent Gets Its Earliest Provisional Date, Schwartz Renders Claims 1, 3-13, and 15-21 Obvious	119
1.	Schwartz Is Entitled to Its Provisional Date	119
2.	Schwartz Renders the Claims Obvious, Based Solely on What is in the Schwartz Provisional	121
C.	Ground 2: Obviousness over Schwartz, Katircioglu, and Avegliano ’430 (Claims 2, 14)	123
1.	Motivation to Combine Schwartz, Katircioglu, and Avegliano ’430	123

1. Claims 2, 14	127
a. Claim 2: “The method of claim 1, wherein capturing one or more images comprises capturing the one or more images with a camera attached to a moveable robot.”	127
b. Claim 14: “The method of claim 10, wherein capturing one or more images of a portion of a shelf with a camera comprises capturing the one or more images with a camera attached to a moveable robot.”	128
D. Ground 3: Avegliano ’430 in view of Avegliano ’157 and Rodriguez Renders Obvious Claims 1, 3-6, 8-10, 12-13, 15-18, and 20-21	131
1. Motivation to Combine Avegliano ’157 and Rodriguez	131
2. Independent Claim 1	135
a. Claim 1.0: “A method comprising:”	135
b. Claim 1.1: “capturing one or more images of a portion of a shelf with a camera;”	137
c. Claim 1.2: “detecting and reading content of a shelf label on the portion of the shelf within the one or more images;”	140
d. Claim 1.3: “defining one or more product facing bounding boxes surrounding one or more corresponding facings of a product within the one or more images;”	142
e. Claim 1.4: “associating the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;”	147
f. Claim 1.5: “cross-correlating the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;”	150

g.	Claim 1.6: “building training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes;”	158
h.	Claim 1.7: “using the training data to build a product classifier; and”	160
i.	Claim 1.8: “using the product classifier to classify an item as the product.”	162
3.	Independent Claim 10	162
4.	Independent Claim 18	164
5.	Claims 3, 15	167
a.	Claim 3: “The method of claim 1, further comprising using manual input to identify potential inventory designated in the one or more product facing bounding boxes.”	167
b.	Claim 15: “The method of claim 10, further comprising using manual input to identify potential inventory designated in the one or more bounding boxes.”	167
6.	Claim 4: “The method of claim 1, wherein using the training data to build a product classifier comprises integrating training to: (1) learn product image descriptors and (2) classify items into a neural network.”	168
7.	Claim 5: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises manually associating the shelf label content with each of the one or more product facing bounding boxes.”	169
8.	Claims 6, 17	171
a.	Claim 6: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”	171

b.	Claim 17: “The method of claim 10, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”	171
9.	Claim 16: “The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.”	172
10.	Claim 20: “The system of claim 18, wherein instructions configured to capture one or more further images comprise instructions configured to capture the one or more further images subsequent to capturing the one or more images.”	175
11.	Claim 21: “The system of claim 18, wherein instructions configured to capture one or more further images of another portion of a shelf comprise instructions configured to capture the one or more further images.”	176
E.	Ground 4: Avegliano ’157, Rodriguez, and Avegliano ’430 Renders Obvious Claims 2, 7-9, 11-14, 16, and 19	177
1.	Motivation to Combine Avegliano ’157 and Rodriguez with Avegliano ’430	177
2.	Claims 2, 14	182
a.	Claim 2: “The method of claim 1, wherein capturing one or more images comprises capturing the one or more images with a camera attached to a moveable robot.”	182
b.	Claim 14: “The method of claim 10, wherein capturing one or more images of a portion of a shelf with a camera comprises capturing the one or more images with a camera attached to a moveable robot.”	182
3.	Claim 7	185
a.	Claim 7.0: “The method of claim 1, wherein capturing one or more images of a portion of a shelf comprises capturing one or more images of a portion of a shelf in a first store; and further comprising:”	186

b.	Claim 7.1.1: “capturing one or more further images of portion of another shelf in a second store;”	186
c.	Claim 7.1.2: “reading content of another shelf label on the portion of the other shelf within the one or more further images;”	188
d.	Claim 7.1.3: “defining one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and”	188
e.	Claim 7.1.4: “associating the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;”	188
f.	Claim 7.2: “wherein cross-correlating comprises cross-correlating the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the second store by matching the shelf label content with the other shelf label content; and”	189
g.	Claim 7.3: “wherein building training data comprises building the training data based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store.”	189
4.	Claim 11	189
5.	Claim 19	190
6.	Claims 8, 12	191
a.	Claim 8: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images subsequent to capturing the one or more images.”	191
b.	Claim 12: “The method of claim 11, wherein capturing one or more further images comprises capturing the one	

or more further subsequent to capturing the one or more images.”	191
7. Claims 9, 13	192
a. Claim 9: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”	192
b. Claim 13: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”	192
8. Claim 16: “The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.”.....	192
XII. SECONDARY CONSIDERATIONS	195
XIII. CONCLUSION.....	196

I, Andrew Wolfe, declare as follows:

I. INTRODUCTION

1. I have been retained by VusionGroup SA (“VusionGroup” or “Petitioner”) as an independent expert consultant in this proceeding before the United States Patent and Trademark Office (“PTO”).

2. I am being compensated at a rate of \$650/hour for my services in this proceeding, which was my regular and customary rate when starting this analysis.

3. 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.

4. I have been asked to consider whether certain references disclose or suggest the features recited in the claims of U.S. Patent No. 11,087,272 (“the ’272 patent”) (Ex. 1001).¹ My opinions are set forth below.

II. BACKGROUND AND QUALIFICATIONS

5. I am an independent consultant. All of my opinions stated in this declaration are based on my own personal knowledge and professional judgment. In forming my opinions, I have relied on my education, experience, and knowledge

¹ Where appropriate, I refer to exhibits I understand will be attached to the petition for *inter partes* review of the ’272 patent (the “Petition”).

regarding electrical engineering, computer science, user interface design, and experience with consumer product design.

6. I am over 18 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein. A copy of my current curriculum vitae (“CV”), which details my education and professional and academic experience, as well as a list of all publications I have authored in the past ten years, is included as Ex. 1004 in this proceeding. The following provides an overview of some of my experience that is relevant to the matters set forth in this declaration.

A. Relevant Academic Experience

7. In 1985, I earned the B.S.E.E. degree in Electrical Engineering and Computer Science from the Johns Hopkins University. In 1987, I received the M.S. degree in Electrical and Computer Engineering from Carnegie Mellon University and in 1992, I received the Ph.D. degree in Computer Engineering from Carnegie Mellon University. My doctoral dissertation proposed a new approach for the architecture of a computer processor.

8. I have taught at some of the world’s leading institutions in the fields of processor technology, computer systems, consumer electronics, software, design tools, data security, cryptography and intellectual property issues. These institutions include Stanford University, Princeton University, Carnegie Mellon University, and Santa Clara University.

9. From 1991 through 1997, I served on the Faculty of Princeton University as an Assistant Professor of Electrical Engineering. At Princeton, I taught undergraduate and graduate-level courses in Computer Architecture, Advanced Computer Architecture, Display Technology, and Microprocessor Systems, and conducted sponsored research in the area of computer systems and related topics. From 1999 through 2002, I taught a Computer Architecture course to both undergraduate and graduate students at Stanford University multiple times as a Consulting Professor. At Princeton, I received several teaching awards, both from students and from the School of Engineering. I have also taught advanced microprocessor architecture to industry professionals in IEEE and ACM sponsored seminars. I am currently a lecturer at Santa Clara University teaching courses on Microprocessor Systems, Real-time Embedded Systems, Advanced Logic Design, System Prototyping, Product Development, and Mechatronics. I have been working on image-processing applications since the mid-90s. I have supervised numerous undergraduate research projects incorporating image processing, computer vision, image recognition, camera-based text recognition, and machine-learning-based classifiers. These have included both custom algorithm development and the use of open-source libraries such as Open-CV and YOLO.

10. I have published more than fifty peer-reviewed papers in computer architecture and computer systems and IC (integrated circuit) design. My

contributions in the areas of low-power systems and power management have been widely recognized. My CV lists all publications that I have authored in the previous 10 years.

11. I have also chaired IEEE and ACM conferences in microarchitecture and IC design and served as an associate editor for IEEE and ACM journals. I served on the IEEE Computer Society Awards committee. I am an IEEE Fellow, and IEEE Computer Society Distinguished Contributor, and a Member of ACM. My elevation to IEEE Fellow was based on contributions in hardware code compression of embedded software, power consumption analysis, and optimization. I am a named inventor on at least fifty-seven U.S. patents and thirty-seven foreign patents including patents related to image processing and pattern recognition.

12. I have been the invited keynote speaker at the ACM/IEEE International Symposium on Microarchitecture and at the International Conference on Multimedia. I have also been an invited speaker on various aspects of technology or the PC industry at numerous industry events including the Intel Developer's Forum, Microsoft Windows Hardware Engineering Conference, Microprocessor Forum, Embedded Systems Conference, Comdex, and Consumer Electronics Show as well as at the Harvard Business School and the University of Illinois Law School. I have been interviewed on subjects related to technology and the electronics industry by publications such as the Wall Street Journal, New York Times, LA Times, Time,

Newsweek, Forbes, and Fortune as well as CNN, NPR, and the BBC. I have also spoken at dozens of universities including MIT, Stanford, University of Texas, Carnegie Mellon, UCLA, University of Michigan, Rice University, and Duke University.

B. Relevant Professional Experience

13. I am the founder and sole employee of Wolfe Consulting. Through Wolfe Consulting, I provide technical and business analytics to businesses on processor technology, computer systems, consumer electronics, software, design tools, data security, cryptography and intellectual property issues. I have more than thirty years' experience developing products, researching, consulting, and teaching in those fields. In that time, I have worked as a computer architect, computer system designer, and as an executive in the PC and electronics business.

14. I have more than 35 years of experience as a computer architect, computer system designer, personal computer graphics designer, educator, and executive in the electronics industry.

15. In 1983, I began designing touch sensors, microprocessor-based computer systems, and I/O (input/output) cards for personal computers as a senior design engineer for Touch Technology, Inc. During the course of my design projects with Touch Technology, I designed I/O cards for PC-compatible computer systems, including the IBM PC-AT, to interface with interactive touch-based computer

terminals that I designed for use in public information systems. I continued designing and developing related technology as a consultant to the Carroll Touch division of AMP, Inc., where in 1986 I designed one of the first custom touch-screen integrated circuits. I designed the touch/pen input system for the Linus WriteTop, which many believe to be the first commercial tablet computer.

16. From 1986 through 1987, I designed and built a high-performance computer system as a student at Carnegie Mellon University. From 1986 through early 1988, I also developed the curriculum, and supervised the teaching laboratory, for processor design courses.

17. In the latter part of 1989, I worked as a senior design engineer for ESL-TRW Advanced Technology Division. While at ESL-TRW, I designed and built a bus interface and memory controller for a workstation-based computer system, and also worked on the design of a multiprocessor system.

18. At the end of 1989, I (along with some partners) reacquired the rights to the technology I had developed at Touch Technology and at AMP, and founded The Graphics Technology Company. Over the next seven years, as an officer and a consultant for The Graphics Technology Company, I managed the company's engineering development activities and personally developed dozens of touch screen sensors, controllers, and interactive touch-based computer systems. I designed

components for many point-of-sale systems and aided in the system development including products used by Sears, McDonalds, UPS, and other high-volume retailers.

19. I have consulted, formally and informally, for a number of fabless semiconductor companies. In particular, I have served on the technology advisory boards for two processor design companies: BOPS, Inc., where I chaired the board, and Siroyan Ltd., where I served in a similar role for three networking chip companies—Intellon, Inc., Comsilica, Inc., and Entridia, Inc.—and one 3D game accelerator company, Ageria, Inc.

20. I have also served as a technology advisor to Motorola and to several venture capital funds in the United States and Europe. Currently, I am a director at Turtle Beach Corporation, providing guidance in its development of premium audio peripheral devices and user input devices for PCs, mobile devices, and gaming consoles.

21. From 1997 through 2002, I held a variety of executive positions at a publicly held fabless semiconductor company originally called S3, Inc. and later called Sonicblue Inc. I held the positions of Chief Technology Officer, Vice President of System Integration Products, Senior Vice President of Business Development, and Director of Technology, among others. During my time at Sonicblue we launched dozens of PC products and more than 30 new consumer

electronics products including a variety of products incorporating wired and wireless communications technologies.

III. MATERIALS REVIEWED

22. The opinions contained in this declaration are based on the documents I reviewed, my professional judgment, as well as my education, experience, and knowledge regarding electrical engineering, computer science, and communications technologies.

23. In forming my opinions expressed in this declaration, I reviewed the following materials:

- the U.S. 11,087,272 (“’272 patent”) (Ex. 1001);
- the file history of the ’272 patent (U.S. Patent Application 15/471,722) (“’272Pat. File History”) (Ex. 1002);
- U.S. Patent Application Publication 2017/0178060 (“Schwartz”) (Ex. 1005);
- U.S. Provisional Application 62/269,899 (“Schwartz Provisional”) (Ex. 1006);
- U.S. Patent Application Publication 2015/0262116 (“Katircioglu”) (Ex. 1007);
- U.S. Patent Application Publication 2017/0193430 (“Avegliano ’430”) (Ex. 1008);

- U.S. Patent Application Publication 2015/0235157 (“Avegliano ’157”) (Ex. 1009);
- U.S. Patent Application Publication 2015/0055855 (“Rodriguez”) (Ex. 1010);
- U.S. Patent Application Publication 2016/0171707 (“Schwartz ’707”) (Ex. 1011);
- Declaration of June Munford (Ex. 1012);
- “A Statistical Model for General Contextual Object Recognition,” article by Peter Carbonetto, et al. (“Carbonetto”) (Ex. 1013);
- “‘Hello! My name is... Buffy’ – Automatic Naming of Characters in TV Video,” article by Mark Everingham, et al. (“Everingham”) (Ex. 1014);
- “Recognizing Products: A Per-Exemplar Multi-Label Image Classification Approach,” article by Marian George, et al. (“George”) (Ex. 1015);
- “Product Counting Using Images with Application to Robot-based Retail Stock Assessment,” article by Nishant Kejriwal, et al. (“Kejriwal”) (Ex. 1016);

- “Remote Retail Monitoring and Stock Assessment Using Mobile Robots,” article by Swagat Kumar, et al. (“Kumar”) (Ex. 1017);
- “Recognizing Groceries in situ Using in vitro Training Data,” article by Michele Merler, et al. (“Merler”) (Ex. 1018);
- “Ontological Supervision for Fine Grained Classification of Street View Storefronts,” article by Yair Movshovitz-Attias, et al. (“Movshovitz-Attias”) (Ex. 1019);
- “Toward Retail Product Recognition on Grocery Shelves,” article by Gül Varol, et al. (“Varol”) (Ex. 1020);
- “Searching the Web with Mobile Images for Location Recognition,” article by Tom Yeh, et al. (“Yeh”) (Ex. 1021);
- “Machine Learning,” textbook by Tom Mitchell (“Mitchell”) (Ex. 1022);
- U.S. Provisional Application 62/314,785 (“’785 Provisional”) (Ex. 1027);
- U.S. Provisional Application 62/427,509 (“’509 Provisional”) (Ex. 1028);

and any other material I refer to in this declaration in support of my opinions.

24. All of the opinions contained in this declaration are based on the documents I reviewed and my knowledge and professional judgment. My opinions have also been guided by my appreciation of how a person of ordinary skill in the art would have understood the claims and the specification of the '272 patent at the time of the alleged invention, which I have been asked to initially consider as no earlier than March 29, 2016 (the filing date of the earliest provisional patent application 62/314,785). '272 *patent*, Cover.² My opinions reflect how one of ordinary skill in the art would have understood the '272 patent, the prior art to the patent, and the state of the art at the time of the alleged invention.

25. Based on my experience and expertise, it is my opinion that certain references disclose and/or suggest, alone or in combination, all the features recited in claims 1–21 (“Challenged Claims”) of the '272 patent, as I explain in detail below.

² As I explain further below, by my review, the claims of the '272 patent recite subject matter that is not disclosed in its provisional patent applications 62/314,785 or 62/427,509. Counsel has informed me that this difference only affords the '272 patent the benefit of its March 28, 2017 filing date. My opinions and understanding of the knowledge of a person of ordinary skill in the art do not change with respect to these different priority dates.

26. My opinions are based on my years of education, research and experience, as well as my investigation and study of relevant materials. In forming my opinions, I have considered the materials I identify in this declaration and those listed in the Exhibit List at the end of this declaration.

27. I may rely upon these materials and/or additional materials to respond to arguments raised by the Patent Owner. I may also consider additional documents and information in forming any necessary opinions — including documents that may not yet have been provided to me.

28. My analysis of the materials produced in this matter is ongoing and I will continue to review any new material as it is provided. This declaration represents only those opinions I have formed to date.

IV. LEVEL OF ORDINARY SKILL IN THE ART

29. The '272 Patent identifies the relevant field of art as follows: “The present disclosure relates generally to a multiple camera sensor suite capable of accurately monitoring retail or warehouse product inventory without needing an initial planogram” and an embodiment in which “the multiple camera sensor suite can be mounted on an autonomous robot and include onboard processing to provide near real time product tracking.” *'272 patent*, 1:17-23. Based on that description of the field of art, my experience, and the information I have reviewed for purposes of this proceeding, I believe a person of ordinary skill in the art (“POSITA”) in

March 2016 would have had Bachelor's degree in electrical engineering, computer engineering, computer science, or a related field, and two or more years of work experience or research related to computer vision, image analytics, machine learning, artificial intelligence or automated object recognition methods and systems, or equivalent training and work experience in the field. Formal education can substitute for work experience and relevant work experience could substitute for formal education. My level of skill in the art as of March 2016 was at least that of a POSITA. I am qualified to provide opinions concerning what a POSITA would have known and understood at that time, and my analysis and conclusions herein are from the perspective of a POSITA as of that date.

V. RELEVANT LEGAL STANDARDS

30. I am not an attorney and offer no legal opinions, but in the course of my work, I have had experience studying and analyzing patents and patent claims from the perspective of a person skilled in the art.

31. In expressing my opinions and considering the subject matter of the claims of the '272 Patent, I am relying upon certain basic legal principles that have been explained to me.

32. One principle relates to the correct body of law; by my review, the '272 patent was filed on March 28, 2017. Counsel has informed me that the '272 patent is thus subject to the body of patent law governed by the America Invents Act

(“AIA”) enacted and taking effect for patent applications filed on or after March 16, 2013. Thus, I understand the ’272 patent is subject to AIA-governed patent law.

33. I understand that for an invention claimed in a patent to be found patentable, it must be, among other things, new and not obvious in view of what was known before the application for the patent on the invention was filed..

34. I understand the information that is used to evaluate whether an invention is new and not obvious is generally referred to as “prior art” and generally includes patents and printed publications (e.g., books, journal publications, articles on websites, product manuals, etc.).

35. I understand that in this proceeding the Petitioner has the burden of proving that the claims of the ’272 Patent are anticipated by or obvious in view of the prior art by a preponderance of the evidence. I understand that “a preponderance of the evidence” is evidence sufficient to show that a fact is more likely true than it is not.

36. I understand that terms appearing in the patent claims are to be interpreted according to their “ordinary and customary meaning” in an IPR proceeding. In determining the ordinary and customary meaning, the words of a claim are first given their plain meaning as they would have been understood by a POSITA at the time of the alleged invention, in light of the specification and file history. I understand that treatises and dictionaries may be consulted, albeit under

limited circumstances, to determine the meaning attributed by a POSITA at the time of the alleged invention. I have followed this approach in my analysis and have applied the ordinary and customary meaning of those terms throughout my analysis in this Declaration.

37. I understand that in this proceeding, the information that may be evaluated is limited to patents and printed publications. My analysis below compares the claims to patents and printed publications that are prior art to the claims.

38. I understand that there are two ways in which prior art may render a patent claim unpatentable. First, the prior art can be shown to “anticipate” the claim. Second, the prior art can be shown to have made the claim “obvious” to a person of ordinary skill in the art. I understand that a claimed invention is not patentable if it would have been obvious to a person of ordinary skill in the field of the invention at the time the invention was made. In this declaration I focus on “obviousness.” My understanding of the legal standard for “obviousness” is set forth below.

39. I understand that the obviousness standard is defined in the patent statute (AIA 35 U.S.C. § 103) as follows:

A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention

pertains. Patentability shall not be negated by the manner in which the invention was made.

40. I understand that the following standards govern the determination of whether a claim in a patent is obvious. I have applied these standards in my evaluation of whether the claims of the '272 patent would have been considered obvious as of the priority date of the patent.

41. I understand that to find a claim in a patent obvious, one must make certain findings regarding the claimed invention and the prior art. Specifically, I understand that the obviousness question requires consideration of four factors (although not necessarily in the following order):

- The scope and content of the prior art;
- The differences between the prior art and the claims at issue;
- The knowledge of a person of ordinary skill in the pertinent art; and
- Whatever objective factors indicating obviousness or non-obviousness may be present in any particular case.

42. I understand that the obviousness analysis cannot be confined by a formalistic conception of the words “teaching, suggestion, and motivation.” I understand that in 2007, the Supreme Court issued its decision in *KSR Int’l Co. v. Teleflex, Inc.*, 550 U.S. 398 (2007), where the Court rejected the previous requirement of a “teaching, suggestion, or motivation to combine” known elements

of prior art for purposes of an obviousness analysis as a precondition for finding obviousness. It is my understanding that *KSR* confirms that any motivation that would have been known to a person of skill in the art, including common sense, or any motivation derived from the nature of the problem to be solved, is sufficient to explain why references would have been combined.

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

44. I understand that a person of ordinary skill in the art attempting to solve a problem will not be led only to those elements of prior art designed to solve the same problem. I understand that under the *KSR* standard, steps suggested by

common sense are important and should be considered. Common sense teaches that familiar items may have obvious uses beyond their primary purposes, and in many cases a person of ordinary skill will be able to fit the teachings of multiple patents together like pieces of a puzzle. As such, the prior art considered can be directed to any need or problem known in the field of endeavor as of the priority date of the patent and can provide a reason for combining the elements of the prior art in the manner claimed. In other words, the prior art does not need to be directed towards solving the same problem that is addressed in the patent. Further, the individual prior art references themselves need not all be directed towards solving the same problem.

45. In addition, I understand that the obviousness inquiry should not be done in hindsight, but must be done using the perspective of a person of ordinary skill in the relevant art as of the effective filing date of the patent claim.

46. I understand the objective factors indicating obviousness or non-obviousness may include the following: (1) commercial success of products covered by the patent claims; (2) a long-felt need for the invention; (3) failed attempts by others to make the invention; (4) copying of the invention by others in the field; (5) unexpected results achieved by the invention; (6) praise of the invention by those in the field; (7) the taking of licenses under the patent by others; (8) expressions of surprise by experts and those skilled in the art at the making of the invention; and (9) the patentee proceeded contrary to the accepted wisdom of the prior art. I also

understand that any of this evidence must be specifically connected to the invention rather than being associated with the prior art or with marketing or other efforts to promote an invention. I am not presently aware of any evidence of “objective factors” suggesting the claims are not obvious, and I reserve my right to address any such evidence if it is identified in the future.

47. I understand that the obviousness analysis need not seek out precise teachings directed to the specific subject matter of the challenged claim, but instead can take account of the “ordinary innovation” and experimentation that does no more than yield predictable results, which are inferences and creative steps that a POSITA would employ.

48. I understand that sometimes it can be necessary to look to interrelated teachings of multiple patents; the effects of demands known to the design community or present in the marketplace; and the background knowledge possessed by a person having ordinary skill in the art. I understand that all these issues may be considered to determine whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue.

49. I understand that obviousness does not require that the features of a secondary reference be bodily incorporated into the structure of the primary reference. Rather, the test is what the combined teachings of those references would have suggested to a POSITA. The disclosures of the prior art references need not be

physically combinable. Combining the teachings of the references should be the focus of the analysis.

50. I understand that an invention that might be considered an obvious variation or modification of the prior art may be considered non-obvious if one or more prior art references discourages or lead away from the line of inquiry disclosed in the reference(s). A reference does not “teach away” from an invention simply because the reference suggests that another embodiment of the invention is better or preferred. My understanding of the doctrine of teaching away requires a clear indication that the combination should not be attempted (e.g., because it would not work or explicit statements saying the combination should not be made).

51. I understand that a person of ordinary skill is also a person of ordinary creativity.

52. I further understand that in many fields, there may be little discussion of obvious techniques or combination, and it often may be the case that market demand, rather than scientific literature or knowledge, will drive design trends. When there is such a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within their technical grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was obvious to

try might show that it was obvious. The fact that a particular combination of prior art elements was “obvious to try” may indicate that the combination was obvious even if no one attempted the combination. If the combination was obvious to try (regardless of whether it was actually tried) or leads to anticipated success, then it is likely the result of ordinary skill and common sense rather than innovation.

53. I understand that a petition must identify, for any terms with means-plus-function limitations, the specific portions of the specification that describe the structure, material, or acts corresponding to each claimed function. I also understand a rebuttable presumption exists that without use of the term, “means,” a claim term is not interpreted as means-plus-function, but that other nonce words like “component” have been found to satisfy the rebuttable presumption and invoke a means-plus-function claim construction of such a claim term.

54. I also understand that a petitioner seeking to demonstrate that the prior art discloses a means-plus-function limitation must prove that the corresponding structure—or an equivalent—was present in the prior art and performing the same claimed function.

55. I understand that proving structural equivalence does not require a component-by-component analysis, but rather that it is sufficient to show structure, materials, or acts that perform the claimed function in substantially the same way to

achieve substantially the same result because the claim limitation is the overall structure corresponding to the claimed function and not individual components.

56. I understand that the standard for determining whether the prior art discloses the corresponding structure or an equivalent is whether there are insubstantial differences, which means that if the assertedly equivalent structure performs the claimed function in substantially the same way to achieve substantially the same result as the corresponding structure described in the specification, it is equivalent.

57. I understand that for a claim to be valid, its scope must be definite, and that improperly introducing a term can lead to an unascertainable scope of a term, rendering it indefinite. I understand that any claim element first addressed in the claims by the article, “a” may then be identified later in the claims by the articles “such” or “the.” I understand that when referred to by the articles “such” or “the,” the antecedent basis for the referenced object is the object previously referred to by the article “a.” I understand that failing to follow this convention may lead to ambiguity in the identification of the particular claim element in question, leading to an indefinite scope of the claim’s term. For purposes of this proceeding, I identified and provided my best understanding of any such terms despite any such unascertainable scope.

VI. SUMMARY OF OPINIONS

58. For the reasons I explain below, it is my opinion that claims 1–21 of the '272 patent are disclosed in and rendered obvious by the prior art.

VII. BRIEF DESCRIPTION OF THE TECHNOLOGY

A. Background of the Technology

59. As I explain more fully below, the Schwartz, Katircioglu, Avegliano '430, Avegliano '157, and Rodriguez prior art references generally relate to image capture and product recognition technology using machine learning.

60. The '272 patent at issue in this matter is related to capturing and processing images of products and their affiliated shelf labels to build a training data set correlating the two, which is used to train a product classifier for future product recognition. '272 *patent*, Fig. 6, 1:65-2:3, 7:27-56, 8:1-41, 10:65-12:60.

61. In order to provide context for the '272 patent and the state of the relevant technology at the time it was filed, I provide the following high-level review and examples of known, machine learning-assisted image capture and product recognition technology preceding March 29, 2016—the earliest listed priority date of the '272 patent. '272 *patent*, Cover.

1. *Machine Learning* by Tom Mitchell (“Mitchell”)

62. Mitchell (Exhibit 1022) discloses that machine learning “encompass[es] computer programs that improve from experience.” *Mitchell*, 4. For example, a

computer program that learns to play checkers could improve its ability to win through experience gained by playing checkers against itself. *Id.*, 2. Another example of machine learning includes a database system in which users update data entries, and the system improves its performance at answering inquiries by using these database updates as training data. *Id.*, 4.

63. According to Mitchell, crafting a well-defined learning problem requires identifying the following three aspects: the class of tasks to be completed, the source of experience, and the measure of performance to be improved. *Id.*, 2–3. To learn the target concept, the learner, or classifier, is given a set of training examples, which may contain errors or noise. *Id.*, 23, 28, 175. In view of the set of training data, the learner must estimate the target concept value of the testing data. *Id.*, 23. Generally, machine learning is “most reliable when the training examples follow a distribution similar to that of future test examples.” *Id.* at 6.

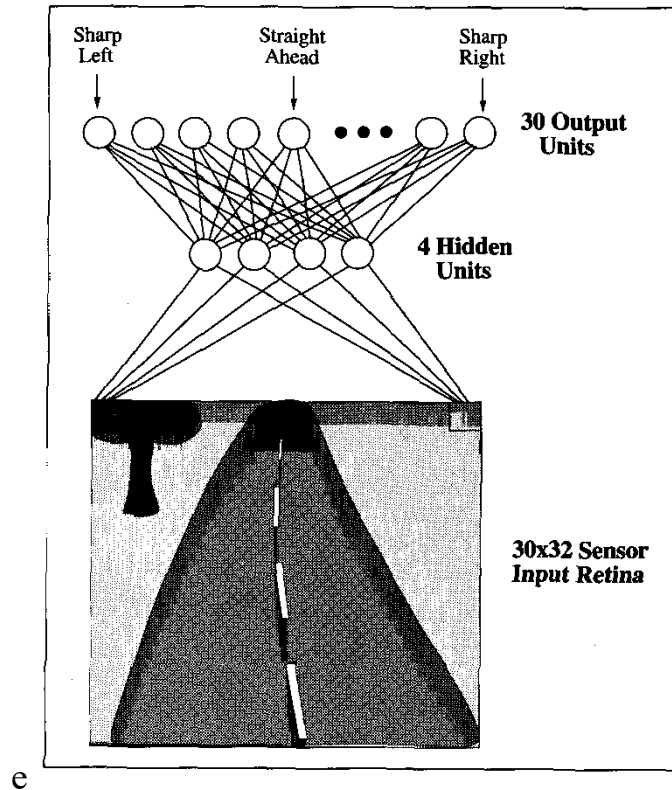
64. Mitchell discloses that neural networks “provide a general, practical method for learning real-valued, discrete-valued, and vector-valued functions from examples.” *Id.*, 81. Inspired by the complicated system of interconnected neurons in people’s brains, neural networks comprise a densely interconnected collection of simple units, where each unit takes real-valued inputs and generates a single real-valued output. *Id.*, 82. Neural network learning can be especially useful when the training data contains noisy, complex sensor data, including inputs from cameras

and microphones. *Id.*, 83. For instance, by inputting images of people as training data, neural networks can conduct facial recognition with high frequency. *Id.*, 112.

65. Neural networks are built of networks of “perceptrons,” each of which takes a vector of real-valued inputs, calculates a linear combination of those inputs (multiplying each input by a coefficient, also called a weight, and summing the results). *Id.*, 86. The perceptron then outputs a result based on the result of the linear combination. *Id.* Mitchell provides the following formula to show these inputs and outputs:

$$o(x_1, \dots, x_n) = \begin{cases} 1 & \text{if } w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n > 0 \\ -1 & \text{otherwise} \end{cases}$$

Id. Perceptrons are built into a layered network, which includes an input layer, a number of “hidden layers,” and an output layer, as shown in Mitchell’s Figure 4.1, which I reproduced below:



e

Id. at 84.

66. Mitchell discloses at least one well-known algorithm for training neural networks incorporating “backpropagation.” *Id.*, 85. In one implementation, backpropagation calculates the error rate for the data in the training examples given the current coefficients in the perceptrons. *Id.*, 88-92. Backpropagation then incrementally reduces that error by using the derivative (or an approximation of the derivative) of the error function, and adjusting the weights accordingly. *Id.*, 97-100.

2. *A Statistical Model for General Contextual Object Recognition* by Peter Carbonetto, et al.

67. Carbonetto (Ex. 1013) discloses a “spatially consistent probabilistic mapping between continuous image feature vectors and the supplied word tokens.”

Carbonetto, 1. Carbonetto recognized that manually labeling images for training a machine learning system “is both time-consuming and problematic.” *Id.*, 2. To address that issue, Carbonetto takes advantage of images that are already labeled. *Id.* For example, Carbonetto noted that there are “thousands of Corel images with keywords, museum images with meta data, news photos with captions, and Internet photo stock agencies....” *Id.* Carbonetto’s model learns word-to-region associations and object relations. *Id.*, 1. As seen in Figure 1 that I reproduced below, the data contains images and associated text, already provided by captions for the images.



Id., 2–3. The images contain blobs, which are features that describe an object. *Id.*, 2. The training data includes only the text associated with the entire images. *Id.* These captions were restricted to English nouns, such as “bear” and “floor.” *Id.* Annotations predicted by the model on the training and testing sets are shown in Figure 6, which I reproduced below.

training images



training images



test images



test images



Id., 11–12. Carbonetto proposes an algorithm for approximating the parameters when the model is not fully observed, and the partition function is intractable. *Id.*, 3.

3. *Hello! My name is... Buffy*’ – Automatic Naming of Characters in TV Video by Mark Everingham, et al. (“Everingham”)

68. Everingham (Ex. 1014) discloses a method of automatically identifying characters in television or film by using textual annotation. *Everingham*, 1. Everingham’s proposed method was tested on two episodes of the television series *Buffy the Vampire Slayer* and an accuracy percentage was calculated. *Id.*, 8–9. Instead of manually identifying the characters on the video, Everingham used information already available to train a machine learning system. In particular, Everingham used the two sources of textual annotation: subtitles featuring timestamps and transcripts featuring the speakers’ identities, which were aligned. *Id.*, 2. In the video processing step, a frontal face detector was run on every video frame, yielding the face’s approximate location and scale. *Id.*, 3–4. Next, the facial features were located in the detected face region by training a feature classifier based on consumer photographs. *Id.*, 4–5. Descriptors extracted a representation of the face appearance. *Id.*, 5. For each face detection, a bounding box containing the character’s clothing was predicted relative to the position and scale of the face detection with the bounding box containing a color histogram computed as a descriptor of the clothing. *Id.*, 5–6. Using the subtitles and transcripts, possible speaker names were proposed. *Id.*, 6.

4. *Recognizing Products: A Per-Exemplar Multi-Label Image Classification Approach* by Marian George, et al. (“George”)

69. George (Ex. 1015) discloses an “efficient approach for per-exemplar multi-label image classification, which targets the recognition and localization of products in retail store images.” *George*, 440. George describes creating a supermarket products dataset that can be used in multi-label, fine-grained cross-dataset object recognition. *Id.*, 443. The training images were taken in ideal conditions with a white background, while the testing images were taken in real-life scenarios using a mobile phone. *Id.*, 444. The approach was tested on 27 classes in the “Food” category and on the background class, which represented shelves and price tags. *Id.* George explains that “[t]he ground truth of each test image specifies bounding boxes with a corresponding single product label for each bounding box” where a “single bounding box covers a group of instances of the same product in a test image as shown in figure 3.” *Id.* George’s approach includes multi-class ranking, fast dense pixel matching, and global optimization. *Id.*, 453. In addition to performing quantitative and qualitative evaluation on the proposed dataset, the approach’s performance was also evaluated using the GroZi-120 dataset. *Id.*, 448–49. The results showed that with each dataset, the approach “effectively recognize[d] and infer[red] the locations of the objects in a test image.” *Id.*, 451–52.

5. *Product Counting Using Images with Application to Robot-based Retail Stock Assessment* by Nishant Kejriwal, et al. (“Kejriwal”)

70. Kejriwal (Ex. 1016) discloses two methods “for obtaining product count directly from images recorded using a monocular camera mounted on a mobile robot.” *Kejriwal*, 1. Figure 1, which I reproduced below, depicts the robot’s on-board cameras capture video of the shelves containing products.

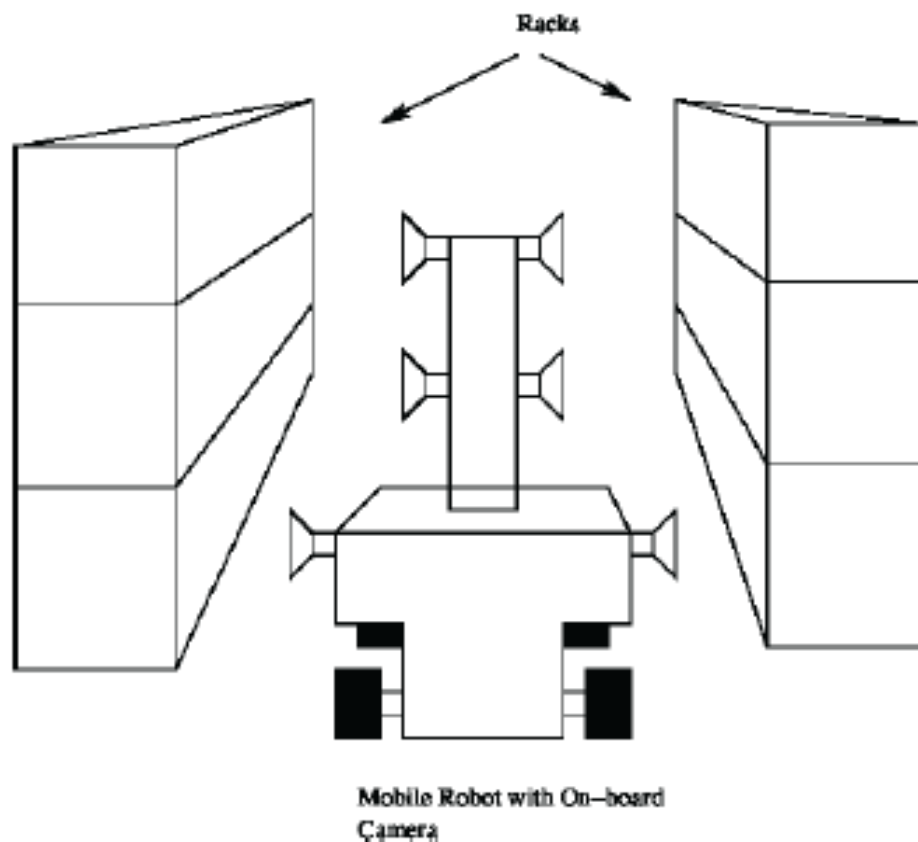


Fig. 1. Retail stock assessment using robots

Id., 2. Product recognition occurs by completing a nearest-neighbor search in the feature space using a k-d tree, where the feature space consists of the product templates’ SURF descriptors. *Id.*, 1. Assuming that the same SURF descriptors will

be repeated if there are several products of the same type, the first method comprises determining the maximum repeatability of a given product feature. *Id.*, 3. In the second method, rectangular bounding boxes are fitted around each product using SURF correspondence and homography, removed sequentially, and counted. *Id.*, 1, 3. The methods were evaluated using two datasets that were created using videos taken by a camera mounted on a mobile robot and two datasets that were created using videos taken with a mobile phone. *Id.*, 4. As noted in Table 1, these experiments demonstrated that the proposed methods reliably and efficiently observe the stock levels on shelves.

TABLE I
PERFORMANCE SUMMARY

Dataset	No. of images	No. of products	Avg. No. of descriptors / frame	Maximum Recall with Precision $\geq 90\%$	
				Method I	Method II
D1	861	22	1021	74.95	72.47
D2	315	23	891	70.17	56.00
D3	370	24	1268	98.86	97.36
D4	1135	27	669	84.79	85.59

Id., 6.

6. Remote Retail Monitoring and Stock Assessment Using Mobile Robots by Swagat Kumar, et al. (“Kumar”)

71. Kumar (Ex. 1017) discloses a system where a “robot is designed to facilitate automatic detection of Out-of-Stock (OOS) situations.” *Kumar*, 1. After

the images and videos of the shelves are collected via the robot's on-board cameras and transmitted to a remote server over the internet or a Local Area Network, they are processed, generating analytics. *Id.*, 2. Then, manufacturers and customers can obtain product-related information, including survey reports and real-time availability information, by contacting a call center or the nearest store or by using a mobile app. *Id.* As proof-of-concept and to conduct experiments using the system, the authors created a mock store with a robot. *Id.*, 2–5. They created a 3D virtual environment that mirrors the real store according to a planogram, constantly updates the position of the robot, and contains a graphical user interface. *Id.* Via the graphical interface, an operator can control or monitor the robot. *Id.*, 3. This system detects Out-of-Stock and misplaced items and shelf status. *Id.*

7. *Recognizing Groceries in situ Using in vitro Training Data* by Michele Merler, et al. (“Merler”)

72. Merler (Ex. 1018) discusses creating the GroZi-120 database of 120 grocery products featuring both *in vitro* and *in situ* images. *Merler*, 3. I reproduced Figures 1 and 2 below, which depict examples in which the training data includes *in vitro* images that were captured under ideal conditions.



Figure 1. Sample of *in vitro* images of multiple views for a product.



Figure 2. Sample of *in vitro* images for different products.

Id. The training images were pulled from the internet by searching based on “UPC codes associated with the products of a local grocery store.” *Id.* In other words, because the images were collected by searching for UPC codes, there would have been no need to manually identify each image. Conversely, the testing data consists of *in situ* images, which were taken in real-world situations in the grocery store and can be seen in Figure 3, which I reproduced below.



Figure 3. Sample of *in situ* images for the above products.

Id., 3–4. Popular object detection and recognition approaches, which include color histogram matching, matching SIFT keypoints (scale and rotation invariant descriptors representing an image), and boosted Haar-like features, were applied to the dataset. *Id.*, 4–5. Additionally, with *in vitro* images used as the training set, localization experiments were conducted by attempting to determine the products’ location in the videos. *Id.*, 5.

8. ***Ontological Supervision for Fine Grained Classification of Street View Storefronts*** by Yair Movshovitz-Attias, et al. (“Movshovitz-Attias”)

73. Movshovitz-Attias (Ex. 1019) “propose[s] a method for creating large scale labeled training data for fine grained storefront classification.” *Movshovitz-*

Attias, 1694. The training dataset was created by matching images of businesses taken from Google Street View with relevant category labels. *See id.* at 1695–96. This matching was achieved using the image’s textual information, which was transcribed using OCR, and its geographical location. *See id.* at 1696. The classifier was taught that multiple categories are possible for a business. *See id.* at 1696. Movshovitz-Attias used 1.2 million images for training the Deep Convolutional Network and 100,000 images for testing. *See id.* at 1696. Movshovitz-Attias discloses that the classifier was able to “classify these images with their high level categories and with fine grained, specialized classes.” *Id.* at 1698. Human performance studies showed that the classifier was able to achieve human level accuracy. *See id.* at 1693, 1698–99. Additionally, the article noted various challenges in storefront classification, such as the difficulty of acquiring a large amount of high-quality labeled data for training. *See id.* at 1695.

9. *Toward Retail Product Recognition on Grocery Shelves* by Gül Varol, et al. (“Varol”)

74. Varol (Ex. 1020) “addresses the problem of retail product recognition on grocery shelf images.” *Varol*, 1. Varol describes creating a database of shelf and product images taken at about 40 grocery stores. *Id.*, 2. Training images include controlled product images, while testing images include product images cropped from shelves. *Id.*, 3. The product detection module generates bounding boxes of the products on the shelves. *Id.*, 4. This module includes a cascade of boosted classifiers

working with histogram of oriented gradients features. *Id.*, 3. The authors introduced a technique to establish shelf boundaries, which can then be counted. *Id.*, 4. Using a bag of words approach, the brand recognition module only employs the logo parts of the images in the training phase. *Id.* at 5. SIFT features (vectors densely sampled from a regular grid overlain on the image) are used for the labels' shape description, and HSV values are used for the color description. *Id.* The experiments evaluating the product detection and brand recognition performances yielded "promising result[s]." *Id.*, 6.

10. *Searching the Web with Mobile Images for Location Recognition* by Tom Yeh, et al. ("Yeh")

75. Yeh (Exhibit 1021) discusses the objective to "recognize [a] location by finding web pages with an image matching the image of the location." *Yeh*, 1. In particular, Yeh discloses an approach to recognize a location from a mobile device through an image-based web search. *Id.* In this approach, an image-based web search is performed to obtain images that are both visually and textually relevant. *Id.*, 1, 3. With a mobile image of a landmark, similar images can be attained by determining the k nearest neighbors in the database using either the image-matching metric based on energy spectrum or the one based on wavelet decompositions. *Id.*, 3–4. Next, useful textual information from the web documents containing the images is extracted, and these keywords can be used to run a text-based search to uncover more related images. *Id.*, 4. Lastly, the resulting image set is filtered to collect images

close to the original query. *Id.*, 1. To test this approach, Yes describes building an image database as the bootstrap set by running a web crawler on mit.edu to gather images and their source URLs. *Id.*, 5. An image classifier was employed to classify the resulting set of 33,000 images as landmarks or non-landmarks with non-landmarks being removed from the database. *Id.* The test images include images that were taken with a camera-equipped mobile device. *Id.* For each query image, the k nearest images were acquired from the database to yield the search result. *Id.* Volunteers evaluated the search results and concluded that there was a “reasonably high” percentage of successful searches. *Id.*

B. Summary of Points on the Background of the Technology

76. A POSITA in this field and technology would have known and understood several points described in the textbook and articles above, which I further correlate to the concepts discussed in the prior art references relied upon in the grounds as I discuss more below.

77. Training data is a set of paired inputs and outputs. In other words, training data is a set of labeled objects. For example, training data may include a set of images, which have been labeled with what the objects each contain or some other descriptor to be correlated with the images. *Carbonetto*, 2-3 (labeling images based on their captions); *Movshovitz-Attias*, 1693 (labeling stores in images); *Everingham*, 3-4 (labeling characters that appear in video frames of television show); *George*, 443

(labeling images of products with appropriate categories); *Merler*, 3 (images labeled with the products they contain). These labels are sometimes referred to as “ground truth,” in that they are the objective truth as observed on the ground. *Varol*, 6; *Movshovitz-Attias*, 1697; *Carbonetto*, 7; *George*, 444; *Merler*, 5; *Everingham*, 8; *Rodriguez*, [0085], [0166]; *Schwartz*, [0003], [0056], [0114], [0131], [0134].

78. A POSITA would have understood that manually labeling training data, where a human applies the labels, can be costly, time-consuming, and complex. *Carbonetto*, 2. A POSITA, therefore, would have further understood that training data can be built automatically and/or synthetically, using some objective form of ground truth to label the training data or a reliable estimator thereof. *Mitchell*, 351 (using prior knowledge to create “synthetic training samples” for neural network to learn how to steer an autonomous vehicle); *Schwartz*, [0134]. This has been tackled in numerous ways in numerous references, and would have been well-known in the art. For example, *Carbonetto* refers to the “large quantities of data” in the form of images that already have labels, such as “thousands of Corel images with keywords, museum images with meta data, news photos with captions, and Internet photo stock agencies....” *Carbonetto*, 2. Instead of having a human identify which character was speaking in each frame of a television show, *Everingham* used screenplays and subtitles to provide this information. *Everingham*, 2-3. *Movshovitz-Attias* used

address and geographical information to identify storefronts in images. *Movshovitz-Attias*, 1695-96.

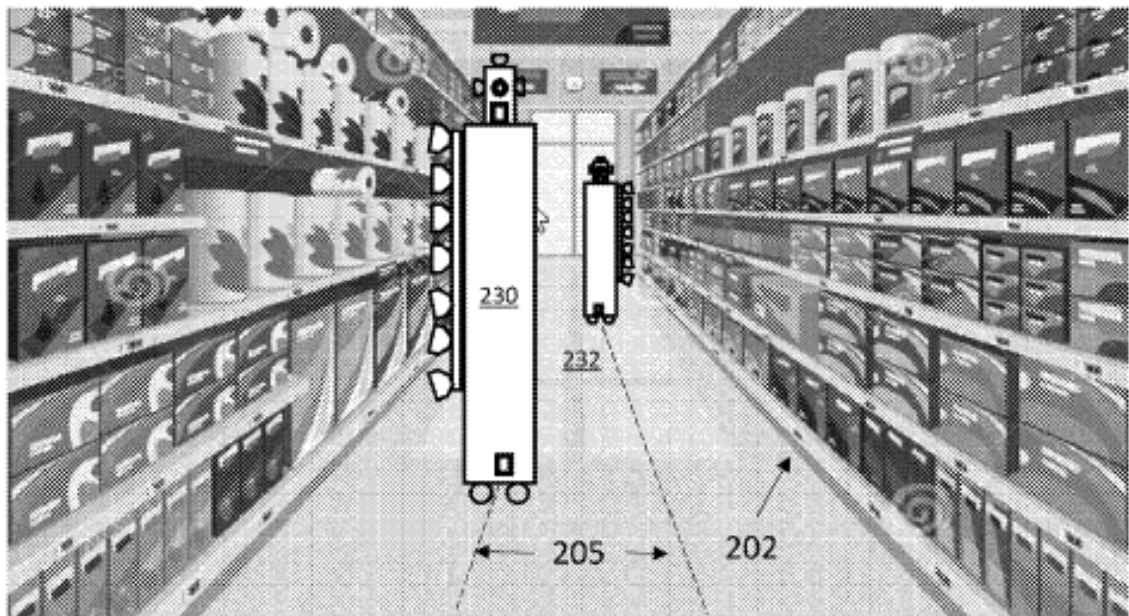
79. In the area of product classification, the reading of a barcode (such as that for a UPC) has been used to automatically identify products—as a source of objective truth for training a machine learning system—without human interaction would have been well known. For example, Merler had a “Perl script” that “automatically crawl[ed] the web looking for images of the products by using the UPC code.” *Merler*, 3. Rodriguez taught scanning the barcode of a product, while simultaneously taking images of the product, to train a machine learning system to identify that product. *Rodriguez*, [0083] (“[O]nce the barcode is identified, then all of these images are known to depict a box of Cheerios cereal. Each of the images can be tagged with label information indicating that it depicts a box of Cheerios cereal, and can be provided to a learning system as a training image.”).

80. Thus, it is my opinion in view of my review above that a POSITA would have understood and recognized several machine learning-assisted image capture and product recognition systems existed and disclosed both problems and solutions to the accuracy, efficiency, and robustness of such systems prior to the earliest listed priority date of the ’272 patent.

VIII. OVERVIEW OF THE '272 PATENT

A. Specification

81. The '272 patent is directed to a “System and Method for Locating, Identifying and Counting Items.” '272*Pat.*, Title. The patent describes time-consuming and expensive problems with manually monitoring and tracking product inventory using planograms. '272*Pat.*, 1:27-53. “Planogram” is an industry term that refers to a drawing or plan for displaying merchandise in a retail store. The '272 patent proposes monitoring and inventory management by mounting an image capture unit and processing module on an autonomous robot depicted in Figure 2, below.



'272*Pat.*, Fig. 2, 2:4-28, 3:10-23.

82. This robot-mounted image capture system stitches captured shelf label and product images together into panoramic pictures used to construct and locally store an updateable “realogram”—another industry term that refers to what is actually on the shelves, as opposed to what is planned by the planogram—mapping retail inventory within a store and across like store locations according to the following steps:

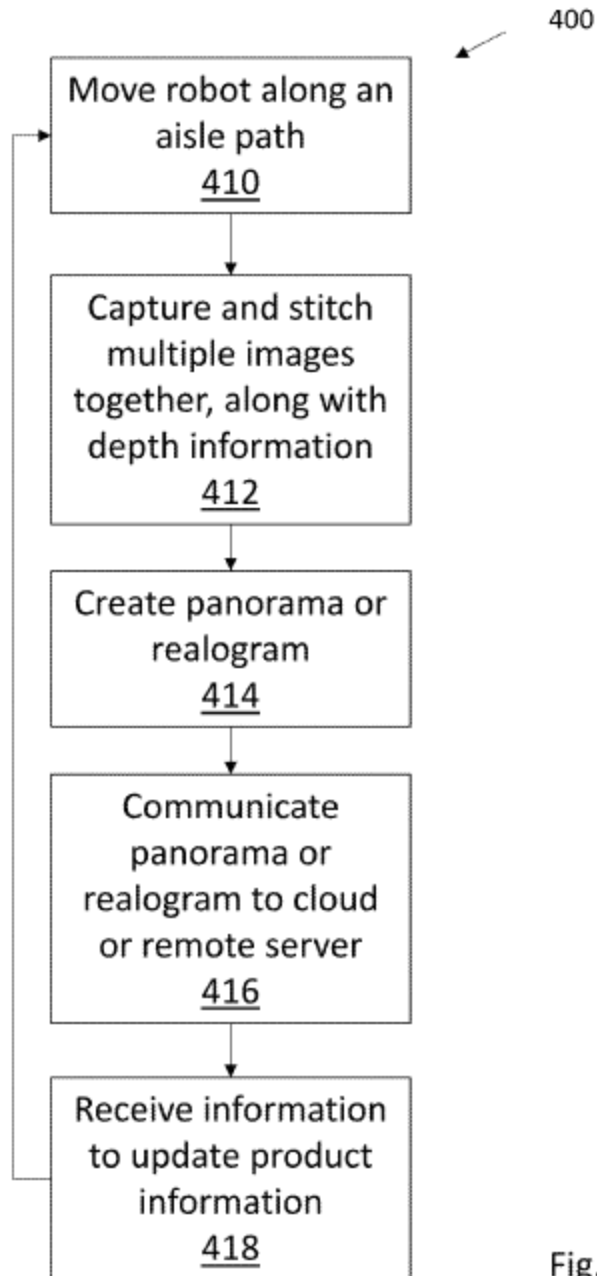
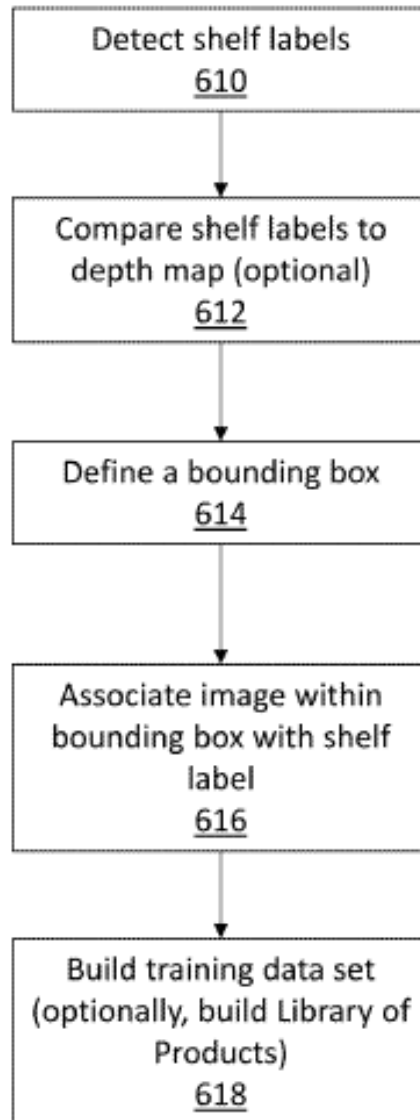


Fig. 4

'272*Pat.*, Fig. 4, 5:48-54, 7:27-56. Realograms may contain product, shelf, 3D model and location attributes that are stored and used to inform and seed subsequent realogram updates and constructions. *Id.*, 5:11-16, 6:51-61.

83. The robot-mounted system further captures and processes images of product-affiliated shelf labels with images contained within bounding boxes to build a training data set, which is used to train a product classifier according to the following steps:



Id., Fig. 6, 1:65-2:3, 7:27-56, 8:1-41, 10:65-12:60.

B. Prosecution History

84. I understand the '272 patent issued from U.S. Application 15/471,722 ("’272 Application"), and claims priority to two provisional patent applications: 62/314,785 (Ex. 1027), filed March 29, 2016, and 62/427,509 (Ex. 1028) , filed on November 29, 2016. '272 *patent*, Cover. In a first Office Action, I see that the Examiner issued a requirement for Applicant to elect a species of claims for examination. '272*Pat. File History*, 80-84. Following, I see that the Applicant elected to prosecute exclaims 1-10 and 15-26. *Id.*, 89-95.

85. By my review, over a series of Office Actions and Responses, the Examiner rejected the claims over different combinations of prior art references including Zimmerman, Zhu, Nakamura, Wicks, Schwartz, Wu, Ren, Calhoon, and Ihara. *Id.*, 105-15, 154-69, 218-32. I see that the Applicant amended the claims each time, narrowing the claim scope to overcome the rejections. Specifically, the Applicant added limitations reciting features including "moving a robot along a shelf...[with] attached cameras," training and using a product classifier, defining and associating product facing bounding boxes with shelf labels, cross-correlating product facing bounding boxes across stores by matching the content of their associated shelf labels, building a training data set based on that associated data, and building and using a product classifier from the training data set. *Id.*, 136-41, 205-12.

86. I further see that the Examiner issued a Notice of Allowance with an Examiner's Amendment to clerical aspects of the claims. *Id.*, 268-77. Reviewing the Reasons for Allowance, I see that the Examiner distinguished the closest prior art of record (Zimmerman, Ihara) over issued claims limitations 1.4-1.5 because (i) Zimmerman does not associate product facing bounding boxes that include and define each of the one or more corresponding product facings, and (ii) Ihara does not match shelf label content with other shelf label content associated with product facing boxes across stores. *Id.*, 275-77. Following, the Examiner issued the claims of the '272 patent on August 10, 2021. '272 *patent*, Cover.

87. I see and confirm that none of the prior art relied upon in the grounds that follow are included in or are related to a reference included in the '272 patent's file history.

C. Neither the '785 nor '509 Provisionals Provide Written Description Support for the '272 Patent Claims

88. When reviewing the '785 and '509 provisional patent applications preceding the '272 patent, I noticed substantive differences in the respective disclosures.

89. First, the earlier '785 Provisional does not disclose building training data based on bounding box images as claimed. Specifically, I do not see any disclosure in the '785 Provisional of "training data" (or even "training"). *See generally, '785 Provisional*. The '785 Provisional discloses an approach based on

extracting descriptors using known algorithms such as “Scale Invariant Feature Transform” (known as “SIFT”). *Id.*, 4. As I explained in my background of the technology section above, both Merler and Varol disclose SIFT keypoint extraction and matching relying upon the feature vector characteristics of a single captured image. *See* Sections VII.A.7, VII.A.9. Conversely, a POSITA would have understood training data to be a set of known mappings between inputs and outputs.

90. For example, the claims of the ’272 patent require building training data based on the “association of the shelf label content with the ... product facing bounding boxes....” ’272*Pat*, claims 1, 10, 18. In other words, the training data required by the claims of the ’272 patent is a set of known mappings between an identifier of the product (e.g., “Cheerios”) and a set of images of the product (pictures of boxes of Cheerios captured from the store shelves). Conversely, the algorithms disclosed by the ’785 Provisional use only the images themselves, and do not use the association with the shelf label content. ’785 *Provisional*, 4.

91. The ’272 patent also expressly distinguishes that product descriptor approach from the claimed approach of training classifiers using bounding box images. ’272*Pat*, 12:8-40 (“An *alternative to extracting product descriptors*³ is to use the bounding boxes as labeled categories and train classifiers on the images

³ All emphasis in this declaration is my own, unless I note otherwise.

contained in the bounding boxes.”). The latter is described in the latter ’509 Provisional (at 17:20-18:2), but not in the earlier ’785 Provisional.

92. Second, from my review, I see that each challenged claim is directed to methods (claims 1, 10) and system (claim 18) that expressly includes a **data cross-correlating step** before **building training data** with the cross-correlated data and using that training data to build and use a **product classifier**:

[1.5/10.5/18.6] cross-correlating[/cross-correlate] the one or more product facing bounding boxes and one or more other product facing bounding boxes **across** one or more stores...

[1.6/10.6/18.7] building[/build] training data based on...the cross-correlated product facing bounding boxes...

[1.7/10.7/18.8] using[/use] the training data to build a product classifier...

[1.8] using the product classifier to classify an item as the product. [10.8] using the product classifier to perform an inventory management activity related to the product [18.9] use the product classifier to perform one or more of: (1) an inventory management activity related to the product or (2) classifying an item as the product.]

93. I do not see anywhere in either of the ’785 or ’509 Provisionals that discloses the “cross-correlating” limitations 1.5/10.5/18.6. At most, I found sections

of the applications that disclose deploying a robot-implemented method throughout several stores and warehouses for purposes of building a library of products using storewide price tags and labels to an *instant* store rather than cross-correlating *product facing bounding boxes* generated in real-time *across stores* per claims 1, 10, and 18. '785 *Provisional*, 4, 6, 10 (describing deployment of the *mobile bases and method* across retail stores and warehouses and stating “the price tag or product label associated with the same product *are identical* across locations” without cross-correlating data to build more data); '509 *Provisional*, 5, 24 (e.g., Claims 6-7 describing association of the bounding box with an image provided by the image capture unit mounted on an autonomous robot to train image classifiers).

94. In my opinion, because the Provisionals fail to disclose the correct type of data component expressly disclosed in 1.5/10.5/18.6, they also fail to disclose the data built from that data and recited in 1.6-1.8/10.6-10.8/18.7-18.9.

D. The '272 Patent Is Entitled Only to a March 28, 2017, Priority Date Based on Its Originally-Filed Claims

95. Counsel has informed me that when ancestor applications do not disclose the subject matter of claims issued from a later-filed application, the claims are only afforded the priority date corresponding to the original disclosure of that subject matter. Because the Provisionals fail to disclose the subject matter recited in the '272 patent claims I identified above, it is my understanding that the '272 claims are not entitled to either provisional date, and therefore the earliest possible priority

date is the March 28, 2017 filing date of the non-provisional application. '272 Patent, Cover.

IX. OVERVIEW OF THE PRIOR ART REFERENCES

96. Below is an overview of the primary prior art references that I considered when analyzing the validity of the claims of the '272 Patent.

A. Schwartz

1. Analogous Art

97. In my opinion, Schwartz (EX1005) is analogous art to the '272 patent. Schwartz is directed toward capturing images of shelved retail products and using image recognition analysis to determine a state of inventory, and I find is therefore directed toward the same field of endeavor as the '272 patent. *Compare Schwartz*, Abstract, [0006]; *Schwartz Provisional*, Abstract, [0003]; *with 272 patent*, 1:57-64. Because Schwartz seeks to automate retail inventory management to overcome the inefficiencies of manual methods, I find that Schwartz is also reasonably pertinent to the problem faced by the '272 patent inventor. *Compare Schwartz*, [0005]; *Schwartz Provisional*, [0002]; *with '272 patent*, 1:27-53.

2. Schwartz is Entitled to the Earlier Effective Filing Date of its Provisional Application

98. I see that Schwartz was filed as U.S. Patent Application No. 15/164,633 ("Schwartz Application") on May 25, 2016, published on June 22, 2017, and claims

priority to provisional application 62/269,899 (the “Schwartz Provisional”) filed on December 18, 2015. *Schwartz*, Cover.

99. Schwartz is a published patent application that contains subject matter that, by small exception, was first disclosed in the Schwartz Provisional filed December 18, 2015. As I noted above when discussing the relation between the ’272 patent and its provisional application, Counsel has informed me that, for purposes of disclosure, a later-filed application is entitled to the priority date of an earlier-filed application if the earlier-filed application discloses the same subject matter as the later-filed application. In my opinions that follow, I provided an itemized table correlating the subject matter disclosed in the Schwartz Provisional to that disclosed in the Schwartz publication.

100. Because it is my opinion the ’272 patent does not disclose all the subject matter it claims until its original filing on March 28, 2017, that published as the ’272 patent itself and that the Schwartz publication was publicly accessible on May 25, 2016, I understand that Schwartz predates the rightful priority date of the ’272 patent and have considered it to be prior art in my following opinions.

101. To the extent Patent Owner argues that the ’272 patent should be afforded its earliest listed March 29, 2016, date, I cite the subject matter from the Schwartz Provisional filed on December 18, 2015, that still predates the subject

matter of the '272 patent. In all scenarios, it is my opinion that Schwartz first disclosed the subject matter the '272 patent claims.

3. Overview

102. As I noted above, the Schwartz Provisional and Schwartz publication share nearly the same disclosure, with the former disclosing the same—and therefore supporting—the latter. I cite both here to illustrate this aligned disclosure.

103. Schwartz is directed toward “planogram matching.” *Schwartz*, Title; *Schwartz Provisional*, Title. Schwartz discloses methods and systems of capturing store shelf images with a client device camera “stitched together into a linear panoramic image” to create a scene of products it describes as a “realogram.” *Schwartz*, Fig. 1 (115a), [0045], [0071], [0128], [0130]; *Schwartz Provisional*, Fig. 1 (115a), [0046], [0066], [0122], [0124]. Captured images include products, e.g., “breakfast cereal boxes in a retail supermarket.” *Schwartz*, [0053]; *Schwartz Provisional*, [0050].

104. An image recognition application 103 located in either the client device (103a) or recognition server 101 (103b) determines which objects are included in the images using feature extraction and database matching of object attributes including its identifier, location, metadata, or image itself. *Schwartz*, [0059], [0072], Fig. 3 (step 304), [0116]; *Schwartz Provisional*, [0048], [0067], Fig. 3 (step 304), [0111]. Specifically, the image identifier may include a “symbolic identifier”: a universal

product code (UPC), product name, brand, packaging. *Schwartz*, [0073], [0132]; *Schwartz Provisional*, [0068], [0125]. The captured images used to identify products include shelf label information, which I reproduced below:

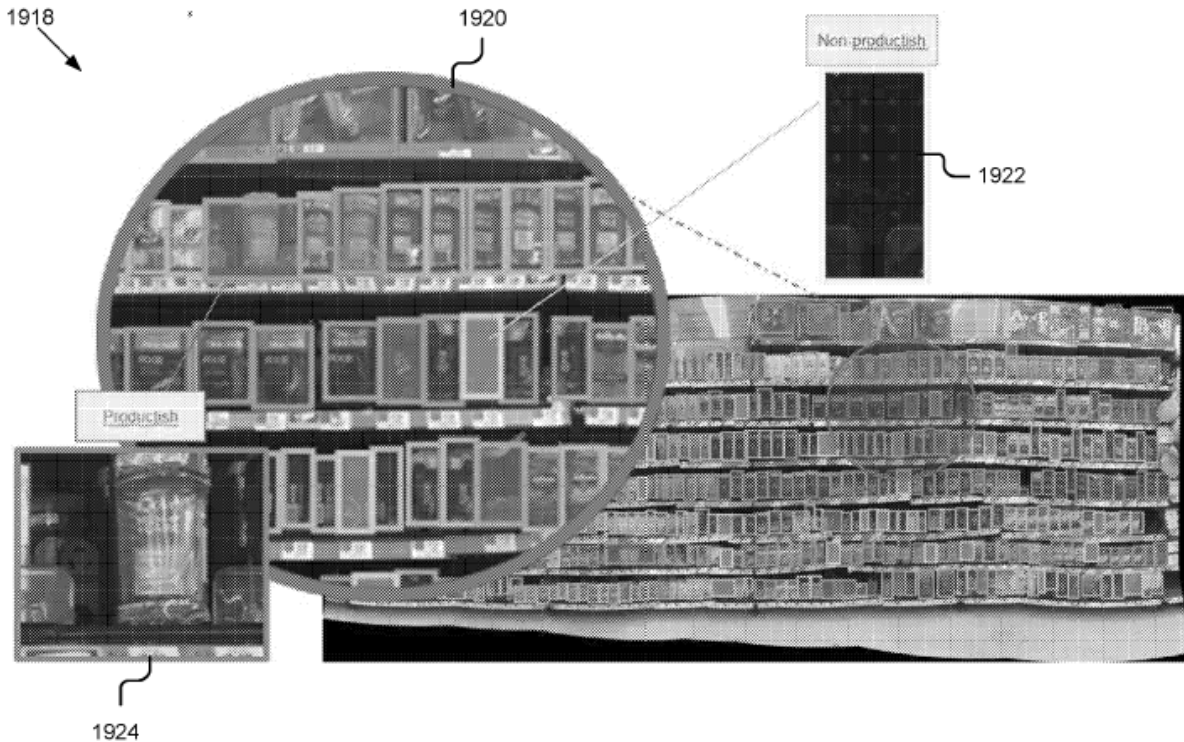


Figure 19D

Schwartz, Fig. 19D;⁴ *Schwartz Provisional*, Fig. 19D. The image attributes are collected across multiple stores and at multiple times by to generate a data stream of object information. *Schwartz*, [0080]-[0081]; *Schwartz Provisional*, [0075]-[0076].

⁴ Wherever I cite and produce a figure disclosed by both *Schwartz* and the *Schwartz Provisional*, I include only the figure to *Schwartz* to avoid unnecessarily duplicating

105. A data stream generator feeds this information to a classification module to identify a product. *Schwartz*, Fig. 3 (step 308), [0117]; *Schwartz Provisional*, Fig. 3 (step 308), [0112]. The classification module then extracts portions of the data stream based on a condition. *Schwartz*, [0082]-[0083]; *Schwartz Provisional*, [0077]-[0078]. The classification module may, e.g., identify an unindexed product and its features from an image and combine that information with information from other stores to correctly index that product as a new product (e.g., where a competitor's new product appears as "unindexed" in store A, combining the information extract from Store A with the information known about competitor's product from information captured at competitor's store B to update and index the product information at store A). *Schwartz*, [0090], [0100]; *Schwartz Provisional*, [0085], [0095].

106. Different system components conduct the matching step in different embodiments. *Schwartz*, [0086], [0090] (classification module matches recall products and determines no match of an unindexed product), [0156] (analytics module matches best fit candidates). *Schwartz Provisional*, [0081], [0085], [0141]. Different algorithms can be used to identify a product from a realogram by matching

the same figure replicated in the disclosures of both *Schwartz* and the *Schwartz Provisional*.

them to known and weighted best-fit candidates, such as product-specific heuristics and a “productish” algorithm. *Schwartz*, [0142]-[0143],[0156], [0158], [0165]-[0167]; *Schwartz Provisional*, Fig. 33 (step 3210), [0130]-[0131], [0141], [0143], [0150]-[0152]. Such algorithms infuse “real world ground truth data” to inform and improve the algorithm. *Schwartz*, [0080], [0166]; *Schwartz Provisional*, [0075], [0151].

B. Katircioglu

1. Analogous Art

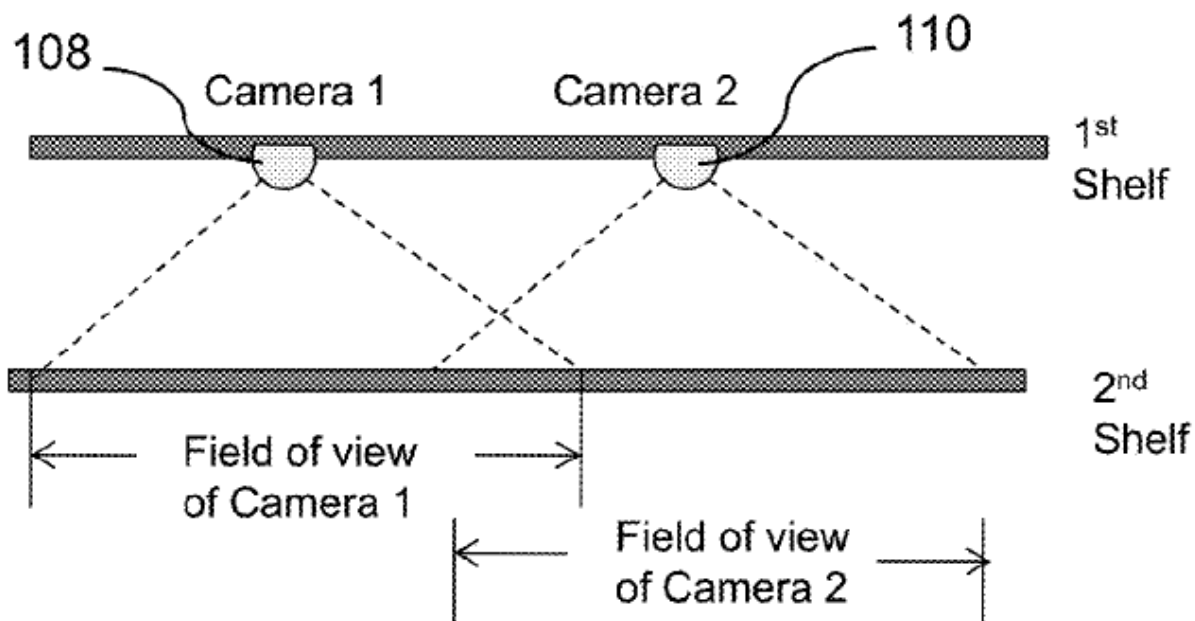
107. In my opinion, Katircioglu (EX1007) is analogous art to the '272 patent. Katircioglu is directed toward capturing images of shelved retail products and using image recognition analysis to determine a state of inventory and is therefore directed toward the same field of endeavor as the '272 patent. *Compare Katircioglu*, Abstract, [0004]; *with '272 patent*, 1:57-64. Because Katircioglu seeks to automate retail inventory management to overcome the inefficiencies of manual methods, it is also reasonably pertinent to the problem faced by the '272 patent inventor. *Compare Katircioglu*, [0003]; *with '272 patent*, 1:27-53.

2. Overview

108. Katircioglu is directed toward “machine vision technology for shelf inventory management.” *Katircioglu*, Title. Katircioglu’s system includes (i) cameras that captures an image of items on a shelf, (ii) an image-count correlation database that stores historical images of product types on a shelf, and (iii) a product

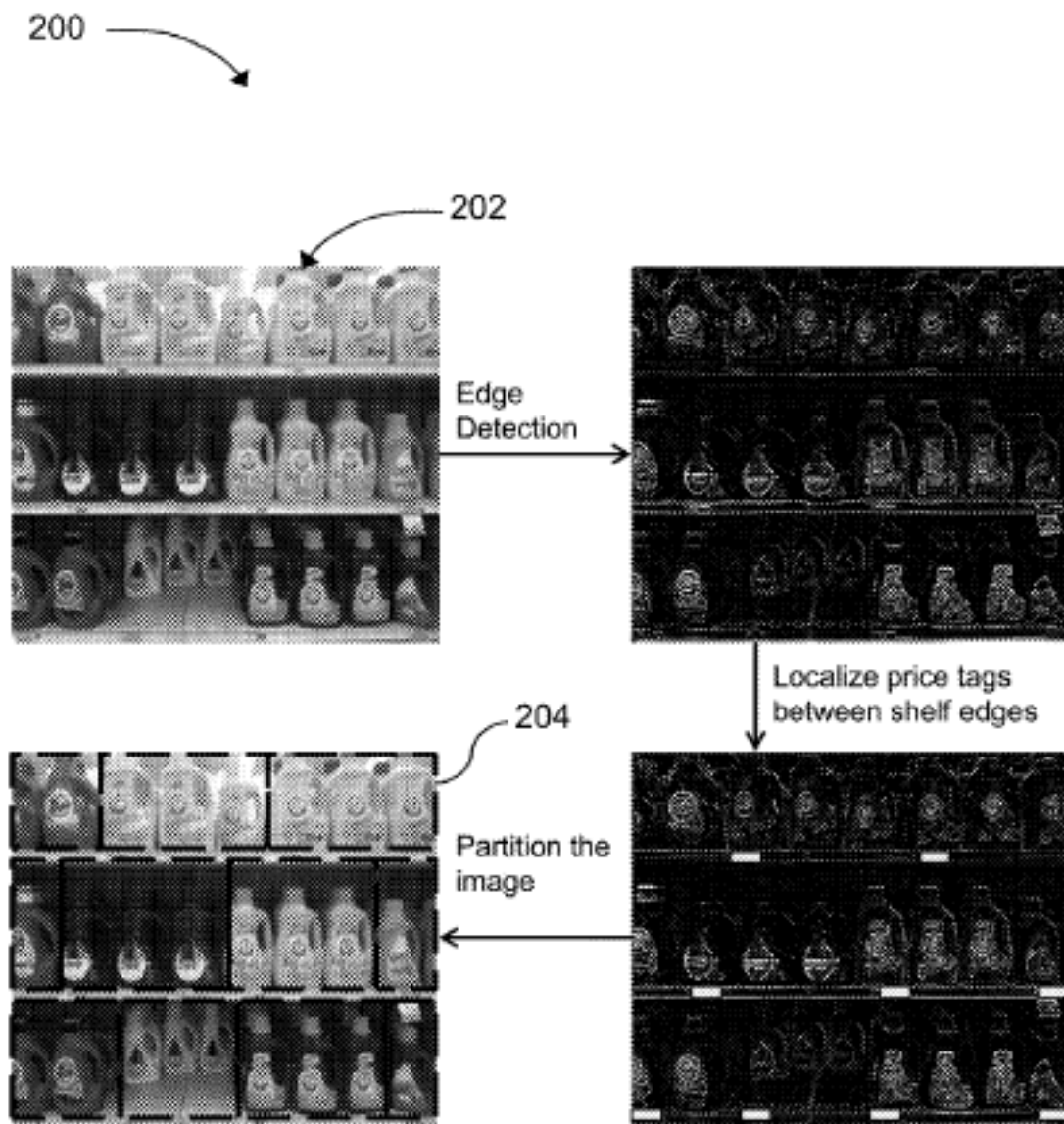
inventory database that stores a product shelf inventory count associated with each of the products. *Katircioglu*, [0004]. The cameras are arranged on shelves to capture images of products and their identifier tags—e.g., price tags—on a shelf from across an aisle as shown in the figures I reproduce below. *Id.*, [0022], [0024].





Id., Figs. 1A-B (combined).

109. The system uses edge detection to partition a captured digital image into products and their affiliated identifier tags—“patches”—using a process of segmentation shown in the figure I reproduced below. *Id.*, [0024].



Id., Fig. 2.

110. Segmented patches (also termed product inventory images), are correlated to product inventory database-stored images of product types (Fig. 7) quantified (Fig. 3A) in a matching process shown in the figures I reproduced below. *Id.*, [0025] (describing image count correlation); [0038] (describing product identification using barcodes on identifier tags).

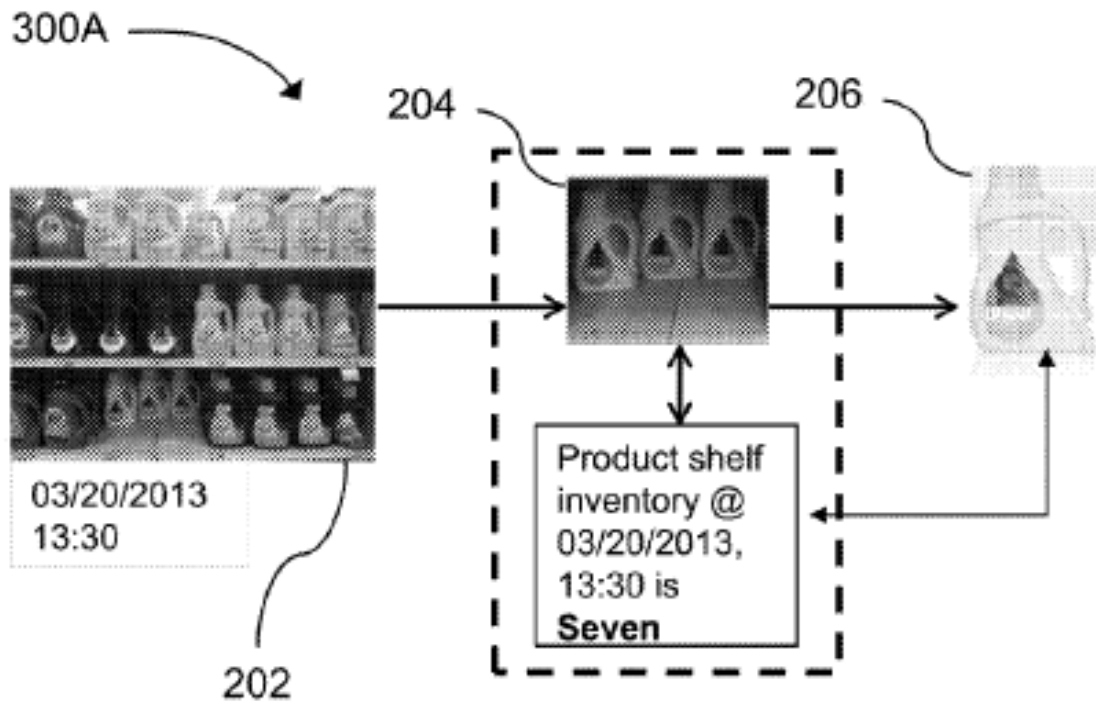
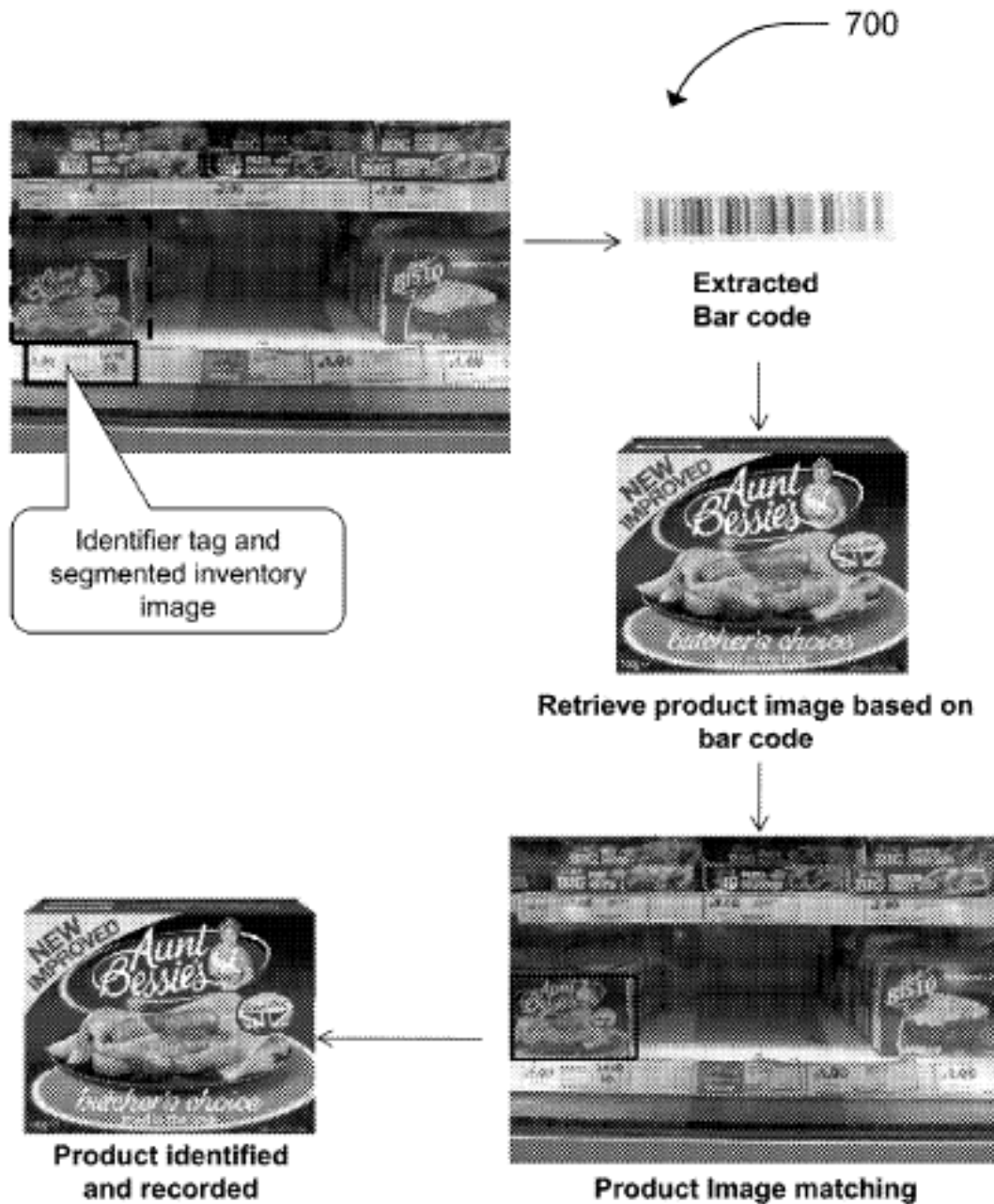


Fig. 3A

Id., Fig. 3A (depicting image count correlation).



Id., Fig. 7 (depicting product identifier tag-based identification). The system builds product image and shelf inventory data correlation pairs over time, creating an expansive volume of data processed with machine learning techniques. *Id.*, [0027]-[0028], [0040].

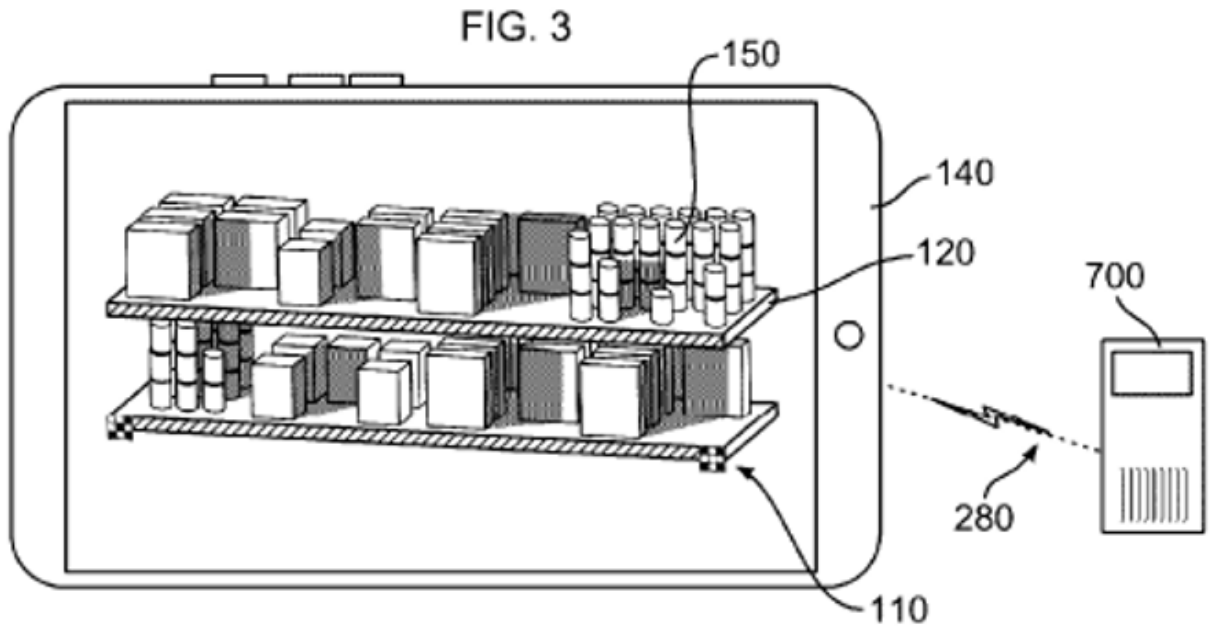
C. Avegliano '430

1. Analogous Art

111. In my opinion, Avegliano '430 (EX1008) is analogous art to the '272 patent. Avegliano '430 is directed toward capturing images of shelved retail products and using image recognition analysis to determine a state of inventory and is therefore directed toward the same field of endeavor as the '272 patent. *Compare Avegliano '430*, Abstract, [0003]; *with '272 patent*, 1:57-64. Because Avegliano '430 seeks to automate retail inventory management to overcome the inefficiencies of manual methods, it is also reasonably pertinent to the problem to be solved by the '272 patent inventor. *Compare Avegliano '430*, [0002]; *with '272 patent.*, 1:27-53.

2. Overview

112. Avegliano '430 discloses a method of “identifying spaces on shelves which are not full of objects” by “obtaining an image of the shelf...using a device...having a camera and a computer” containing “software configured to analyze the image” to identify a shelf, its storage area and any vacant areas on it. *Avegliano '430*, [0003]. In one embodiment, Avegliano '430 captures images of products 150 sitting on shelves 120 using a digital camera in a smartphone “or other portable computing device” 140 and uses an application to process the image data as shown in the figure I reproduced below. *Id.*, [0022].



Id., Fig. 3. In another embodiment, Avegliano '430 avoids the need for an “employee [to] carry out such image gathering” by flying an unmanned aerial vehicle (UAV) or drone 200 carrying cameras 220 and depth measuring devices 222 along rows of displayed product and gathering image data. *Id.*, [0034]. The UAV/drone and its orientation with respect to product rows are shown in the figures I reproduced below.

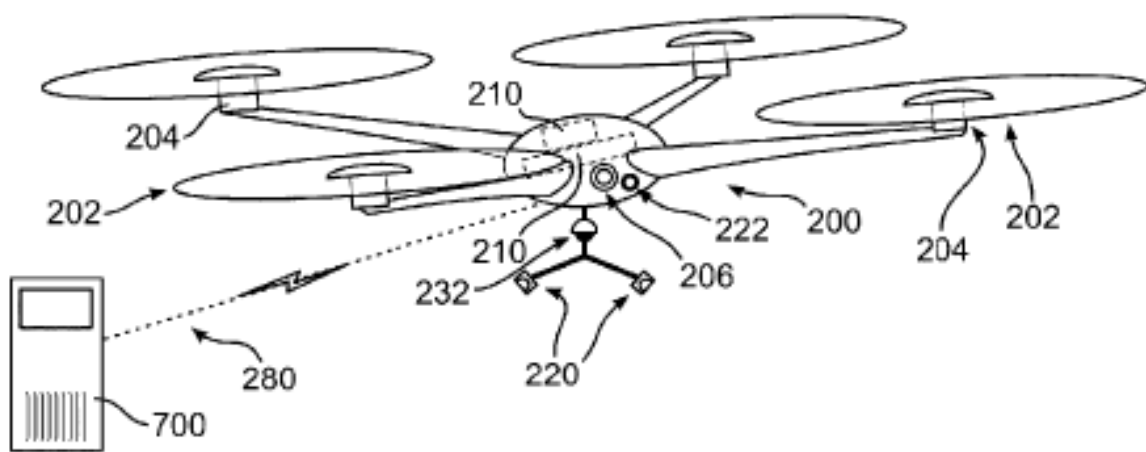
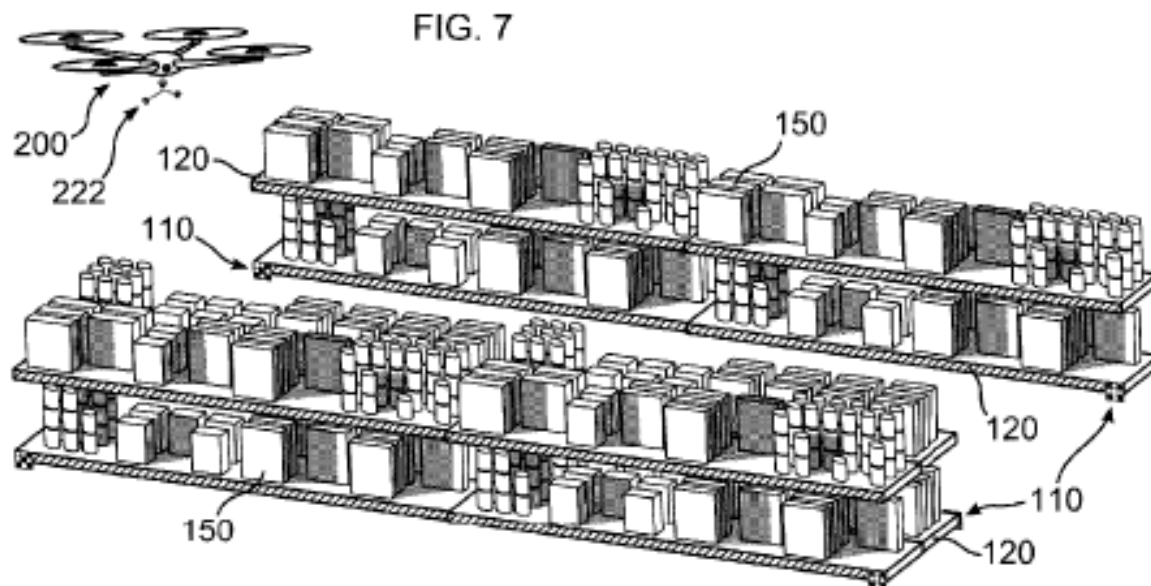


FIG. 5



Id., Figs. 5, 7.

D. Avegliano '157

1. Analogous Art

113. In my opinion, Avegliano '157 (EX1009) is analogous art to the '272 patent. Avegliano '157 is directed toward capturing images of shelved retail products and using image recognition analysis to determine a state of inventory, and is therefore directed toward the same field of endeavor as the '272 patent. *Compare Avegliano '157*, Abstract, [0016]; *with '272Pat.*, 1:57-64. Because Avegliano '157 seeks to automate retail inventory management to overcome the inefficiencies of manual methods, it is also reasonably pertinent to the problem faced by the '272 patent inventor. *Compare Avegliano '157*, [0030], [0045]; *with '272Pat.*, 1:27-53.

2. Overview

114. Avegliano '157 is directed toward “mobile device[-]based inventory management and sales trends analysis in a retail environment.” *Avegliano '157*, Title. Avegliano '157 discloses a retail store server receiving a product’s availability data. *Id.*, Abstract, [0016]. With an image of a product on the shelf taken by a mobile device, products are identified “using tags attached to the shelves.” *Id.*, Abstract, [0016], Figs. 3-4. The tags can be identified “using tag template information and information regarding a distribution of tags on the retail store shelves that is read from a tag database.” *Id.*, [0018], Fig. 3. The identified tags and products can be mapped to a shelf “to define which products are displayed in a rack of shelves.” *Id.*, [0060]. Thus, the products can be identified in the image by “reading product templates related to the identified tags,” and “a list of identified products and their quantities” can be transmitted to the server. *Id.*, [0062]. A first image processing application identifies the tags, while a second image processing application identifies the products in the image. *Id.*, [0019]. Based on product dimensions, the results of the identified products can be filtered to “eliminate false positives.” *Id.*, [0020], Fig. 3.

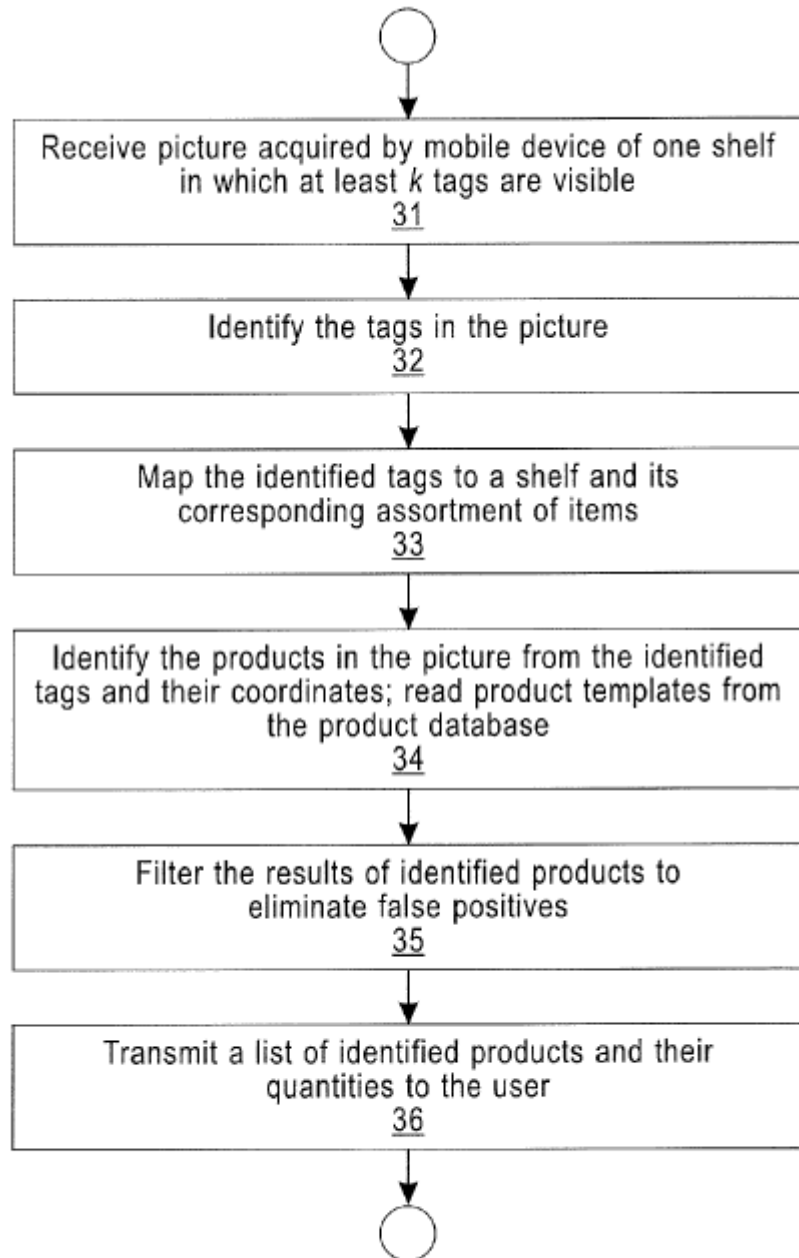


FIG. 3

Shelf rack S1

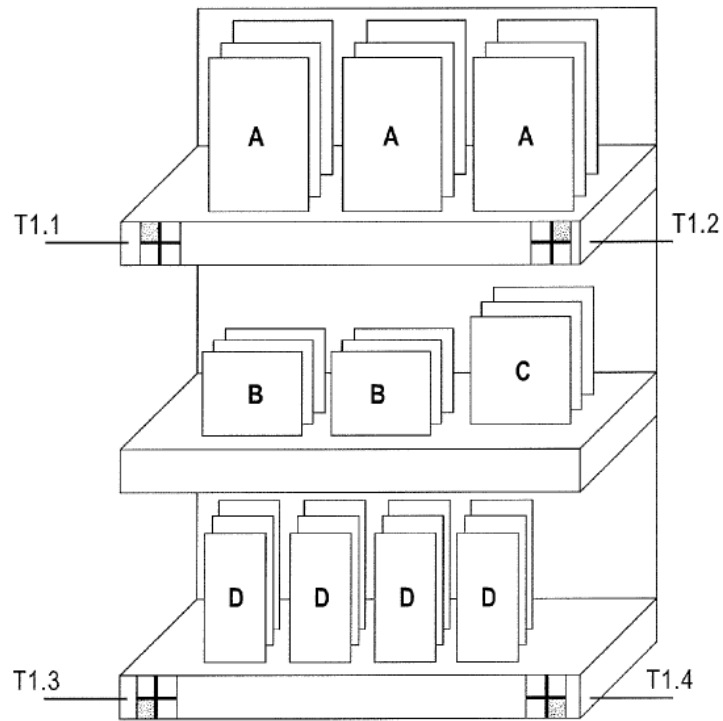


FIG. 4

Id., Fig. 3-4.

115. The images to be processed are collected by applications on mobile devices, used by both customers and employees (“stockists”). *Id.*, [0046]. For example, these images can be captured as a “stockiest courses through a store....” *Id.*, [0059]. Based on these images, the product’s sales velocity and sales trends (e.g., product turnover and demand forecasts) are calculated and transmitted “to one or more third parties’ systems in a supply chain of said retail store, where products and their locations on retail store shelves have been catalogued in a product database.”

Id., Abstract, [0016], [0022]-[0024]. A warning message is be triggered “if the retail store shelf is empty.” *Id.*, [0017]. Avegliano ’157 also discloses “a non-transitory program storage device readable by a computer, tangibly embodying a program of instructions executed by the computer to perform the method steps for managing inventory and sales trends in a retail environment using mobile devices.” *Id.*, [0030].

E. Rodriguez

1. Analogous Art

116. In my opinion, Rodriguez (EX1010) is analogous art to the ’272 patent. Rodriguez is directed toward capturing images of shelved retail products and using image recognition analysis to determine a state of inventory and is therefore directed toward the same field of endeavor as the ’272 patent. *Compare Rodriguez*, Abstract, [0009]; *with ’272Pat.*, 1:57-64. Because Rodriguez seeks to automate retail inventory management to overcome the inefficiencies of manual methods, it is also reasonably pertinent to the problem faced by the ’272 patent inventor. *Compare Rodriguez*, [0009]; *with ’272Pat.*, 1:27-53.

2. Overview

117. Rodriguez is directed toward “improving object recognition, e.g., at supermarket checkouts—such as by machine learning or image fingerprinting.” *Rodriguez*, [0009]. It emphasizes that “[i]n the prior art, training of machine-learning systems typically proceed[ed] with reference to . . . high quality imagery” even though the testing data comprised of images of “relatively inferior quality.” *Id.*,

[0076]. According to Rodriguez, these prior art systems, which were trained exclusively with high-quality images, performed poorly when tested with low-quality images. *Id.* Therefore, Rodriguez discloses capturing images of retail products in a real-world retail setting, such as at point-of-sale checkout stations or by customers during shopping and using these images as training data to build a product classifier. *Id.*, [0009], [0160]-[0162], [0186].

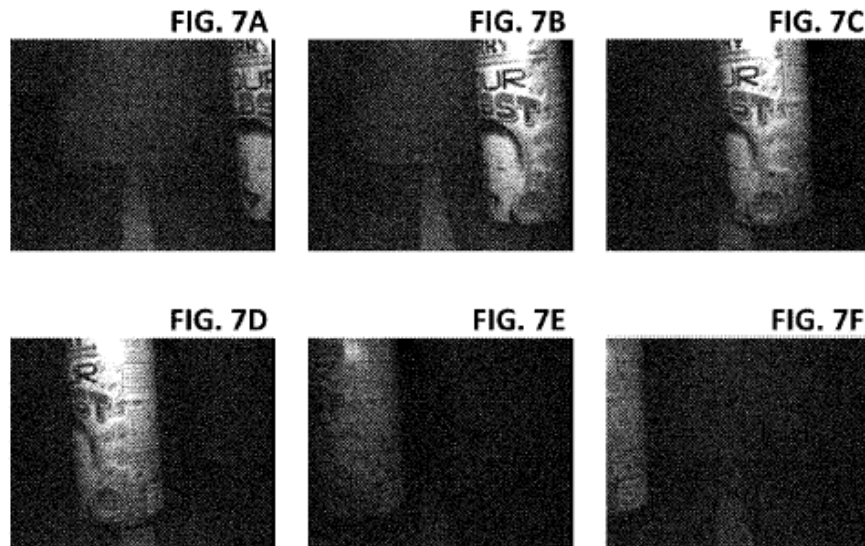
118. At checkout, the point-of-sale product scanner captures different images of the same item with varying angles, focus, and lighting, yielding “degraded depiction[s] of the item” (e.g., blurry images or images taken with a glare). *Id.*, [0079]-[0080], [0159]-[0161]. Rodriguez notes that “within store environments, imagery may be captured by other than cameras at point-of-sale checkout stations.” *Id.*, [0186]. For instance, with their smartphones or other mobile devices, customers can “capture still and/or video imagery from aisles of a supermarket, while doing their shopping.” *Id.* Unlike in the prior art, “imagery gathered during normal operation of a retail store is used as training imagery for a machine-learning system.” *Id.*, [0081]. Through processing these images, a machine-readable identifier, such as a barcode, is detected, and data can be decoded from the identifier. *Id.*, [0009], [0159], [0162].

119. Once the item is identified, such as by its barcode, the images and associated product identification are used to train a classifier that can recognize the

products even in “degraded images,” with examples shown in Figures 7A through 7F. *Id.*, Abstract, [0009], [0018], [0159], Figs. 6, 7A-7F. Thus, the product image and identifier (e.g., barcode) can serve as ground truth data “to label the product depicted in the sequence of imagery.” *Id.*, [0085], [0166]. These “input samples” are then submitted “as a training set to a classifier.” *Id.*, [0173]-[0174]. In other words, “[t]hese image frames—together with the identification information—are provided to a [classification system]..., to train it to recognize such items [in a test image] without resort to barcode decoding, watermark decoding, reference fingerprint data, etc.” *Id.*, [0186].



FIG. 6



Id., Figs 6, 7A-F.

X. CLAIM CONSTRUCTION

120. I have attributed all but two means-plus-function terms according to their plain and ordinary meaning as would have been understood by a person of ordinary skill in the art, at the time of the alleged invention, which I treat as no earlier than March 29, 2016 for claims 1–21, having taken into consideration the language of the claims, the specification, and the prosecution history of record as reflected in the '272 patent and its file history. *See generally* '272 *patent* (Ex. 1001), '272 *patent file history* (Ex. 1002).

121. I have identified two terms—"processor" and "product classifier"—that may potentially be construed as means-plus-function terms. I have identified the means-plus-function constructions proposed by Petitioner below. The charts I provide below identify the specific portions of the specification that best describe

the corresponding structure for each means-plus-function limitation. I have found these limitations to be present in the prior art as explained below whether they are simply construed with their plain meaning or whether means-plus-function applies.

Limitation	Function/Corresponding Structure
“processor” [Claim 18]	<u>Function</u>: capture images with a camera; detect and read content of a shelf label; associate shelf label content with product facing bounding boxes; cross-correlate product facing bounding boxes and other product facing bounding boxes across stores; build training data, use the training data to build a product classifier, and use the product classifier to perform one or more of an inventory management activity related to the product or classify an item as the product. <u>Corresponding structure</u>: Processing modules for computing, a server. <i>See, e.g.</i>, ’272 patent, Abstract, Fig. 1 (122), Fig. 3 (318), Fig. 4 (416), 1:61-64, 2:6-15, 2:21-27, 3:14-23, 5:4-20, 5:67-6:3, 6:61-7:9, 7:9-26, 7:43-47, 12:63-13:2.
“product classifier” [Claims 1, 10, 18]	<u>Function</u>: classifying an item as a product [claim 1]; performing an inventory management activity related to a product [claim 10]; performing one or more of: (1) an inventory management activity related to a product or (2) classifying an item as a product [claim 18]. <u>Corresponding structure</u>: a robot, set of robots, or cloud server implementing a convolutional, feedforward, or recurrent neural network. <i>See, e.g.</i>, ’272 patent, 12:24-40, 12:61-66.

XI. DETAILED EXPLANATION OF GROUNDS

A. Ground 1: Schwartz in view of Katircioglu Renders Claims 1, 3-13, 15-21 Obvious

1. Motivation to Combine Schwartz and Katircioglu

122. In my opinion, a POSITA would have been motivated to combine Schwartz and Katircioglu. Both references describe the need to mitigate out-of-stock

inventory situations that lower sales—and therefore store revenue—because customers cannot buy products that are not on a shelf. *Schwartz*, [0005]; *Schwartz Provisional*, [0002]; *Katircioglu*, [0003]. Both propose automated, camera-capture and image recognition solutions that reduce restocking latency stemming from more time-consuming and laborious methods of manual inventory auditing and shelf checking. *Schwartz*, [0005]-[0006]; *Schwartz Provisional*, [0002]-[0003]; *Katircioglu*, [0003]-[0004]. Schwartz generates a data stream of shelf inventory images captured by cameras in multiple stores (even competitor stores) at multiple times to inform a given store’s inventory status—this creates a massive volume of noisy image data—data containing both product-affiliated information (e.g., packaging and labels) and non-product information (e.g., gaps between products)—that needs to be analyzed and classified (i.e., processed) for human interpretable results to take any meaningful action (e.g., identifying and restocking out-of-stock inventory). *Schwartz*, [0081]-[0086]; *Schwartz Provisional*, [0076]-[0081]. In particular, Schwartz discloses using the UPC label on a product to identify the “ground truth” of the identity of the product in the image. *Schwartz*, [0134]. This mapping of an identifier (based on UPC label) to a product image is then used to train a machine learning system. *Id.*

123. A POSITA would have looked to Katircioglu to improve the training of this machine learning system because it discloses using “certain data clustering

or machine learning approaches...and machine learning techniques” to manage the increasing volume of data into its correlation database. *Katircioglu*, [0028]. In particular, instead of relying solely on the UPC label on the *product* itself (which may not always be visible or readable) as ground truth for the identity of the product, as taught by Schwartz, a POSITA would have used Katircioglu’s system of capturing *shelf label content*, as shown in Figure 7, to provide the ground truth of the identity of the products proximate to it. That association of product identifier from the shelf label, with the images of the products themselves, would, in the combination of Schwartz and Katircioglu, be used as training data for the product classifier disclosed by Schwartz.

124. A POSITA would have had a reasonable expectation of success for this combination because “[o]ne skilled in the art will recognize the method of implementing [these] clustering and machine learning techniques” according to the volume of techniques and methods I described in the background of the technology section, above. *Katircioglu*, [0028]; *see also* Section VII.A.

2. Independent Claim 1

a. Claim 1.0: “A method comprising:”

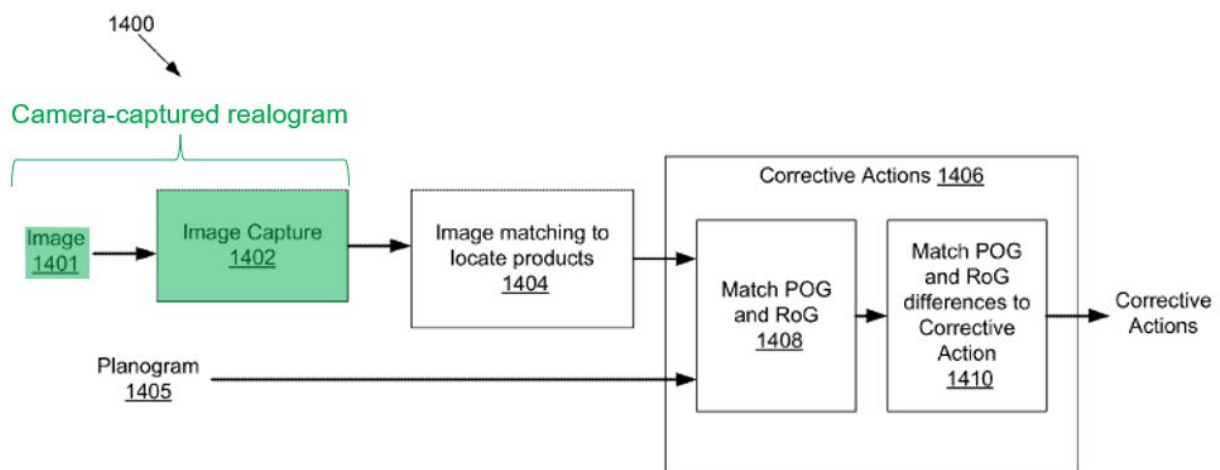
125. If limiting, it is my opinion that Schwartz teaches limitation 1.0. Schwartz discloses “methods and systems for planogram matching” including steps of image capture and processing to identify a product based on a symbolic identifier

and comparing that image-based product to known products in a known planogram. *Schwartz*, Abstract; [0006]. The system includes an image recognition application 103 that may run on either/both a client device 115a and/or recognition server 101. *Id.*, Abstract, Figs. 1-2, [0006], [0051].

126. Accordingly, it is my opinion that Schwartz teaches a method comprising steps.

b. Claim 1.1: “capturing one or more images of a portion of a shelf with a camera;”

127. In my opinion, Schwartz teaches limitation 1.1. Schwartz discloses capturing an image of a store shelf—“realogram” 1401—with a client device (e.g., mobile phone) camera as the first step 1402 of a planogram matching process. *Schwartz*, [0049], [0128] (defining “realogram,” “as “an image of a product shelf or display”), [0130]. This step starts an image capture and planogram comparison process shown in my annotated version of the figure I reproduced below:

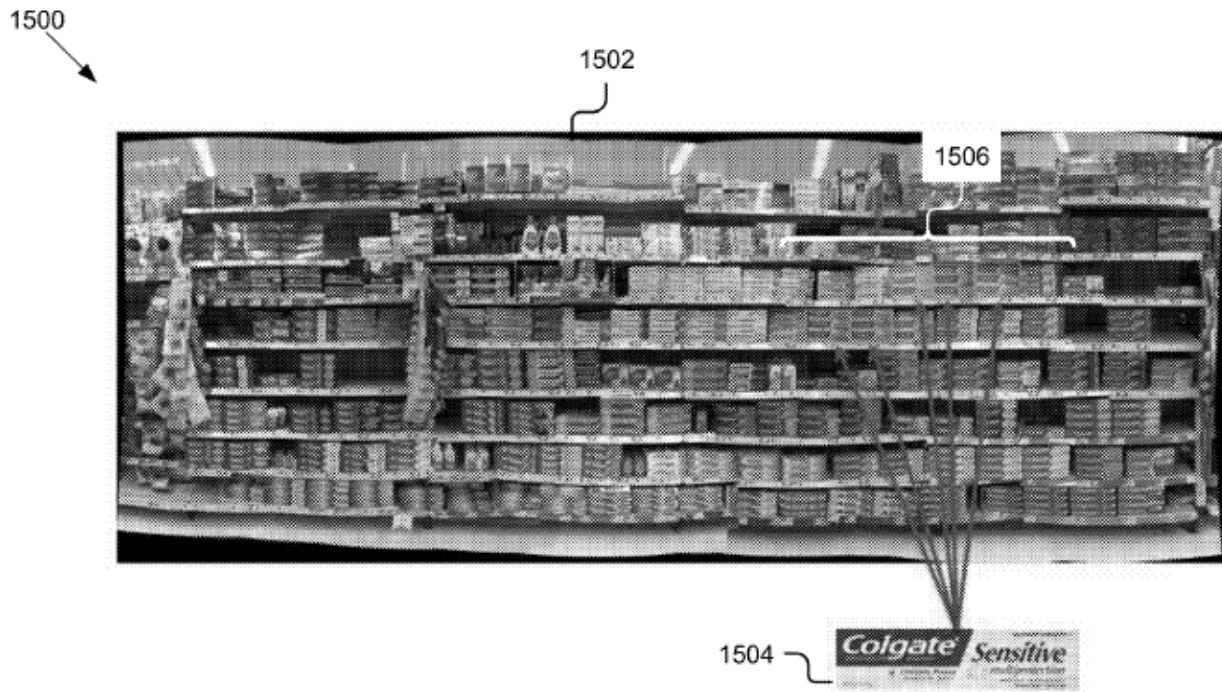


Id., Fig. 14A (item 1401, step 1402). The realogram may be one or several images “stitched together into a linear panoramic image.” *Schwartz*, [0130].

128. Accordingly, it is my opinion that Schwartz teaches capturing a realogram (“one or more images of a portion of a shelf”) with a camera.

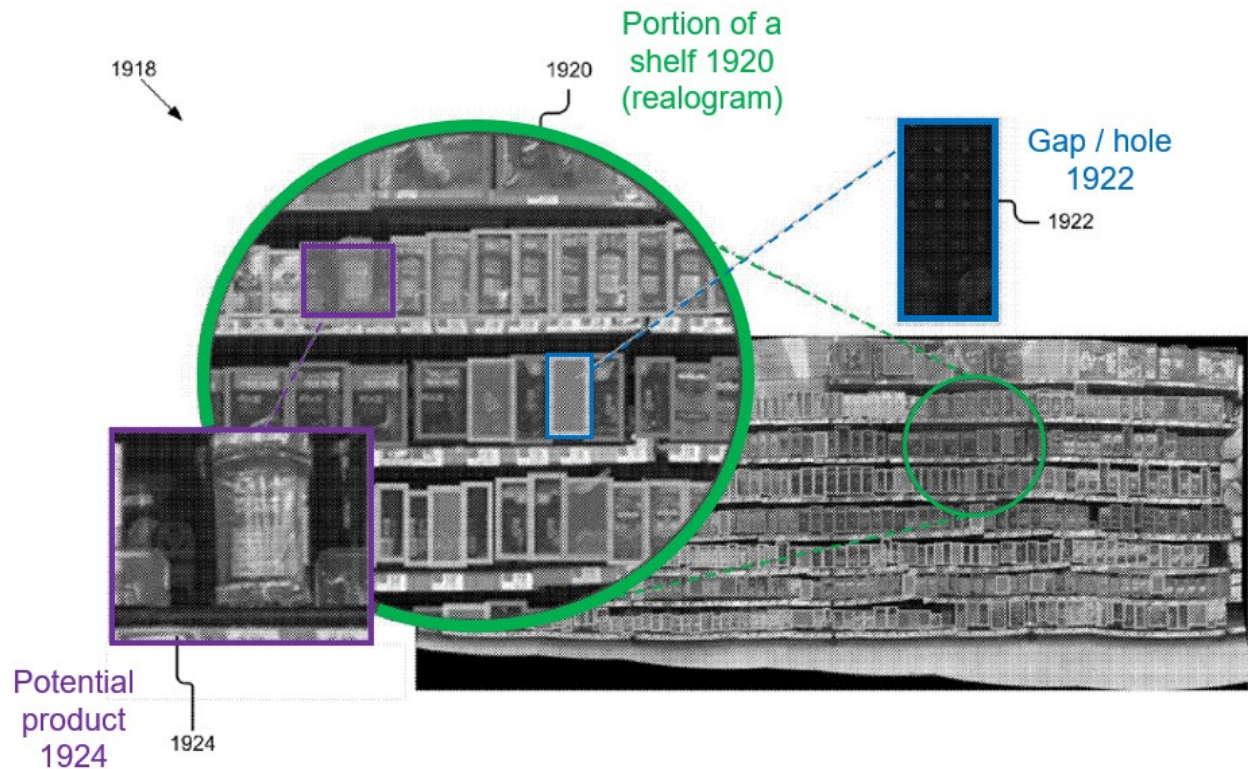
c. Claim 1.2: “detecting and reading content of a shelf label on the portion of the shelf within the one or more images;”

129. In my opinion, Schwartz alone, or in combination with Katircioglu, teaches limitation 1.2. *Schwartz*’s image processing module “determines a unique identifier for an object identified from the one or more received images.” *Schwartz*, [0073]. The system can also determine metadata such as “packaging identifier,” “shelf identifier,” and other “product attributes” from the image. *Id.*, [0076]. The unique identifiers can include a product’s specific UPC, which can be used to determine other product characteristics. *Id.*, [0137]. By example, Schwartz identifies a Colgate toothpaste product 1504 in a realogram using a packaging-type symbolic identifier common to products 1506 in the figure I reproduced below:



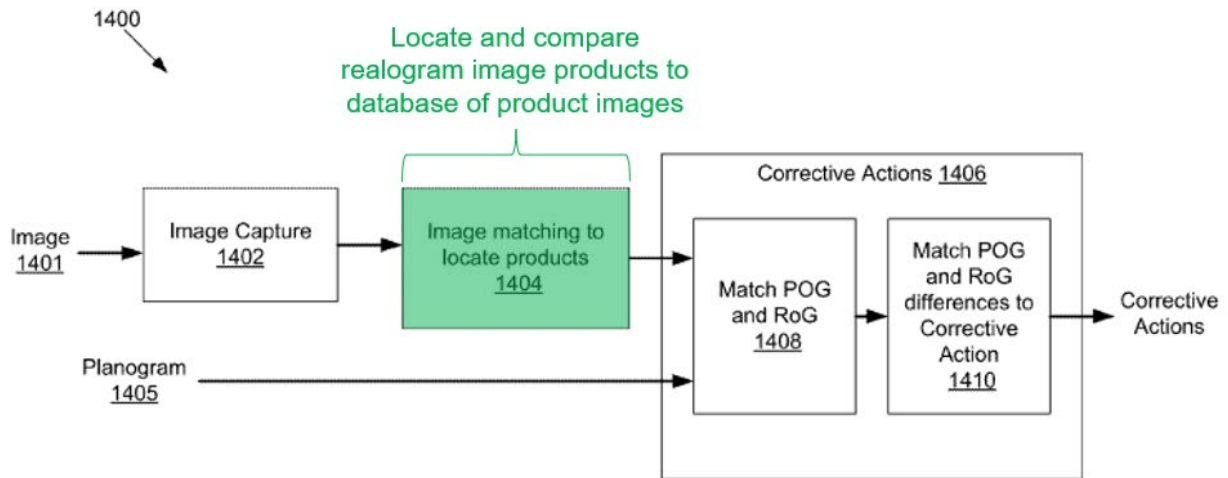
Id., Fig. 15A, [0137]. Although Schwartz discloses that product identifiers are generally on the products themselves, Schwartz suggests it also detects and reads content from shelf labels near the products.

130. In particular, Schwartz discloses identifying “gaps” in realogram images, lacking potential products:



Id., Fig. 19D, [0137], [0151]. In this situation, a POSITA would have understood that the Schwartz system cannot identify what product should have been shelved within that gap without reading the content on the shelf label below the gap.

131. The image processor locates and identifies realogram product candidates as known planogram products using matching algorithms and resultant recognition scores in the step shown in my annotated version of the figure I reproduced below. *Id.*, [0130], [0142]–[0143].



Id., Fig. 14A.

132. A POSITA would have understood, and found it obvious that Schwartz's image processor suggests detecting and reading product-affiliated (UPC) data on a product's shelf label in addition to or instead of a UPC on product packaging because (i) placement of aisle-facing shelf labels with UPC information was a well-known and common technique used at the relevant time allowing store employees to manually scan and place products at appropriate shelf locations with handheld scanners, (ii) Schwartz identifies gaps, where no products including UPCs can be seen, and (iii) Schwartz seeks to improve such manual planogram inventory compliance methods by automating the shelf label reading process with a camera capture technique requiring the same information to be presented in the same configuration to achieve the same, but more efficient result while preserving the manual technique for additional compliance checks. *Schwartz*, [0166]. Specifically and advantageously, Schwartz automates an otherwise manual product recognition

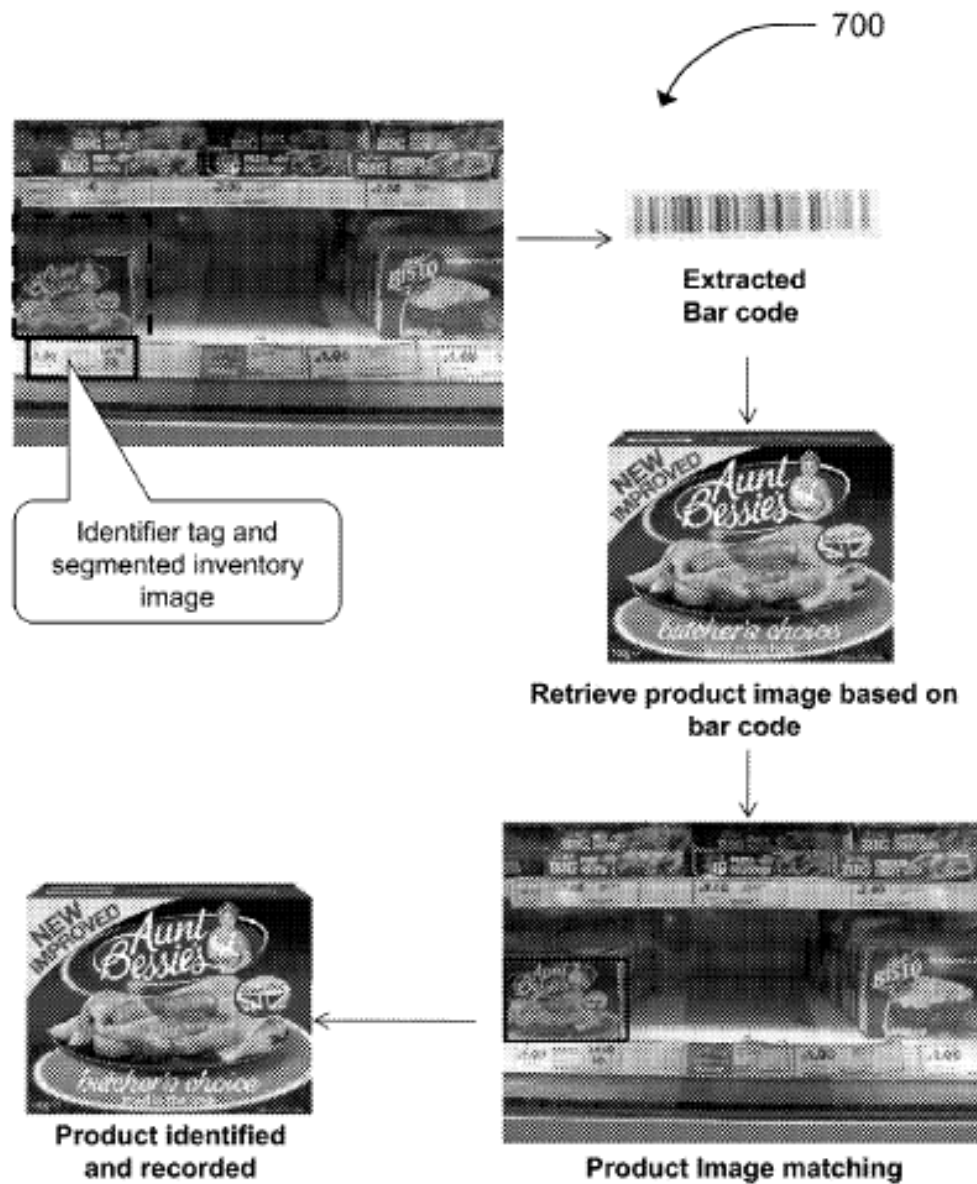
activity using predictable image capture and feature extraction improvements that lead to increased inventory management efficiencies and reduced manual labor hours and costs, while still preserving the manual option for, e.g., routine auditing.

133. Further, a POSITA would have recognized Schwartz analyzes the shelf label content of each product captured in a realogram because its image processing module determines product metadata that includes a “price of the product as sold in the retailer store,” which at the relevant time was commonly displayed only on a shelf label unique to a store and not the product packaging that remains uniform regardless where that product is sold to account for local economics and store-based discounts and is further updateable local to the store alone. *Schwartz*, [0076], [0109], [0117]. For example, a 20 oz. bottle of Coke soda that is centrally manufactured, distributed, then vended by a regional retailer would have the same appearance whether retailed in Seattle or Atlanta if received from the same distributor, but its price will likely vary between stores in the two different locations offering different sale promotions.

134. To the extent Patent Owner argues Schwartz does not teach detecting and reading content of a shelf label, Katircioglu does, and it would be present in the combination of Schwartz and Katircioglu. Katircioglu expressly discloses product identifier tags used for product identification and/or as price tags. *Katircioglu*, [0024]. These tags are “positioned proximate to the items on the shelves” and include

“barcodes...associated with a segmented inventory image and pre-captured product image” that are “extracted from the identifier tag to identify the product type of the items on the shelf.” *Id.*, [0038]. As shown in the figure I reproduced below, the pre-captured image barcode is matched to barcode data in an inventory database to retrieve the associated product image, and “the product identification is confirmed.”

Id.



Id., Fig. 7.

135. A POSITA would have found it obvious to modify Schwartz with Katircioglu for the reasons and with the reasonable expectation of success discussed in Section XI.A.1, above, and would have understood UPCs (as taught by Schwartz) and barcodes (as taught by Katircioglu) both encode product identifiers commonly

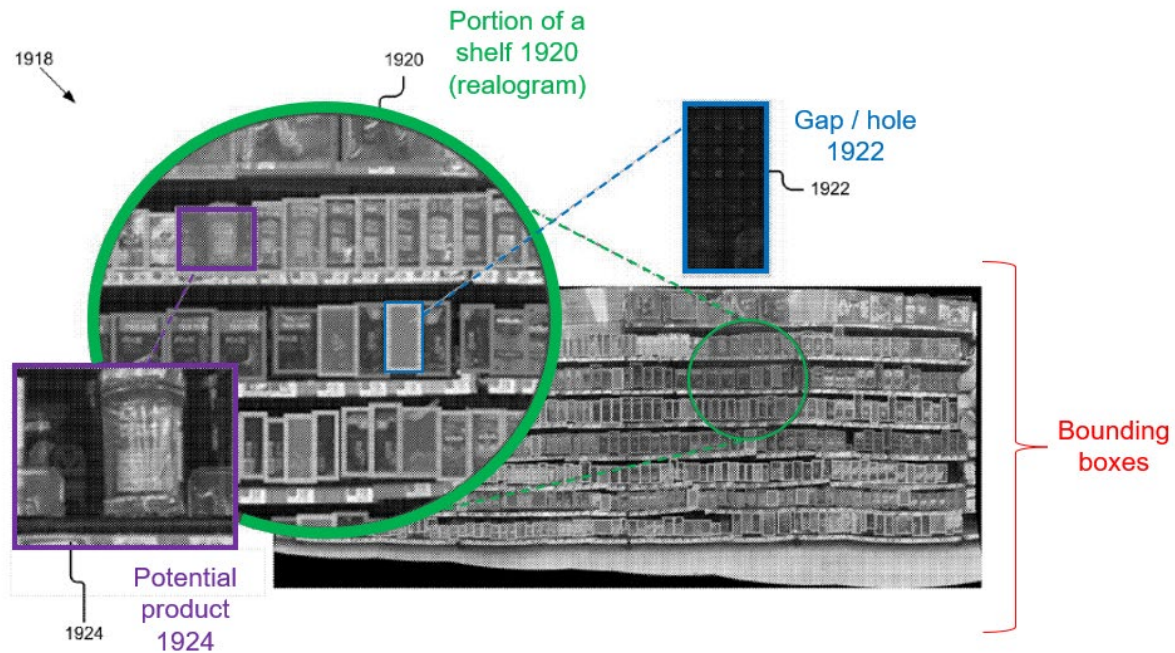
associated with and affixed proximately to products on their packaging (e.g., box, carton, hanging tab, sticker label) and/or corresponding rack or shelf tag and/or label to identify that product. *See Schwartz*, Fig. 19D, [0073], [0076], [0137]; *Katircioglu*, Figs. 7–8, [0038]–[0039].

136. Accordingly, it is my opinion that Schwartz alone, or in combination with Katircioglu, teaches capturing (“detecting”) and extracting (“reading”) a product-specific identifier from the tag on a shelf proximate to the product (“shelf label”) from within the realogram (“images”).

d. Claim 1.3: “defining one or more product facing bounding boxes surrounding one or more corresponding facings of a product within the one or more images;”

137. In my opinion, Schwartz, alone or in combination with Katircioglu, teaches limitation 1.3. As I explained in Section XI.A.2.b, above, a realogram includes a plurality of images taken by a camera of aisle-facing products.

138. Schwartz’s system further includes a gap detection module that processes a realogram with three processing filters including color evaluation, an edge/line detector, and an image feature corner detector to delineate between potential products, non-products, and shelf gaps. *Schwartz*, [0151]. Different portions of a realogram receive respective coloring, e.g., products are highlighted in purple, non-products in yellow, and gaps in dark coloring. *Id.* Potential products are bounded in boxes as shown in my annotated version of the figure I reproduced below:



Schwartz, Fig. 19D.

139. In addition to Schwartz, Katircioglu also teaches defining product facing bounding boxes within images. Katircioglu teaches a segmentation process for portioning a digital image into patches using a variety of edge detection methods (Sobel or Laplacian operators, phase congruency-based or differential edge detection, thresholding, and linking) and recognition of identifier tags “located on the shelves and are captured in the shelf image.” *Katircioglu, [0024]*. A POSITA would have recognized these edge detection techniques and the creation of bounding products around boxes in captured images were well-known. A POSITA would have recognized these techniques and steps as well-known techniques. For example, Varol describes generating bounding boxes around product images cropped from

images. *Varol*, 3-4; Section §VII.A.9. As a result, “each image patch only contains one specific product” as shown in my annotated version of the figures I reproduced below. *Id.*, [0024]-[0025], [0037].

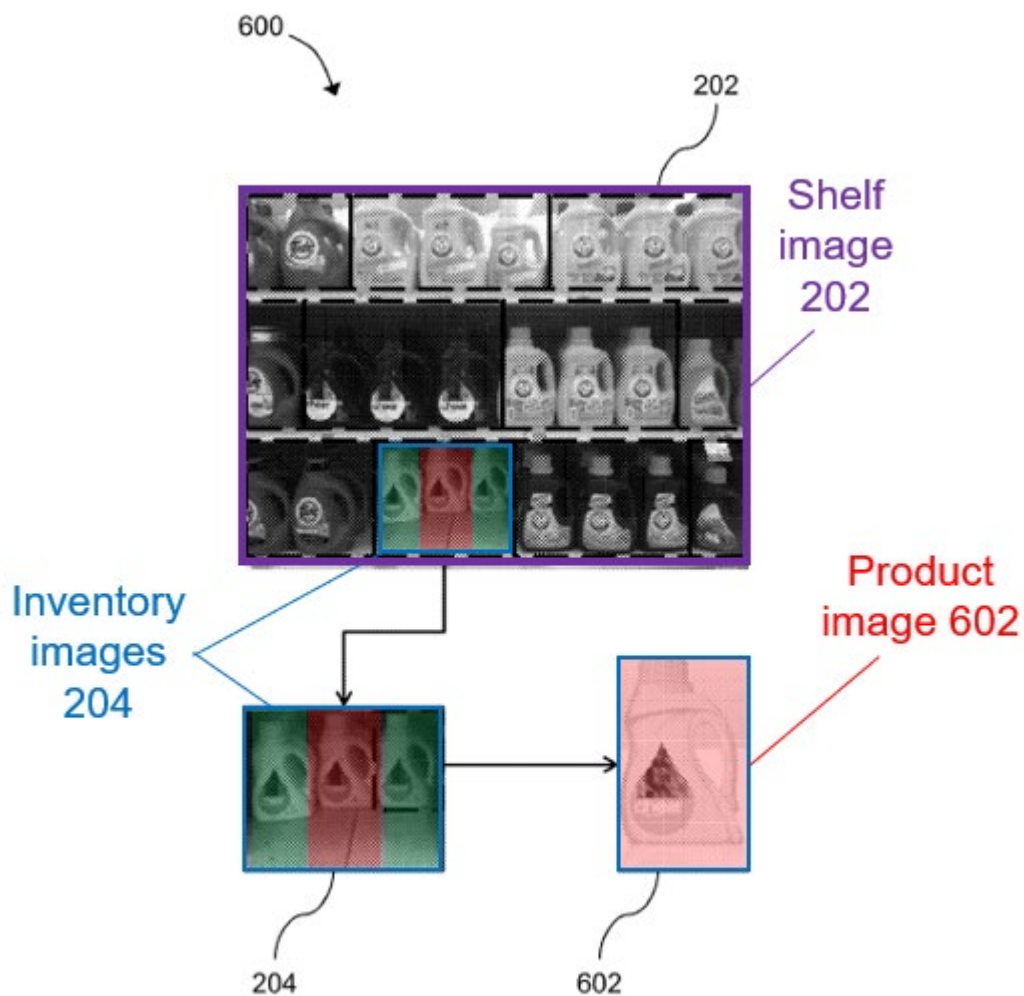
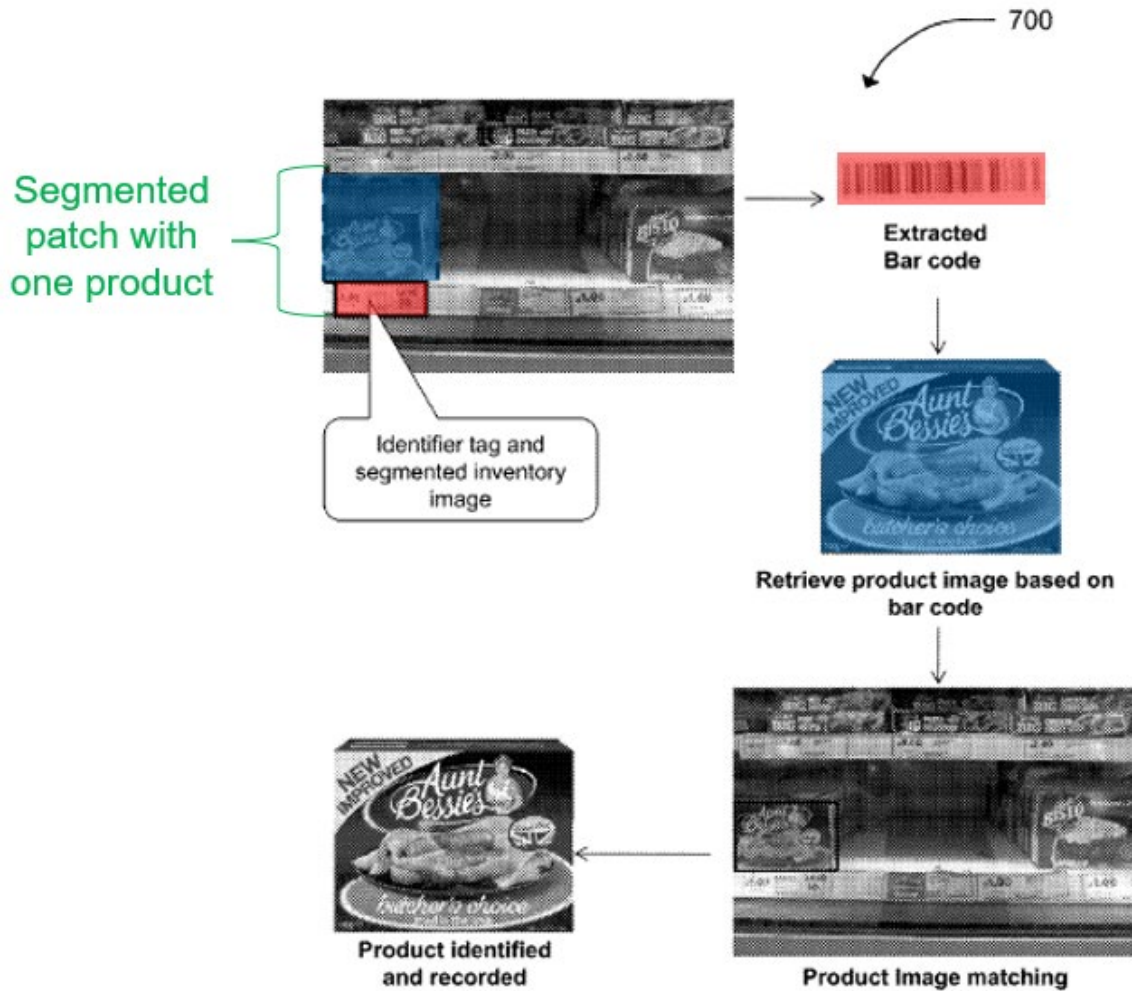


Fig. 6

Id., Fig. 6.

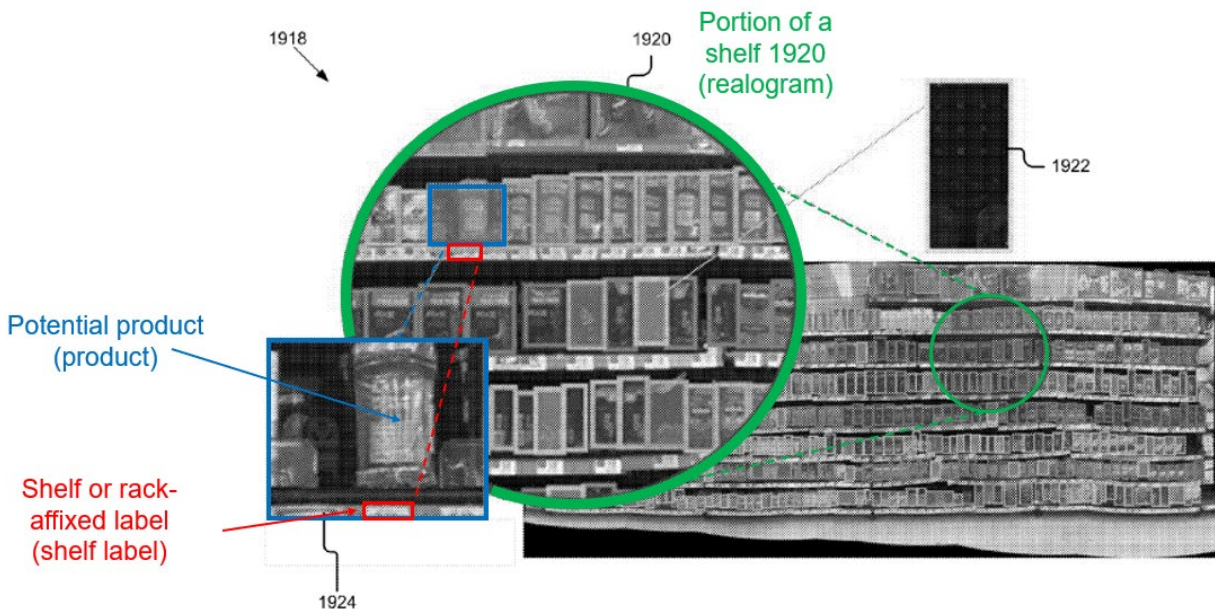


Id., Fig. 7.

140. Accordingly, it is my opinion that Schwartz alone, or in combination with Katircioglu, teaches edge detecting rectangular product boundaries (defining one or more product facing boundary boxes) in captured realograms.

- e. **Claim 1.4: “associating the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;”**

141. In my opinion, Schwartz alone, or in combination with Katircioglu, teaches limitation 1.4. Schwartz’s captured image 1401 (realogram) shows a label affixed to a rack located beneath a product as shown in my annotated version of the figure I reproduced below.



Schwartz, Fig. 19D (annotated).

142. A POSITA would have understood this to be a product-affiliated symbolic identifier—i.e., a product UPC 1416—according to well-known and conventional retail shelving arrangements like those I explained in the background of the technology and because Schwartz further discloses matching the UPC in image 1401 with a planogram 1405 to “identify visually similar products and

generate real world ground truths” as shown in my annotated version of the figure I reproduced below. *Id.*, [0131].

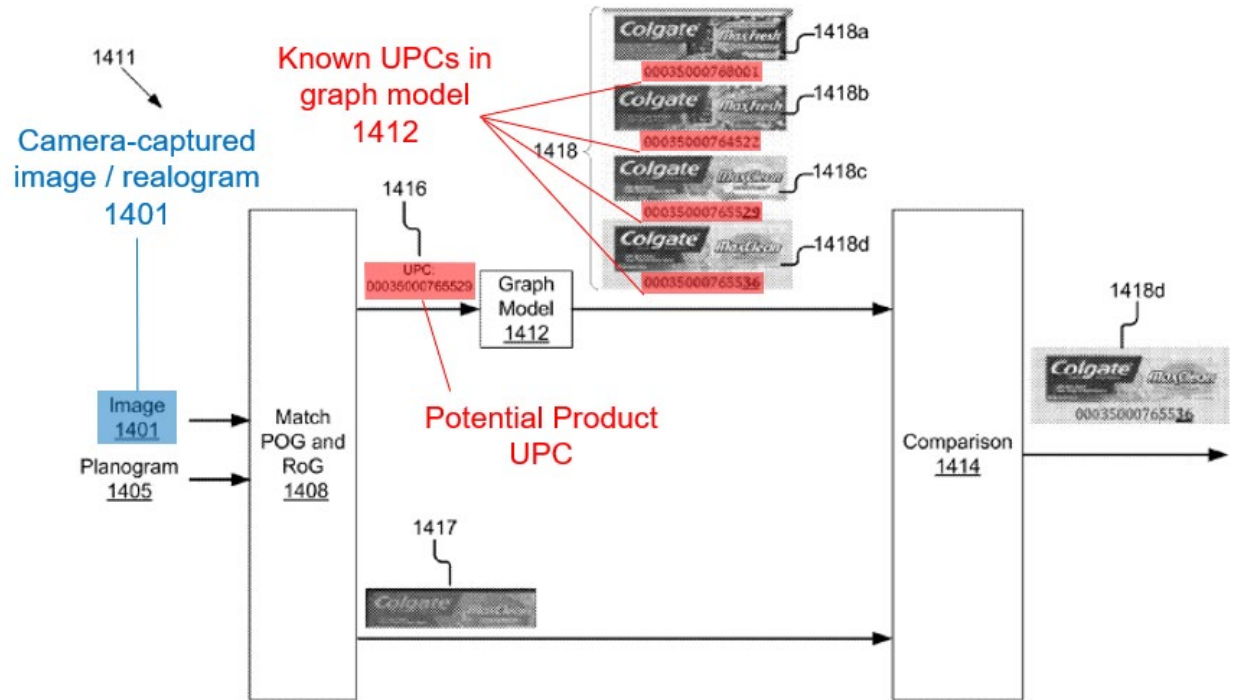
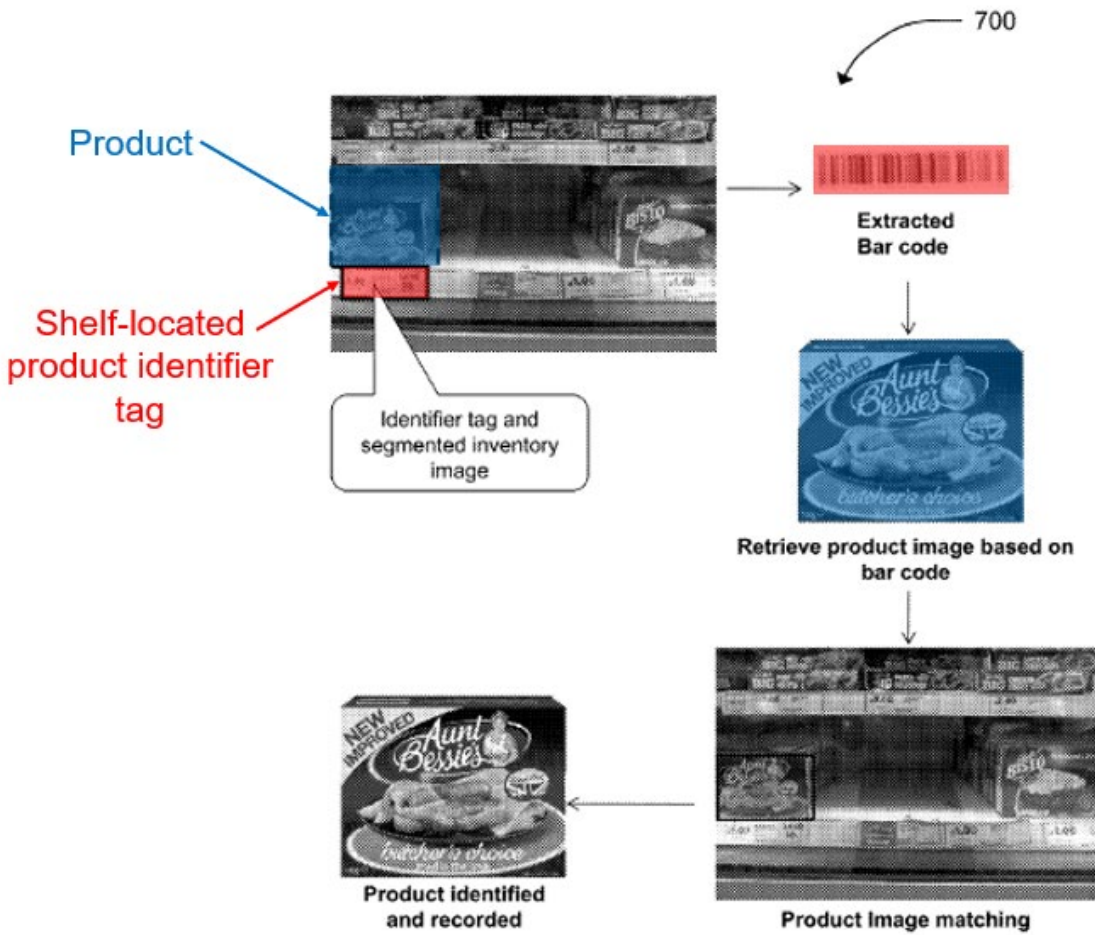


Figure 14B

Id., Fig. 14B. Real world ground truth data includes images of “actual scenes” “with corresponding labels such as product locations and UPCs that are used for algorithm performance valuation, system validation and testing and training of matching learning algorithms.” *Schwartz*, [0134]. Schwartz uses a “similar product graph model” 1412 to compare the UPC of the expected (i.e., potential) product 1416 to that of known symbolic identifiers (i.e., UPCs) of known products in graph model 1412. *Id.*, [0132]. An image detection block 1432 recognizes a product UPC from image 1401 to identify a product and further “output[s] bounding boxes or bounding

polygons indicating locations of products.” *Id.*, [0135]. A POSITA would have recognized that because Schwartz discloses an image detection step 1432 that analyzes a realogram to (i) identify a product UPC and (ii) draw bounding boxes around that product, that each UPC is therefore associated with the single bounded product because Schwartz determines product-affiliated information along with the product within this finite image. *Schwartz*, [0076] (describing metadata including shelf identifiers are determined for image-captured potential products) [0151] (describing edge-detection based image filtering to determine potential products).

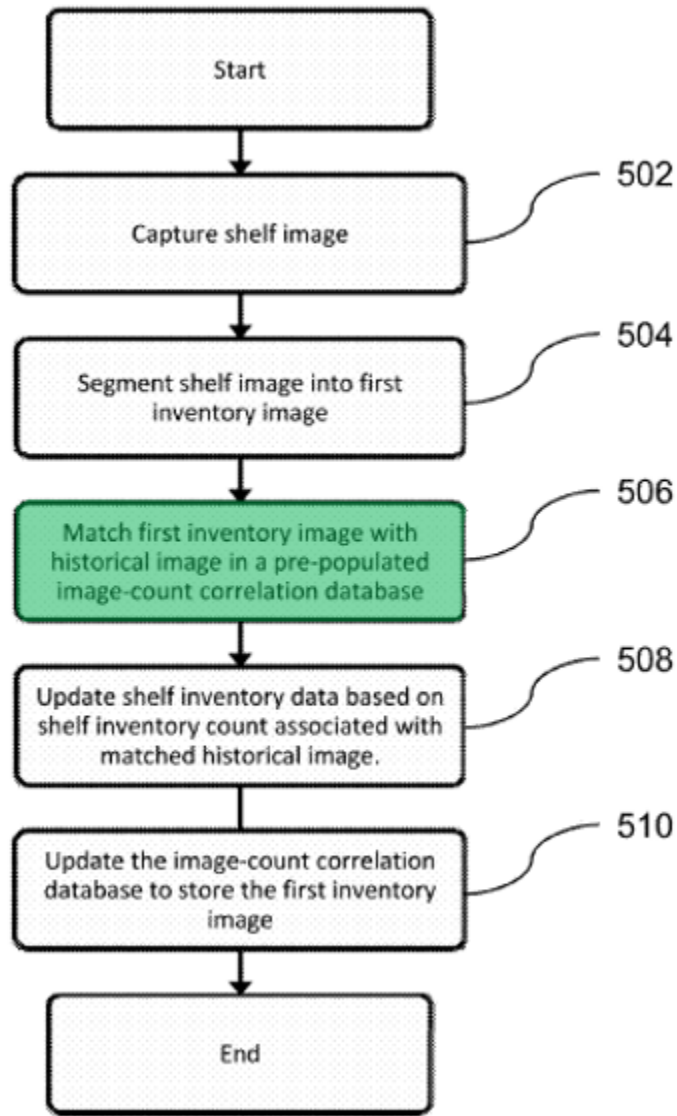
143. To the extent Patent Owner argues Schwartz does not teach associating shelf label content with product facing bounding boxes, the combination with Katircioglu does. Katircioglu teaches that “identifier tags *associated with the items* are located on the shelves and *are captured in the shelf image 202*” as shown in the annotated version of the figure I reproduced below. *Katircioglu*, [0024].



Katircioglu, Fig. 7.

144. In particular, Figure 7 depicts a process “for identifying and verifying products utilizing identifier tags...positioned proximate to the items on the shelves” that further include “barcodes and are associated with a segmented inventory image and pre-captured product image.” *Id.*, [0038]. The barcode on the identifier tag is extracted and matched with barcode data in inventory database 410 to confirm product identification. *Id.*

145. A POSITA would have recognized this automated processing occurs in step 506 of Figure 5 (highlighted in my annotated version of the figure I reproduced below) of Katircioglu's system for "automated monitoring of shelf inventory." *Id.*, Fig. 5 (506), [0001], [0003], [0032]-[0036]. Katircioglu automated this process to automate the otherwise less efficient, time and labor-hour costly technique requiring "manual staff monitoring [that] is labor intensive and subject to staff availability." *Katircioglu*, [0003].



Id., Fig. 5.

146. Therefore, a POSITA would have recognized the combination of Schwartz and Katircioglu associates shelf label content (as shown in Katircioglu Figure 7) with product images from the Schwartz's realogram, in place of the UPC information disclosed by Schwartz.

147. Accordingly, it is my opinion that Schwartz alone, or in combination with Katircioglu, teaches associating shelf-located identifier tags (“associating the shelf label content”) with the single product segmented patches (“product facing bounding boxes that includes and defines each of the one or more corresponding product facings”) identified in a realogram.

- f. **Claim 1.5: “cross-correlating the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;”**

148. In my opinion, Schwartz alone, or in combination with Katircioglu, teaches limitation 1.5. As I explained in Section XI.A.2.e, Schwartz alone and in combination with Katircioglu captures, analyzes, and edge detects rectangular boundaries of product and their corresponding shelf labels. Further, Schwartz “aggregates and summarizes” object (i.e., product) information received “from images of multiple stores of the same or different type” and “with the same or different local demographics”; the data is “collected over time.” *Schwartz*, [0081]. Specifically, the data comes from images that are “taken from a single visit to a single store, multiple visits to a single store, and multiple visits to multiple stores” where the store may even be substituted for a “warehouse, stockroom, store room, cabinet, etc.” *Schwartz*, [0083].

149. A POSITA would have recognized that Schwartz collects and correlates product identification and (at least in combination with Katircioglu) affiliates shelf label content (e.g., price) information with that of a planogram for a store, and further *cross-correlates* that product and affiliated shelf label information *across* stores, store types, and time because this enables, for example, store A to identify a new, unindexed product C stocked in a different warehouse B and newly appearing in store A to update and ensure planogram compliance showing store A now stock product C to satisfy inventory audits. *Schwartz*, [0083], [0090], [0100]. A POSITA would have understood that retailer chains would have desired uniform quality assurance and control across their subsidiary retail locations to best track and inform global manufacturer and supplier distributors while optimizing regional sales and local inventory auditing processes, all of which Schwartz's cross-store data correlating method would have enabled.

150. Further, a POSITA would have understood that the same product is often positioned in multiple locations of a single store. For example, candy bars may be included in an aisle, as well as near the check-out register. When images of those candy bars are collected, it would have been obvious to a POSITA to keep all of the images of the same product together. The combination of Schwartz and Katircioglu, which associates product images with shelf label content, would provide such cross-correlation.

151. Accordingly, it is my opinion that Schwartz alone, or in combination with Katircioglu, teaches spatial-temporal aggregation of retail product realogram images and affiliated shelf label information to ensure planogram compliance (“cross-correlating...product facing bounding boxes...across one or more stores by matching shelf label content”)

g. Claim 1.6⁵: “building training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes;”

152. In my opinion, Schwartz alone, or in combination with Katircioglu, teaches limitation 1.6. As I explained in Sections XI.A.2.e–XI.A.2.f, Schwartz alone or in combination with Katircioglu teaches associating shelf label content with cross-correlated product facing bounding boxes.

153. In Schwartz, image recognition application 103 “generate[s] real world ground truth results of matching the planogram 1405 and the image 1401 to identify products.” *Schwartz*, [0134]. Schwartz further defines real world ground truth data as “images with corresponding labels such as product locations and UPCs....used for algorithm performance evaluation, system validation and testing and training of

⁵ By my review, the subject matter of this claim limitation does not appear to be disclosed anywhere in the ’272 patent or its provisional application beyond its express listing as the claim itself.

matching learning algorithms.” *Id.* Schwartz also discloses using “synthetic ground truth,” where the “true labels” are known as part of the creation process. *Id.* All this image-derived data product identification, location, and metadata is used by a data generator that “*generates a data stream* including [that] information.” *Schwartz*, [0080]. Expansively, the data stream also includes this same image-derived product information aggregated “from a single visit to a single store, multiple visits to a single store, and multiple visits to multiple stores.” *Schwartz*, [0081].

154. Katircioglu also teaches building training data. Katircioglu teaches capturing product inventory data across time in an “ongoing process” of “image and inventory *data correlation building*.” *Katircioglu*, [0027]. Correlation pairs between camera-captured and correlation database images are stored in a database. *Id.*, [0025]–[0027]. When the “data volume becomes very large” in the correlation database, Katircioglu applies “machine learning approaches...[and] techniques.” *Id.*, [0028].

155. A POSITA would have understood that the combination of Schwartz and Katircioglu uses the association of the product identifier from the shelf label as the ground truth for the identification of the products in the bounding boxes of the realogram images to train a machine learning system, such as a Convolutional Neural Network, to recognize other images of the same product. *Schwartz*, [0080], [0134], [0135]. Mitchell corroborates the usefulness of a neural network for training

data; neural networks are particularly helpful and tuned for “training data [that contains] noisy, complex sensor data” such as “inputs from cameras.” *Mitchell*, 83; *see also* Section VII.A.1. A POSITA would have easily recognized that the same utility neural networks provide for, e.g., “face recognition” using input facial images would apply equally and advantageously to product recognition using input product images. *Mitchell*, 112; *see also* Section VII.A.1.

156. Accordingly, it is my opinion that Schwartz alone, or in combination with Katircioglu, teaches generating (“building”) a stream of correlated captured and known product images across stores and time (“training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes”).

h. Claim 1.7⁶: “using the training data to build a product classifier; and”

157. In my opinion, Schwartz teaches limitation 1.7. Schwartz’s image recognition application includes a classification module 207 that receives the data stream “that includes product information of a retail environment” generated from the cyclical combination of synthetic and real-world ground truths. *Schwartz*, Fig. 2 (207), [0083], [0134]. The classification module “may determine...conditions and extract...data from the data stream” according to a non-exclusive list of predicated

⁶ *See supra* note 5.

conditions including out of stock or stock out, recall, position map, new product, price match, representative performance, planogram adherence, and/or profit adherence conditions. *Schwartz*, [0084]-[0094]. Classification may also occur using a “Convolutional Neural Network (CNN) classifier.” *Schwartz*, [0135].

158. A POSITA would have recognized the step of receiving a generated data stream rich with multi-store, time series product information attributes (i.e., training data such as recalled items) at a classification module—with a Convolutional Neural Network (CNN) classifier—that determines conditions for extracting “a portion of” data that predicates the determination of the relevant condition for extracting (i.e., classifying) that data (i.e., “build[s] a product classifier”) because to be useful, the classification module receives retail production information data with condition-satisfying attributes (so it can identify recalled products from store-generated realograms). *Schwartz*, [0083], [0086]. It would have been well known to a POSITA to use training data consisting of images mapped to product identifiers to build a product classifier that could recognize a later, unidentified image as a particular product. *See Merler*, 3–5 (describing the GroZi-120 database of 120 grocery products containing *in vitro* and *in situ* images); *George*, 440–52 (disclosing classification approach tested on both a proprietary supermarket products dataset and the GroZi-120 dataset); *see also* Sections VII.A.7, VII.A.4.

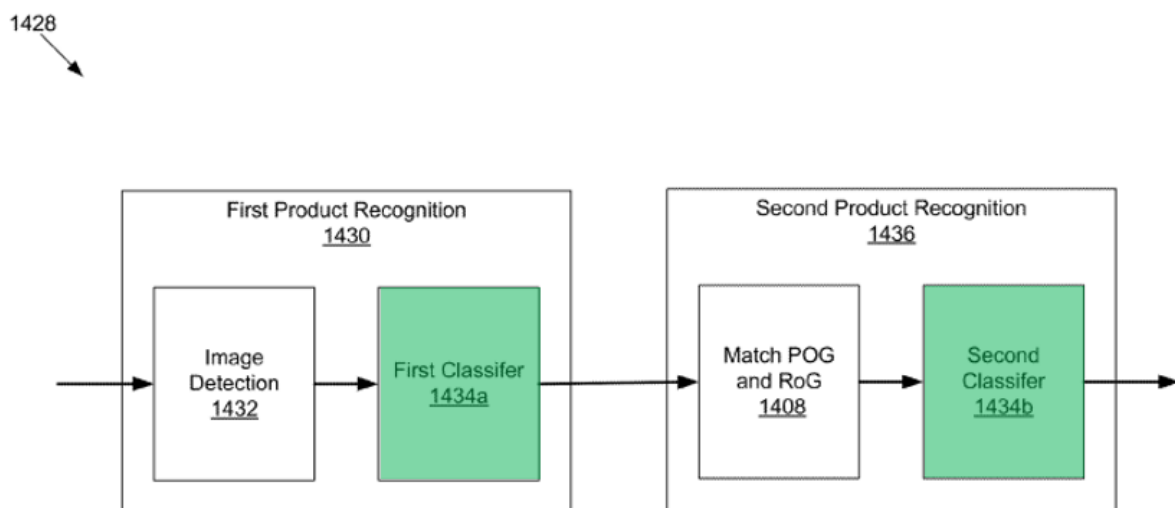
159. If “product classifier” is construed as a means-plus-function term, Schwartz discloses or suggests the structure discussed in Section X. As I previously explained, Schwartz discloses the product classifier being implemented in a “Convolutional Neural Network.” *Schwartz*, [0135]. Schwartz further discloses a “recognition server,” which a POSITA would have understood to be a cloud server or would have been motivated to implement as a cloud server. *Id.*, [0047]-[0048]. The Convolutional Neural Network and the recognition server disclosed in Schwartz each perform the claimed functions, classifying an item as a product [claim 1]; performing an inventory management activity related to a product [claim 10]; and performing one or more of: (1) an inventory management activity related to a product or (2) classifying an item as a product [claim 18]. The disclosed structure is equivalent to the structure disclosed in the ’272 patent, i.e. a robot, set of robots, or cloud server implementing a convolutional, feedforward, or recurrent neural network.

160. Accordingly, it is my opinion that Schwartz teaches sending and receiving a generated data stream rich with product information to a classification module (“build a product classifier”).

i. Claim 1.8: “using the product classifier to classify an item as the product.”

161. In my opinion, Schwartz teaches limitation 1.8. As I explained in Section XI.A.2.b, Schwartz discloses receiving a generated data stream (training

data) at a classification module (product classifier) to extract (classify) “a portion of” data satisfying one of a number of non-exclusive, specific condition. *Schwartz*, [0084]–[0094]. Specifically, classification module 207 includes software instructions “to provide the functionality for *classifying* conditions of a data stream.” *Schwartz*, [0082]. The classification module receives a data stream “that includes product information of a retail environment” and “extracts a portion of data from the data stream based on the condition.” *Schwartz*, Fig. 2 (207), [0083]. Further, the classification module may include two classifiers to both identify a product by UPC and its location using a Convolutional Neural Network (CNN) classifier, shown in my annotated version of the figured I reproduced below. *Schwartz*, [0135].



Id., Fig. 14C (annotated).

162. A POSITA would have recognized the step of extracting condition-satisfying data from the data stream as a step of classifying a product satisfying that

condition because only a “portion of data” is extracted from a realogram to identify products that satisfy that condition. *Schwartz*, [0086] (describing under the recall condition, “the classification module 207 matches product information in the data stream to a list of recall products in a database...[and] determines whether a received image of the data stream includes the recall products” and “flags each recall product on the received image”); *Schwartz Provisional*, [0081]. It would have been well known to a POSITA to use a product classifier that could recognize an image as a particular product. *See Merler*, 3–5 (describing the GroZi-120 database of 120 grocery products containing *in vitro* and *in situ* images); *George*, 440–52 (disclosing classification approach tested on both a proprietary supermarket products dataset and the GroZi-120 dataset); *see also* Sections VII.A.7, VII.A.4..

163. Accordingly, it is my opinion that Schwartz teaches a step of extracting condition-satisfying products (“classify”) using a classification module (“product classifier”).

3. Independent Claim 10

164. In my opinion, Schwartz alone, or in combination with Katircioglu, teaches claim 10 for substantially the same reasons as claim 1, as I explained in Section XI.A.2, because the claims are substantively identical. *Compare* claim limitations 10.0-10.8, *with* claim limitations 1.0-1.8.

165. The table I provide below correlates the reasons why Schwartz alone, or in combination with Katircioglu, renders obvious claim limitations 10.0-10.8 for the reasons rendering claim 1 obvious.

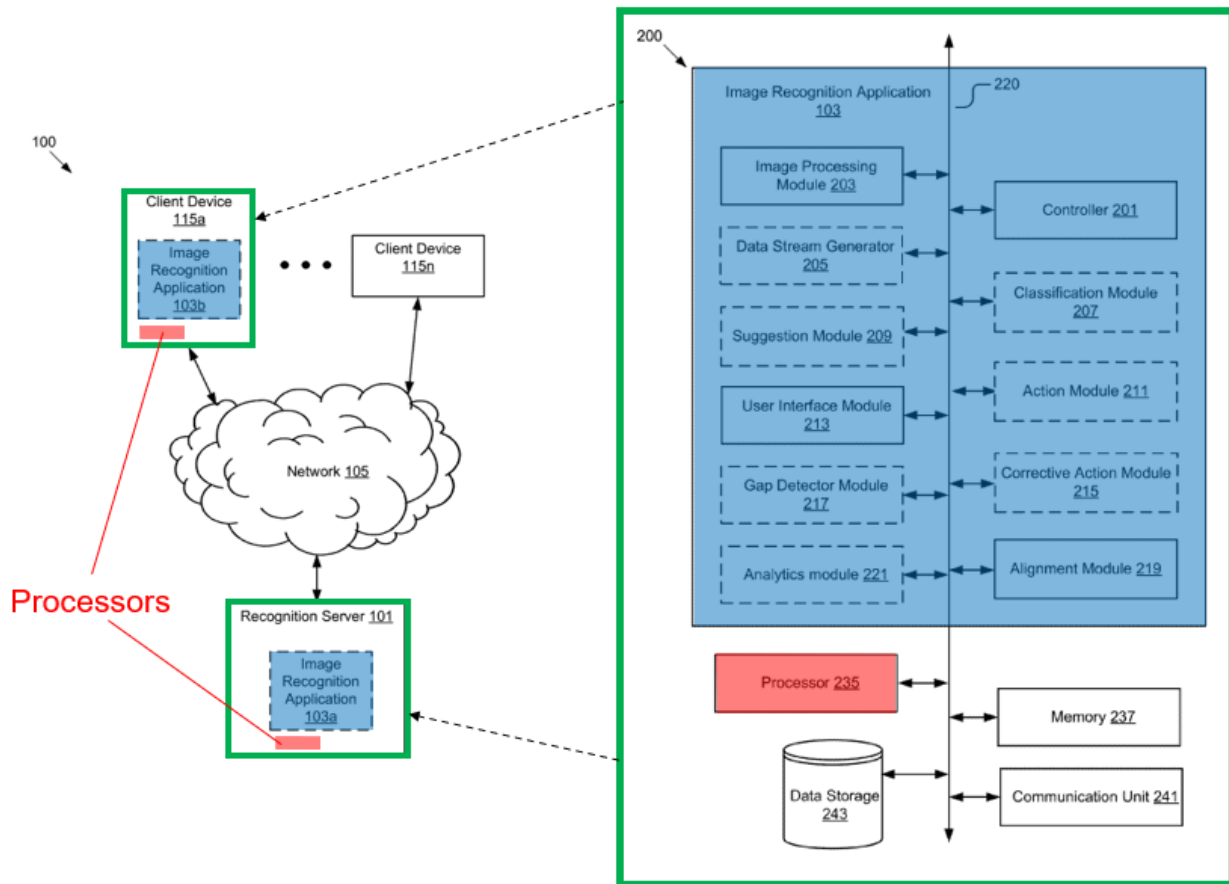
Claim 10		
Limitation	Substantially Similar Limitation(s)	Reference Section
10.0	1.0	XI.A.2.a
10.1	1.1	XI.A.2.b
10.2	1.2	XI.A.2.c
10.3	1.3	XI.A.2.d
10.4	1.4	XI.A.2.e
10.5	1.5	XI.A.2.f
10.6	1.6	XI.A.2.g
10.7	1.7	XI.A.2.h
10.8	1.8	XI.A.2.i

166. To the extent the discussion of limitation 1.8 does not disclose “using the product classifier to perform an inventory management activity related to the product” as in 10.8, that is also disclosed by the combination of Schwartz and Katircioglu. For example, Schwartz teaches the “classification module” determines an out-of-stock condition. *Schwartz*, [0084], [0117]; *Schwartz Provisional*, [0079], [0112]. Similarly, Katircioglu discloses a system for inventory management. *Katircioglu*, [0004]-[0006], [0025]-[0027].

4. Independent Claim 18

167. It is my opinion that Schwartz alone, or in combination with Katircioglu, teaches claim 18 for substantially the same reasons as I explained for claim 1 in Section XI.A.2 because the claims are substantively identical, except that claim limitations 18.0-18.2 recite a “system comprising: a processor” and a “system memory coupled to the processor...storing instructions configured to cause the processor” to execute the same steps as recited in claim 1. *Compare* claim limitations 18.3-18.10, *with* claim limitations 1.1-1.8. To the extent the discussion of limitation 1.8 does not disclose “using the product classifier to perform an inventory management activity related to the product” as in 18.10, that is also disclosed by the combination of Schwartz and Katircioglu as I explained with respect to 10.8.

168. Schwartz’s “method and systems for planogram matching” executes the steps described in claim 1 with an image recognition application running on a computing device with processor 235 coupled via bus 220 to memory 237 storing executable instructions as shown in my annotated version of the figures I reproduced below. *Schwartz*, Abstract, [0058]-[0061], [0072], [0116].



Schwartz, Figs. 1-2 (200, 235), [0058]-[0061].

169. If “processor” is construed as a means-plus-function term, *Schwartz* teaches the structure discussed in Section X. As I explained above *Schwartz* discloses processing modules for computing, *Schwartz*, Fig. 2 (200, 235), [0058]. *Schwartz* further discloses a server in the form of a “recognition server.” *Id.*, [0047]-[0048]. *Schwartz*’s processing modules and server, along with the connected peripherals each perform the claimed function of “capture images with a camera; detect and read content of a shelf label; associate shelf label content with product facing bounding boxes; cross-correlate product facing bounding boxes and other

product facing bounding boxes across stores; build training data, use the training data to build a product classifier, and use the product classifier to perform one or more of an inventory management activity related to the product or classify an item as the product.” The disclosed structures in the prior art are identical to the disclosed structures in the ’272 patent, i.e. processing modules for computing, a server.

170. Additionally, the table I provided below correlates the reasons why Schwartz alone, or in combination with Katircioglu, teaches claim limitations 18.3-18.10 for the reasons rendering claim 1 obvious.

Claim 18		
Limitation	Substantially Similar Limitation(s)	Reference Section
18.3	1.1	XI.A.2.b
18.4	1.2	XI.A.2.c
18.5	1.3	XI.A.2.d
18.6	1.4	XI.A.2.e
18.7	1.5	XI.A.2.f
18.8	1.6	XI.A.2.g
18.9	1.7	XI.A.2.h
18.10	1.8, 10.8	XI.A.2.i, XI.A.3

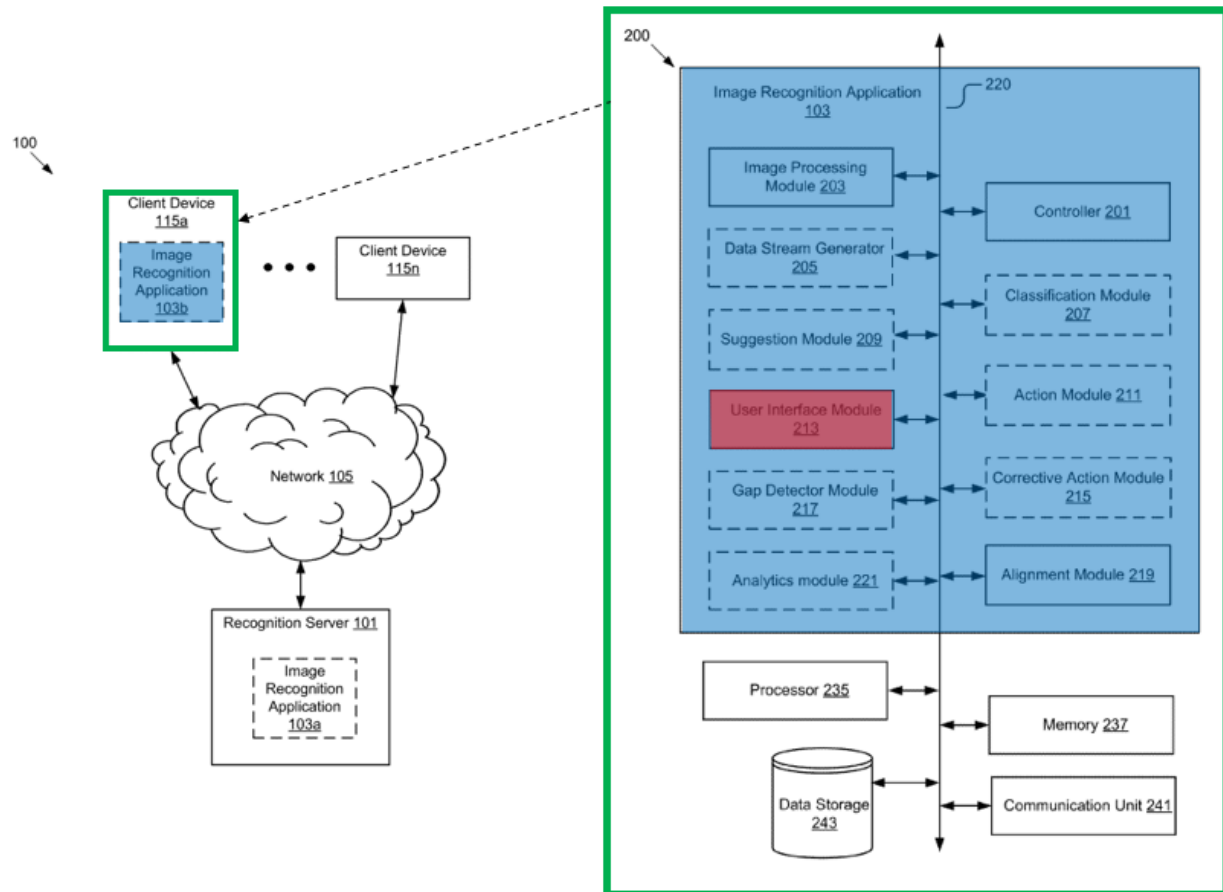
5. Claims 3, 15

- a. Claim 3: “The method of claim 1, further comprising using manual input to identify potential inventory designated in the one or more product facing bounding boxes.”**

b. Claim 15: “The method of claim 10, further comprising using manual input to identify potential inventory designated in the one or more bounding boxes.”

171. Schwartz alone, or in combination with Katircioglu, teaches claims 1 and 10 as I explained above. *See* Section XI.A.2. Specifically, Schwartz teaches delineating rectangular boundaries around recognized realogram products using edge detection. *See* Section XI.A.2.d. In my opinion, Schwartz further teaches the additional limitations of claims 3 and 15. The image recognition application uses a “productish analysis” to match products captured in a realogram with a planogram, but some products remain “potential”; i.e., at times a “product [] cannot be definitively identified,” so “manual scoring of images and evaluation of planogram compliance” may be required. *Schwartz*, [0166]. Specifically, Schwartz’s method discloses, as one option, using real world ground truths in which “humans make final decisions about what the true labels are,” e.g., when products are very similar or a UPC cannot be recognized. *Schwartz*, [0134].

172. A POSITA would have recognized the image data is presented to a user for manual comparison via a graphical user interface on the client device containing both (i) the camera for capturing realograms and (ii) image recognition application for executing the systems and methods of recognizing products in those realograms as shown in my annotated and combined version of the figures I reproduced below. *Schwartz*, Fig. 2 (213), [0066], [0068].



Id., Figs. 1-2.

173. Accordingly, it is my opinion that Schwartz alone, or in combination with Katircioglu, teaches manually comparing potential products in realograms with those in planograms using a client device graphical user interface (“using manual input to identify potential inventory”).

6. **Claim 4: “The method of claim 1, wherein using the training data to build a product classifier comprises integrating training to: (1) learn product image descriptors and (2) classify items into a neural network.”**

174. Schwartz alone, or in combination with Katircioglu, teaches claim 1 as I explained above. *See* Section XI.A.2. Specifically, Schwartz predicates the step of extracting data from a generated data stream—classifying data—with determining which condition to satisfy to conduct that extraction—building the classifier (e.g., knowing recalled products exist, make the determination to use that condition as a basis to extract that information). *See* Sections XI.A.2.g–XI.A.2.i. In my opinion, Schwartz further teaches the additional limitations of claim 4.

175. Schwartz determines a new product condition from the generated data stream, then extracts a portion of that product’s data from the data stream. *Schwartz*, [0090]. Specifically, the classification module identifies an unindexed (new) product from the data stream when its “features...e.g., luminance, color, package dimension) do not match the features stored in the database.” *Id.* The classifier then extracts and *combines* this data with additional information “obtained from other resources” including “competitors’ on-shelf inventory that was collected during one or more visits to the competitors’ store” to then “identify the product and index the product in the database.” *Id.*

176. Schwartz further discloses a product classifier, built from the training data as discussed above, in the form of a “Convolutional Neural Network.” *Schwartz*,

[0135]. A POSITA would have recognized that neural networks are ideally suited for classifying image-captured products. *Mitchell*, 114 (“After training on a set of 260 images, classification accuracy over a separate test set is 90%” as opposed to 25% from randomly guessing); *see also* Section XI.A.2.g.

177. Accordingly, it is my opinion that Schwartz teaches identifying features of unindexed products (“learn product image descriptors”), and combining the extract data with, e.g., observed competitor product data (“integrating training” from a “network”) to index products (“classifying items”) into a product classifier in the form of a convolutional neural network (“neural network”).

7. Claim 5: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises manually associating the shelf label content with each of the one or more product facing bounding boxes.”

178. Schwartz alone, or in combination with Katircioglu, teaches claim 1 as I explained above. *See* Section XI.A.2. Specifically, Schwartz alone and in combination with Katircioglu teaches *automatically* associating shelf label content with product facing bounding boxes using a gap detector module to edge detect and associate boundaries of products and identifier tags in realograms within a larger “productish analysis.” *See* Section XI.A.2.e; *Schwartz*, [0166].

179. In my opinion, Schwartz further discloses or suggests the additional limitations of claim 5 of *manually* associating the shelf label content with product

facing bounding boxes because the “productish analysis” further allows “*manual scoring* of images and evaluation of planogram compliance” when “a product...cannot be definitively identified” by presenting a client device user with the images on the device’s graphical user interface. *Schwartz*, [0166]. Specifically, Schwartz discloses that “[w]hile algorithms may assist in generating ground truth data, *humans make final decisions* about what the true labels are.” *Id.*, [0134]. Additionally, the image recognition application provides real world ground truth data “with assistance from a user,” especially when a user is needed to identify a product lacking a symbolic identifier (e.g., UPC) or when two very similar products are analyzed. *Id.*

180. Accordingly, it is my opinion that Schwartz teaches manually associating identifier tags on shelves (shelf label content) with potential product candidates (“product facing bounding boxes”).

8. Claims 6, 17

- a. Claim 6: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”**
- b. Claim 17: “The method of claim 10, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”**

181. Schwartz alone, or in combination with Katircioglu, renders obvious claims 1 and 10 for the reasons I explained above. *See* Section XI.A.2. In my opinion, Schwartz alone and in combination with Katircioglu renders obvious claims 6 and 17 for the same reasons as discussed in claim limitation 1.4, which I explained makes associations without human intervention, above. *See* Section XI.A.2.e.

182. A POSITA would have therefore recognized that Schwartz automatically associates products with their symbolic identifiers (e.g., price or UPC) containing corresponding identification and price information when classifying products in realograms because Schwartz clearly teaches using synthetic ground truth and steps of extracting both (i) products and (ii) their affiliated shelf identifier and price information in the same realogram given their common and uniform product-over-label paired orientation to allow one instance of image capture to help

a store employee determine whether, e.g., breakfast cereal boxes match their prescribed supermarket price in a given week. *Id.*, Fig. 19D, [0053], [0091], [0124].

183. In the same vein, Katircioglu clearly teaches that price and product “identifier tags [are] *associated* with the items...located on the shelves and are captured in the shelf image 202.” *Katircioglu*, [0024]. Barcodes on the identifier tags are “*associated* with a segmented inventory image and pre-captured product image.” *Id.*, [0038].

184. Accordingly, it is my opinion that Schwartz alone and in combination with Katircioglu renders obvious claims 6 and 17.

9. Claim 7

a. Claim 7.0: “The method of claim 1, wherein capturing one or more images of a portion of a shelf comprises capturing one or more images of a portion of a shelf in a first store; and further comprising:”

185. If limiting, Schwartz teaches limitation 7.0 for the same reasons I explained in limitations 1.1 (“capturing one or more images of a portion of a shelf with a camera”) and 1.5 (“cross correlating . . . product facing bounding boxes...*across one or more stores*”). See Sections XI.A.2.b, XI.A.2.f. Specifically, Schwartz discloses generating a data stream of images including “product information of a retail environment” that are “taken from a single visit to a single store, multiple visits to a single store, and multiple visits to multiple stores.” *Schwartz*, [0083]. A POSITA would have therefore recognized, and found it obvious,

that Schwartz captures realograms from at least a first and second (multiple) stores that are separate (e.g., one is a “competitor” to the other). *Id.*, [0090].

b. Claim 7.1.1: “capturing one or more further images of portion of another shelf in a second store;”

186. Schwartz teaches limitation 7.1.1 for the same reasons as I explained in limitation 7.0, where a POSITA would have recognized, and found it obvious, that capturing images at multiple stores includes a separate store capturing images at a second (e.g., “competitor”) store with its own, separate shelves and correspondingly unique images. *See* Section XI.A.9.a.

c. Claim 7.1.2: “reading content of another shelf label on the portion of the other shelf within the one or more further images;”

187. Schwartz teaches limitation 7.1.2 for the same reasons as I explained in limitations 1.2 and 7.0 because Schwartz discloses executing this step at multiple stores (i.e., “reading content of *another* shelf label...of the *other* shelf”). *See* Sections XI.A.2.c, XI.A.9.a.

d. Claim 7.1.3: “defining one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and”

188. Schwartz alone, or in combination with Katircioglu, teaches limitation 7.1.3 for the same reasons as I explained in limitations 1.3 and 7.0 because Schwartz executes this step at multiple stores (i.e., “defining one or more *other* product facing

bounding boxes...within the one or more *further* images”). *See* Sections XI.A.2.d, XI.A.9.a.

- e. **Claim 7.1.4: “associating the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;”**

189. Schwartz alone, or in combination with Katircioglu, teaches limitation 7.1.4 for the same reasons as I explained in limitations 1.4 and 7.0 because Schwartz executes the step at multiple stores (i.e., “associating the *other* shelf label content with each of the one or more *other* product facing bounding boxes”). *See* Sections XI.A.2.e, XI.A.9.a.

- f. **Claim 7.2: “wherein cross-correlating comprises cross-correlating the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the second store by matching the shelf label content with the other shelf label content; and”**

190. Schwartz alone, or in combination with Katircioglu, teaches limitation 7.2 for the same reasons as I explained in limitations 1.5 and 7.0 because Schwartz executes this step at multiple stores (i.e., “cross-correlating...product facing bounding boxes *and the one or more other* product facing bounding boxes across the *first store and the second* store”). *See* XI.A.2.f, XI.A.9.a.

- g. **Claim 7.3: “wherein building training data comprises building the training data based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store.”**

191. Schwartz alone, or in combination with Katircioglu, teaches limitation 7.3 for the same reasons as I explained with respect to limitations 1.6 and 7.0 because Schwartz executes this step at multiple stores (i.e., “building training data based on the association of the shelf label content *and the other* shelf label content...across the *first store and the second store*”). See Sections XI.A.2.g, XI.A.9.a

10. Claim 11

192. Schwartz alone, or in combination with Katircioglu, teaches claim 11 for substantially the same reasons as I explained with respect to claim 7 because the claims are substantively identical. *Compare* claim limitations 11.0-11.3, *with* claim limitations 7.0-7.3. See Sections XI.A.9.a–XI.A.9.g.

193. The table I provided below correlates the reasons why Schwartz alone, and in combination with Katircioglu, teaches a claim limitations 11.0-11.3 for the reasons rendering claim 7 obvious.

Claim 11		
Limitation	Substantially Similar Limitation(s)	Section Reference
11.0	7.0	XI.A.9.a
11.1.1	7.1.1	XI.A.9.b

11.1.2	7.1.2	XI.A.9.c
11.1.3	7.1.3	XI.A.9.d
11.1.4	7.1.4	XI.A.9.e
11.2	7.2	XI.A.9.f
11.3	7.3	XI.A.9.g

11. Claim 19

194. Schwartz alone, or in combination with Katircioglu, teaches claim 19 for substantially the same reasons as I explained with respect to claims 7 and 11 because the claims are substantively identical. *Compare* claim limitations 19.0-19.3, *with* claim limitations 7.0-7.3. *See* Sections XI.A.9.a–XI.A.9.g.

195. The table I provided below correlates the reasons why Schwartz alone, or in combination with Katircioglu, teaches claim limitations 19.0-19.3 for the reasons rendering claim 7 obvious.

Claim 19		
Limitation	Substantially Similar Limitation(s)	Section Reference
19.0	7.0	XI.A.9.a
19.1.1	7.1.1	XI.A.9.b
19.1.2	7.1.2	XI.A.9.c
19.1.3	7.1.3	XI.A.9.d
19.1.4	7.1.4	XI.A.9.e
19.2	7.2	XI.A.9.f
19.3	7.3	XI.A.9.g

12. Claims 8, 12, 20

- a. **Claim 8: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images subsequent to capturing the one or more images.”**
- b. **Claim 12: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further subsequent to capturing the one or more images.”**
- c. **Claim 20: “The system of claim 18, wherein instructions configured to capture one or more further images comprise instructions configured to capture the one or more further images subsequent to capturing the one or more images.”**

196. Schwartz alone, or in combination with Katircioglu, teaches claims 7, 11, and 18 as I explained above. *See* Sections XI.A.9, XI.A.10, XI.A.4. In my opinion, Schwartz teaches the additional limitations of claims 8, 12, and 20. Specifically, Schwartz’s method and system “aggregates and summarizes” object (i.e., product) information received “at different time[s] and provides *time series data* of objects to the data stream generator.” *Schwartz*, [0078], [0081]. The images are “collected over time,” and are “taken from a single visit to a single store, multiple visits to a single store, and multiple visits to multiple stores.” *Schwartz*, [0081], [0083]. This allows one store to determine an unindexed product shelved at a competitors’ store. *Schwartz*, [0090].

197. A POSITA would have recognized, and found it obvious, that images are captured across time and across stores to create a relative sequence of captured

images respective to a store. For example, if store A is the “competitor’s store” Schwartz first captures a product C image and indexes that image in the database (“prior to”) store B doing the same. Conversely, if store B is running an inventory audit and captures an unindexed image of product C it identifies using images from competitor store A, then store B has captured images of product C after (“subsequent to”) store A.

13. Claims 9, 13

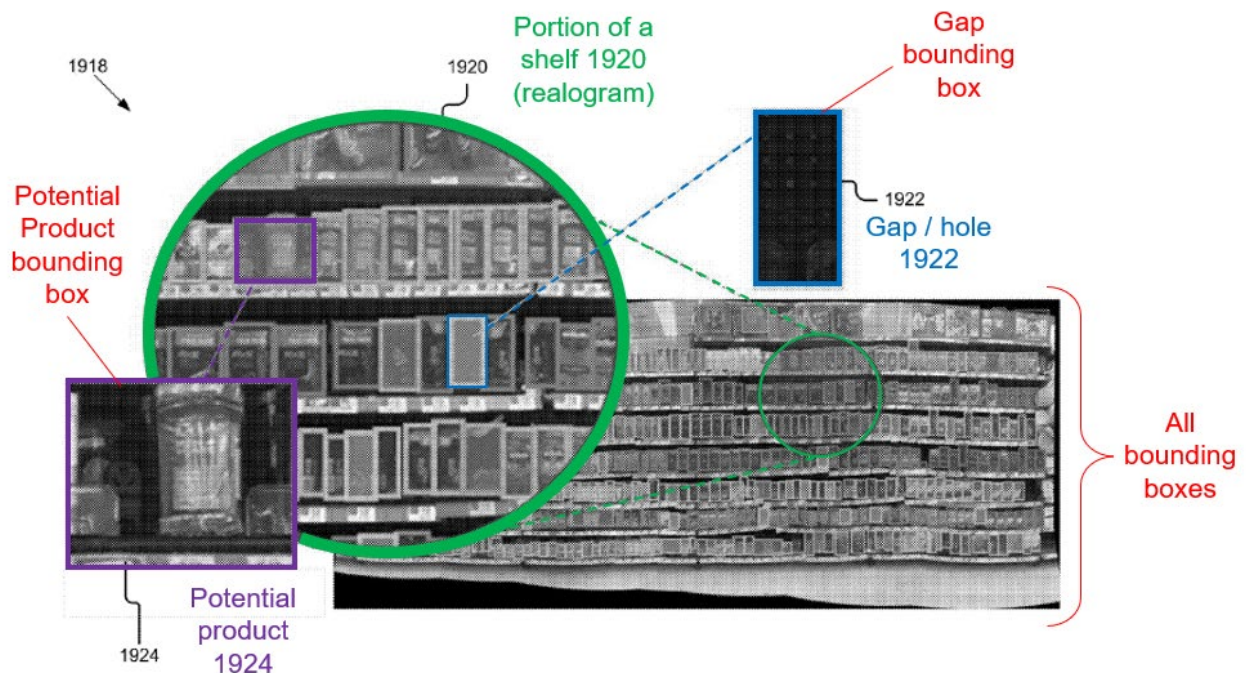
- a. Claim 9: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”**
- b. Claim 13: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”**

198. Schwartz alone, or in combination with Katircioglu, teaches claims 7 and 11 for the reasons I explained above. *See* Sections XI.A.9, XI.A.10. In my opinion, Schwartz teaches the additional limitations of claims 9 and 13 for the reasons and image capture timing (“prior to”) I explained in claims 8, 12, and 20. *See* Section XI.A.12.

14. Claim 16: “The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.”

199. Schwartz alone, or in combination with Katircioglu, teaches claim 10 for the reasons I explained above. *See* Section XI.A.3. In my opinion, Schwartz

teaches the additional limitations of claim 16. Specifically, Schwartz discloses a “gap detector module” that “perform[s] gap detection and ‘productish’ analysis” including “reviewing gaps and doing further analysis to determine the probability that the gaps may include an image of a product that cannot be definitively identified.” *Schwartz*, [0166]. The gap detector module uses filters and edge detection to delineate rectangular boundaries of “gap areas” that are either “holes” 1922 (i.e., bona fide gaps) or “potential products” 1924 as shown in my annotated version of the figure I reproduced below. *Schwartz*, [0151].



Schwartz, Fig.19D.

15. Claim 21: “The system of claim 18, wherein instructions configured to capture one or more further images of another portion of a shelf comprise instructions configured to capture the one or more further images.”

200. Schwartz alone, or in combination with Katircioglu, teaches claim 18 for the reasons I explained above. *See* Section XI.A.4. In my opinion, Schwartz teaches the additional limitations of claims 21. As discussed in claims 8, 12, and 20, Schwartz’s method and system “aggregates and summarizes” object (i.e., product) information received “at different time[s] and provides *time series data* of objects to the data stream generator.” *Schwartz*, [0078], [0081]. The images are “collected over time,” and are “taken from a single visit to a single store, multiple visits to a single store, and multiple visits to multiple stores.” *Schwartz*, [0081], [0083], [0090]. A POSITA would have recognized, and found it obvious, that Schwartz’s method and system captures a plurality of images (“further images”) across time and store space.

B. Even if the ’272 Patent Gets Its Earliest Provisional Date, Schwartz Renders Claims 1, 3-13, and 15-21 Obvious

1. Schwartz Is Entitled to Its Provisional Date

201. As I explained above, Schwartz alone, or in combination with Katircioglu, teaches obvious claims 1, 3-13, 15-21 if Schwartz is entitled to its publication date. *See* Section XI.A.

202. Patent Owner may argue the ’272 patent is entitled to its earliest listed March 29, 2016, provisional application filing date—despite the differences in

subject matter I previously noted. Schwartz would nevertheless be prior art because it is entitled to the date of the Schwartz Provisional. Therefore, Schwartz antedates even the earliest possible priority date of the '272 patent. The Schwartz Provisional discloses or suggests claims 1, 3-13, and 15-21 with like disclosure as shown in the table below.

Claims	Schwartz, Ex. 1005	Schwartz Provisional Ex. 1006
1.0, 10.0	Figs. 1-2, Abstract, [0006], [0051]	Figs. 1-2, Abstract, [0003], [0048]
18.0-18.2	Figs. 1-2, Abstract, [0047]-[0048], [0058]-[0061], [0072], [0116]	Figs. 1-2, Abstract, [0044]-[0045], [0053]-[0056], [0067], [0111]
1.1, 10.1, 18.3	Fig. 14A, [0049], [0128], [0130]	Fig. 14, [0046], [0122], [0124]
1.2, 10.2, 18.4	Figs. 14A, 15A, 19D, [0073], [0076], [0109], [0117], [0130], [0137], [0142]-[0143], [0151], [0166]	Figs. 14, 15A, 19D; [0068], [0071], [0104], [0112], [0124]-[0125], [0130]-[0131], [0136]-[0137], [0151]
1.3, 10.3, 18.5	Fig. 19D, [0151]	Fig. 19D, [0136]
1.4, 10.4, 18.6	Figs. 14B ⁷ , 19D, [0076], [0131]-[0132], [0134]-[0135], [0151]	Fig. 19D, [0136]
1.5, 10.5, 18.7	[0081], [0083], [0090], [0100]	[0076], [0078], [0085], [0095]
1.6, 10.6, 18.8	[0080]-[0081], [0083], [0134]-[0135]	[0075], [0078]
1.7, 10.7, 18.9	Fig. 2, [0047]-[0048], [0083], [0094], [0134]-[0135]	Fig. 2, [0044]-[0045], [0078]-[0089]
1.8, 10.8, 18.10	Figs. 2, 14C, [0082]-[0094], [0117], [0135]	Fig. 2, [0077]-[0089], [0112]
3, 15	Figs. 1-2, [0066], [0068], [0134], [0166]	Figs. 1-2, [0061], [0063], [0151]

⁷ The **highlighted** disclosure only appears in the Schwartz publication.

4	[0090], [0135]	[0085], [0110]
5	[0134], [0166]	[0151]
6, 17	Fig. 19D, [0053], [0091], [0124]	Fig. 19D, [0050], [0071], [0086], [0118], [0125], [0136]
7.0, 11.0, 19.0	[0083], [0090]	[0078], [0085]
7.1.1, 11.1.1, 19.1.1	<i>See</i> 7.0	<i>See</i> 7.0
7.1.2, 11.1.2, 19.1.2	<i>See</i> 7.0, 1.2	<i>See</i> 7.0, 1.2
7.1.3, 11.1.3, 19.1.3	<i>See</i> 7.0, 1.3	<i>See</i> 7.0, 1.3
7.1.4, 11.1.4, 19.4	<i>See</i> 7.0, 1.4	<i>See</i> 7.0, 1.4
7.2, 11.2, 19.2	<i>See</i> 7.0, 1.5	<i>See</i> 7.0, 1.5
7.3, 11.3, 19.3	<i>See</i> 7.0, 1.6	<i>See</i> 7.0, 1.6
8, 12, 20	[0078], [0081], [0083], [0090]	[0076], [0078], [0085]
9, 13	<i>See</i> 8, 12, 20	<i>See</i> 8, 12, 20
16	Fig. 19D, [0151], [0166]	Fig. 19D, [0136], [0151]
21	[0078], [0081], [0083], [0090]	[0076], [0078], [0085]

2. Schwartz Renders the Claims Obvious, Based Solely on What is in the Schwartz Provisional

203. Although very similar, the disclosure of Schwartz (publication) provides additional details beyond what was disclosed in the Schwartz Provisional. This is evident when comparing the Schwartz Provisional (Fig. 14, [0124]-[0125]), with the Schwartz publication (Figs. 14A-C, [0130]-[0137]). Even if, however, the Schwartz Provisional does not provide sufficient written description for the items highlighted above (Figures 14B, 14C, [0131], [0132], [0134], [0135]), the limitations that cite those disclosures (1.4, 1.6, 1.7, 1.8, 10.4, 10.7, 10.8, 18.1, 18.6, 18.9, 3, 5, 15) would have been obvious based on the other disclosures of Schwartz that are supported by the Schwartz Provisional.

204. Limitations 1.4, 10.4, and 18.6 cite to Figure 14B and paragraphs [0131]-[0132] and [0134]-[0135] of Schwartz as disclosing using the UPC captured in an image to identify a product. Even without these disclosures, it would have been obvious to use a UPC code to identify a product. For example, Merler discloses searching for images of products based on UPC codes. Merler, 3-4. And even without such knowledge, this limitation would still be taught by Katircioglu's reading of barcodes on identifier tags.

205. Limitations 1.6, 1.7, and 1.8 (and the corresponding limitations of claims 10 and 18) relate to building training data, building a product classifier, and using a product classifier, respectively. While the analysis cites Figure 14C and paragraph [0134]-[0135] of Schwartz for the disclosure of a Convolutional Neural Network classifier, training and using classifiers based on image data was well known in the art, even without such disclosure, and would have been obvious to use in the combination of Schwartz and Katircioglu. For example, Mitchell teaches training and using neural networks, Mitchell, 81-86; Carbonetto teaches training and using a classifier to recognize items within images, Carbonetto, 2-3; Merler and George each teach training and using a classifier to recognize grocery items, Merler 3-5; George, 440-43. Further, *Katircioglu* depicts and defines image segmentation as “the process of partitioning a digital image into patches so that each image patch only contains one specific product” (Fig. 6 and [0024]), and further discloses that

“image and inventory data correlation building can be an ongoing process...it takes a certain period of time to build a rigorous image-count correlation database” using a “large number of correlation pairs” such that “when the data volume becomes very large in this correlation database, certain data clustering or machine learning approaches can be applied” “[o]ne skilled in the art will recognize” ([0027]-[0028]).

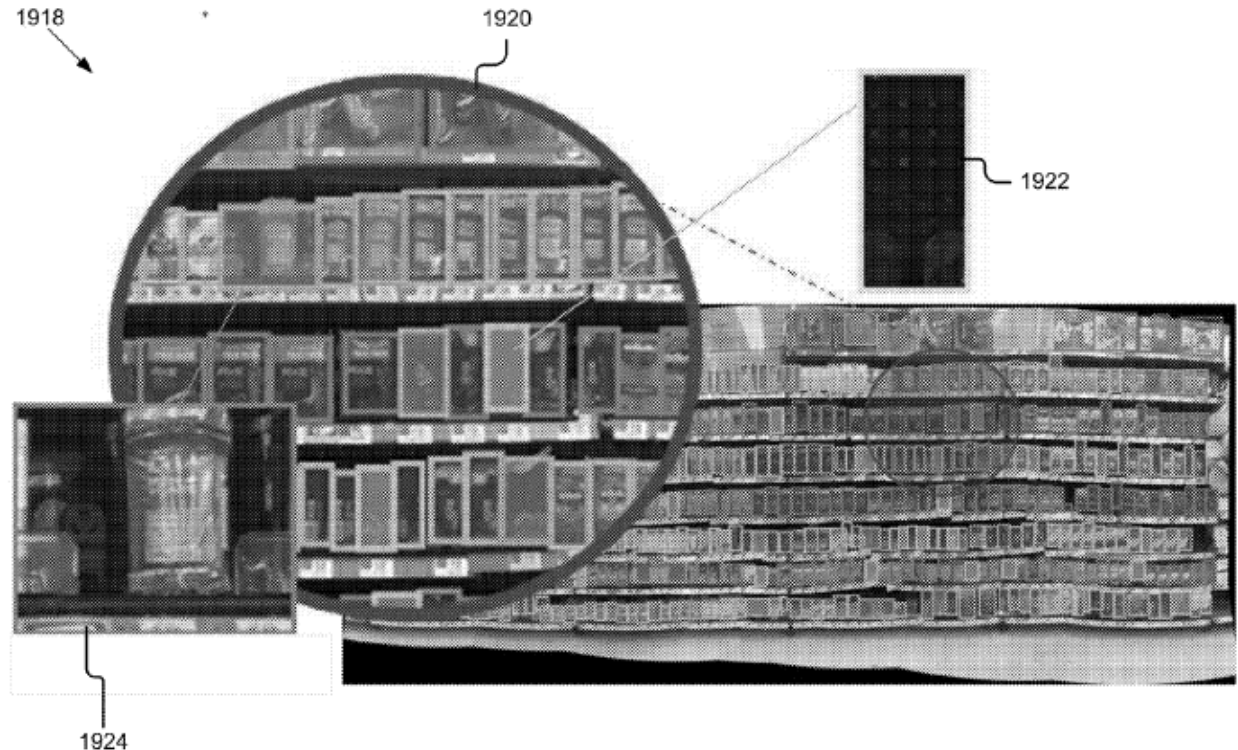
206. Claims 3 and 15 recite using manual input to identify potential inventory. Similarly, claim 5 recites manually associating shelf label content with bounding boxes. While the analyses above cite paragraph [0134], it would have been obvious to a POSITA to have human intervention to make final decisions on determining what labels apply to images. For example, Carbonetto recognizes that manually labeling images was very well known in the art. Carbonetto, 2. While the analysis for claim 4 cites paragraph [0135], the limitation is met even without that disclosure by paragraph [0090].

C. Ground 2: Obviousness over Schwartz, Katircioglu, and Avegliano '430 (Claims 2, 14)

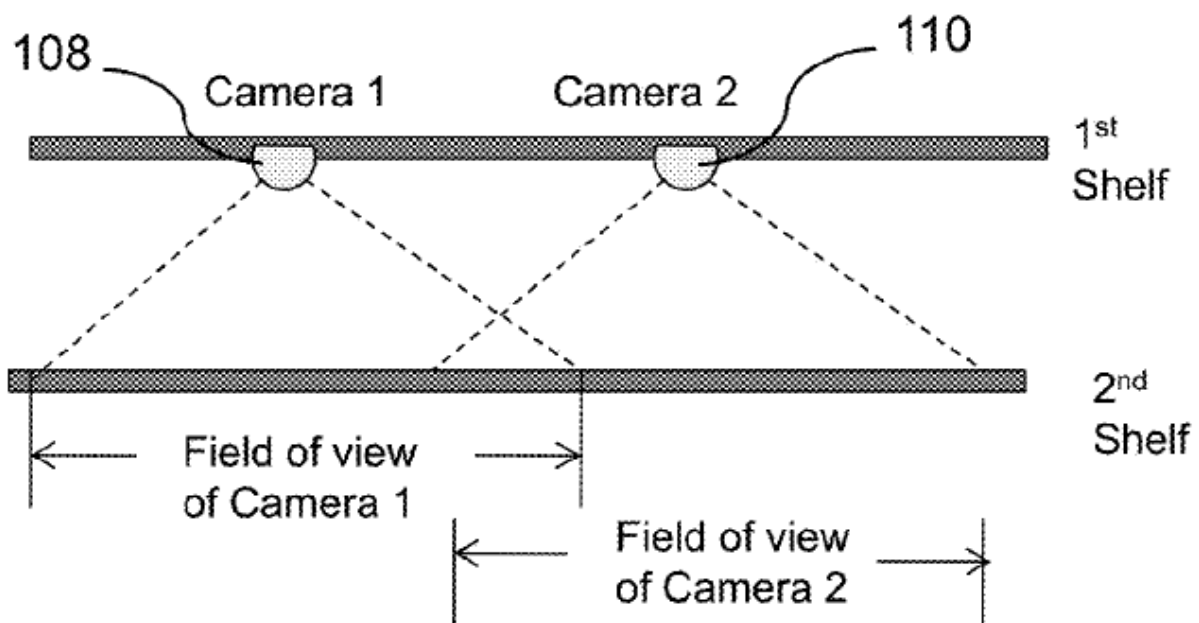
1. Motivation to Combine Schwartz, Katircioglu, and Avegliano '430

207. A POSITA would have been motivated to further modify Schwartz alone or the Schwartz-Katircioglu combination with Avegliano '430. Schwartz alone and in combination with Katircioglu disclose an automated, camera-capture and image recognition solution that reduce the latency of restocking stemming from

manual inventory auditing and shelf checking. *Schwartz*, [0005]-[0006]; *Katircioglu*, [0003]-[0004]. Schwartz generates a data stream of shelf inventory images captured by cameras in multiple stores (even competitor stores) at multiple times to create an expansive library of image data useful for indexing known and unknown products in a database. *Schwartz*, [0081]-[0086]. The systems of Schwartz alone and in combination with Katircioglu use lower elevation, user-held cameras and shelf-installed cameras that can be obscured by normal retail employee and customer foot and cart traffic from the vantage points shown in the figures that I reproduced and combined below. *Schwartz*, [0049], [0128], [0130], [0151]; *Katircioglu*, [0022], [0024].

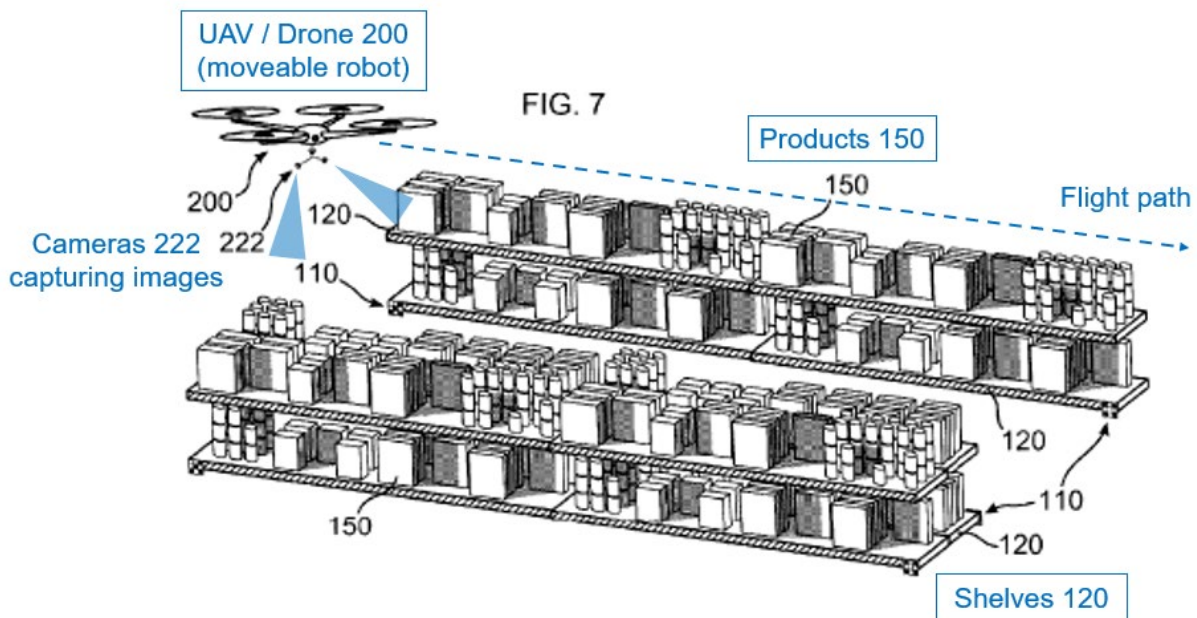


Schwartz, Fig. 19D (showing plan view of retail products and shelves filtered by a edge detection methods).



Katircioglu, Figs. 1A-B (combined).

208. A POSITA would have looked to Avegliano '430 to solve this problem because it flies an unmanned aerial vehicle (UAV), or drone 200, carrying cameras 220 and depth measuring devices 222 along rows of displayed products to gather image data. *Avegliano* '430, [0034]-[0039]. As depicted below, the drone flies at an angle above foot and cart traffic, allowing it to monitor inventory at any time (store open or closed), in intervals (e.g., to re-capture an image if it is unclear), and up and down aisles that may have more foot traffic due to popularity (e.g., seasonal items). *Id.*, [0039].



Id., Fig. 7 (annotated).

209. A POSITA would have been motivated to make this combination to provide real-time stocking level information for quicker replenishment, which creates more retail sales and resulting revenue. *Id.*, [0050].

210. A POSITA would have had a reasonable expectation of success in applying Avegliano '430's drone camera-capture method and system to the Schwartz-Katircioglu combination because Avegliano '430 teaches both embodiments as viable alternatives. *Avegliano '430*, [0022], [0034].

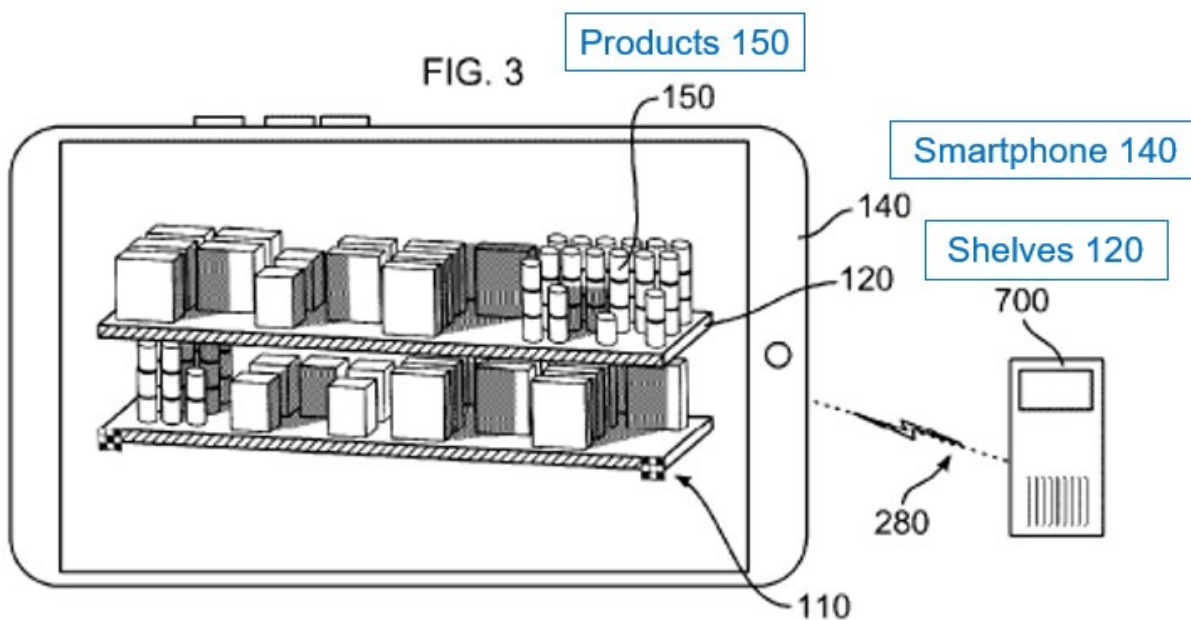
211. A POSITA would have also understood that drone capture or client device capture would be useful in different retail stores selling different products based on type, size, and demand of product in a given retail environment (e.g., drone image capture in dense supermarkets requiring increasingly quick and precise restocking information for high-demand, high turnover, small food items and employee capture in large home improvement stores with lower-demand, large and low-turnover, e.g., lawn mower products easily monitored by roving employees).

1. Claims 2, 14

- a. Claim 2: “The method of claim 1, wherein capturing one or more images comprises capturing the one or more images with a camera attached to a moveable robot.”**

- b. **Claim 14: “The method of claim 10, wherein capturing one or more images of a portion of a shelf with a camera comprises capturing the one or more images with a camera attached to a moveable robot.”**

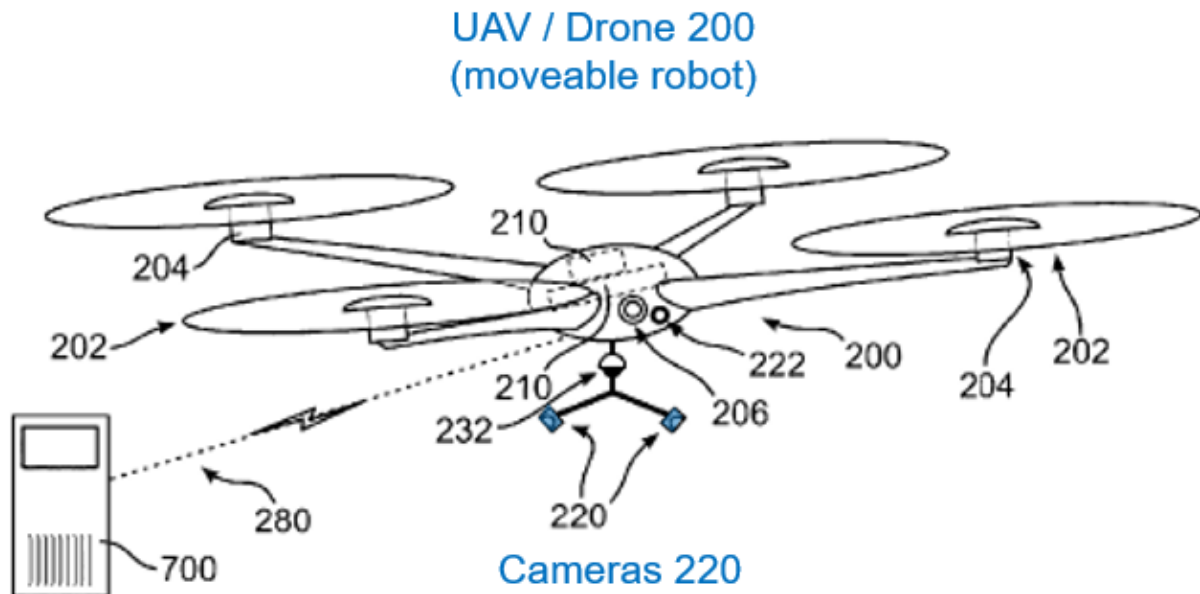
212. Schwartz alone, or in combination with Katircioglu, teaches independent claims 1 and 10 as I explained above. *See* Sections XI.A.2, XI.A.3. In my opinion, Avegliano '430 teaches the additional features of claims 2 and 14. In an embodiment, Avegliano '430 discloses capturing an image of a shelf section using a smartphone or other portable computing device camera as shown in my annotated version of the figure I reproduced below. *Avegliano '430*, [0022].



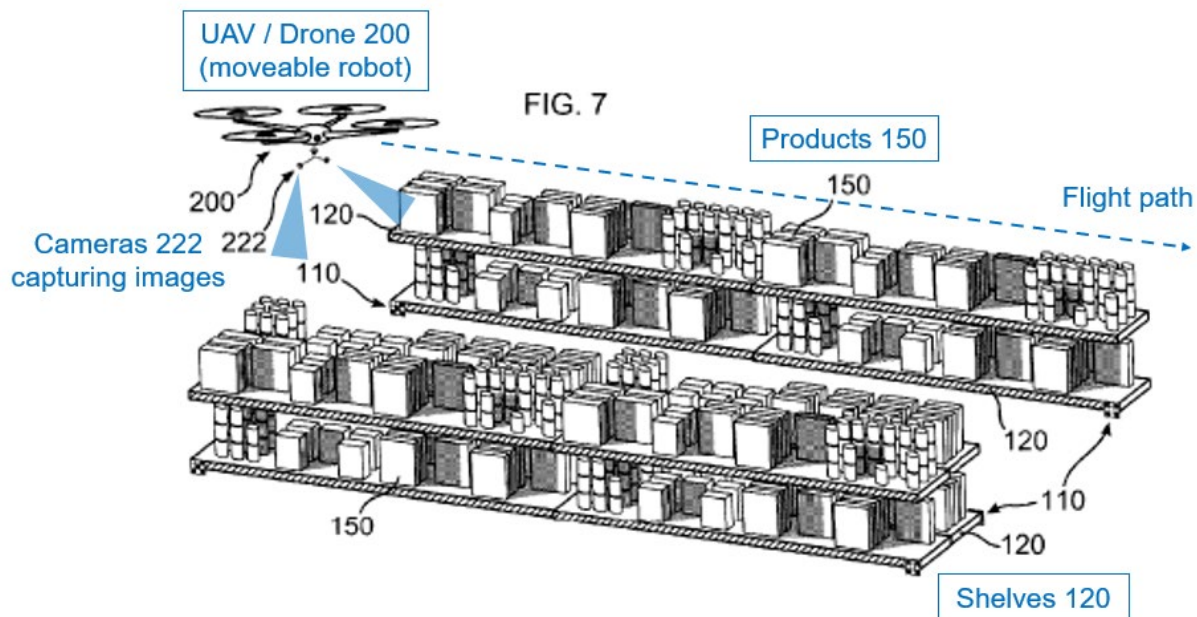
Id., Fig. 3.

213. However, Avegliano '430 also discloses an alternative embodiment that avoids the need for an “employee [to] carry out such image gathering” by flying

an unmanned aerial vehicle (UAV) or drone 200 carrying cameras 220 and depth measuring devices 222 along rows of displayed product and gathering image data. *Id.*, [0034]-[0039]. I show this in my annotated versions of the figures I reproduced below.



Id., Fig. 5.



Id., Fig. 7.

214. Avegliano '430's UAV/drone further includes the control, navigation, and communication circuitry, cameras, sensors, electronics, and circuitry to operate remotely, semi-autonomously, or fully autonomously. *Id.*, [0035], [0037], [0039]. Specifically, the UAV/drone “can conduct flights either autonomously or under human control.” *Id.*, [0039].

215. Accordingly, it is my opinion that Avegliano '430 teaches gathering images with cameras attached to a UAV or drone (“camera attached to a moveable robot”). It would have been obvious to combine Avegliano '430 with Schwartz and Katircioglu for the reasons described in Section XI.C.2.

D. Ground 3: Avegliano '430 in view of Avegliano '157 and Rodriguez Renders Obvious Claims 1, 3-6, 8-10, 12-13, 15-18, and 20-21

1. Motivation to Combine Avegliano '157 and Rodriguez

216. In my opinion, a POSITA would have been motivated to combine Avegliano '157 and Rodriguez. Both references describe the need to identify products in a real-world retail environment. *Avegliano '157*, Abstract, [0016], Figs. 3-4; *Rodriguez*, Abstract, [0009], [0081], [0086]. Both accomplish this task by capturing an image featuring products, such as products placed on shelves in a retail store, with a camera and processing this image. *Avegliano '157*, Abstract, [0016], [0019], Figs. 3-4; *Rodriguez*, [0009], [0186]. Additionally, both references propose identifying products by reading a tag, including tags attached to the shelves and barcodes located on the products. *Avegliano '157*, [0016], [0018]-[0019]; *Rodriguez*, [0009], [0166].

217. Rodriguez discloses capturing images either at checkout or in the aisles of the store. *Rodriguez*, [0009], [0186]. Rodriguez also discloses identifying a product by its barcode. *Id.*, [0009]. A POSITA would have recognized that while the barcode is necessarily seen at checkout, because it must be scanned for purchase, the barcode on the product is not necessarily visible when the product is on the shelf in the store aisle. Therefore, an alternative product identifier would be beneficial to enhance the system disclosed by Rodriguez.

218. In Avegliano '157, the products are identified in captured images by matching objects in the captured images to product templates that were retrieved from a database based on contents of the shelf tags. *Avegliano '157*, [0042]. Therefore, Avegliano '157 teaches an alternative product identifier that a POSITA would have been motivated to add to the system disclosed by Rodriguez. In the combined Rodriguez-Avegliano '157 system, shelf images would be captured in the aisles, and the particular products would be identified by the shelf labels, similar to identifying the products by barcode at checkout. The images of the products associated with the shelf label—as determined by the disclosure in Avegliano '157—would then be used to train a classifier to recognize those products.

219. Although Avegliano '157 discloses using height and width of the product to determine coordinates of a bounding box that surrounds the product facing, which is further used to identify the product, Avegliano '157 does not expressly disclose any detailed image processing algorithms or techniques to complete the matching and identification processes. *Id.*, [0062]-[0063], Figs. 3-7.

220. A POSITA would have looked to Rodriguez for help with this problem. Rodriguez discloses machine-learning based image processing techniques to identify and recognize objects efficiently and reliably. *Rodriguez*, Abstract, [0019], [0159]. It was well known to a POSITA that machine-learning based image processing is a popular solution for object identification in images. Indeed, as

Rodriguez discloses, “[d]eep learning is emerging—and machine learning and neural networks are re-emerging—as promising solutions to a great variety of problems,” including in machine identification of retail store imagery. *Rodriguez*, [0019], [0073]-[0076]. Specifically, Rodriguez discloses machine learning techniques and teaches details of object identification processes in image processing, such as using feature points, weighting functions, and filtering techniques to identify products reliably and accurately in images captured in supermarkets. *Rodriguez*, [0083]-[0092], [0127]-[0131], Figs. 6, 7A-7F.

221. Therefore, in my opinion, combining Avegliano ’157 and Rodriguez would have used a known technique (e.g., object identification based on machine learning) to improve the similar Avegliano ’157 method (e.g., object identification process for images captured in supermarkets, using shelf labels for identification) in the same way (e.g., to ensure reliable and accurate object identification).

222. Also, as noted above, like Rodriguez, Avegliano ’157 discloses capturing images of products in retail stores and reading tags to identify the products. *Avegliano ’157*, [0016]; *Rodriguez*, [0009]. Avegliano ’157 discloses that the images can be captured using either a mobile application designed for stockists or another mobile application for consumers to report that a product is missing on the shelf. *Avegliano ’157*, [0046]. Both mobile application capture images of supermarket aisles using a regular mobile device and during normal operations of a supermarket,

e.g., under typical uneven illumination conditions. A POSITA would have recognized that images captured under these circumstances can be of low quality, which in turn can affect the accuracy of the product identification process. Rodriguez recognizes this issue and provides a solution. According to Rodriguez, “[a]n object of certain aspects of the present technology is to improve object recognition from **degraded imagery**.” *Rodriguez*, [0078]. In particular, Rodriguez discloses that its object identification process works with “images that may be degraded by blur, inferior lighting, etc.” *Rodriguez*, Abstract, [0009], [0161]-[0165]. Rodriguez accomplishes this by training the system on images that are themselves low-quality. *Rodriguez*, [0009]. While Rodriguez focuses on capturing such images at checkout, a POSITA would have understood that capturing images in the aisles would provide valuable training data, where the ultimate product classifier is meant to identify products in that same environment.

223. Therefore, in my opinion, modifying Avegliano ’157 based on Rodriguez would have used a known technique (e.g., object identification in low-quality images) to improve the similar Avegliano ’157 method (e.g., object identification process for images captured in less than ideal conditions) in the same way (e.g., to ensure reliable and accurate object identification).

224. Furthermore, using real-world images of retail products as training data for a classifier was well within the skill of a POSITA and would have required

nothing more than the application of known, conventional techniques (building a classifier with images of retail products as the training data) to yield predictable results (classifying retail products located on shelves using the classifier).

2. Independent Claim 1

a. Claim 1.0: “A method comprising:”

225. If limiting, it is my opinion that Avegliano '157 discloses or suggests limitation 1.0. Avegliano '157 discloses “a method for calculating sales trend of a product at a store shelf” including steps of receiving “picture acquired of the product on the shelf [and] identifying products on the shelf using tags attached to the shelves.” *Avegliano '157*, Abstract, [0016]. Specifically, Figure 3 of Avegliano '157 is “a method...for identifying products on tagged shelves.” *Id.*, [0062]. Figure 3 is reproduced below.

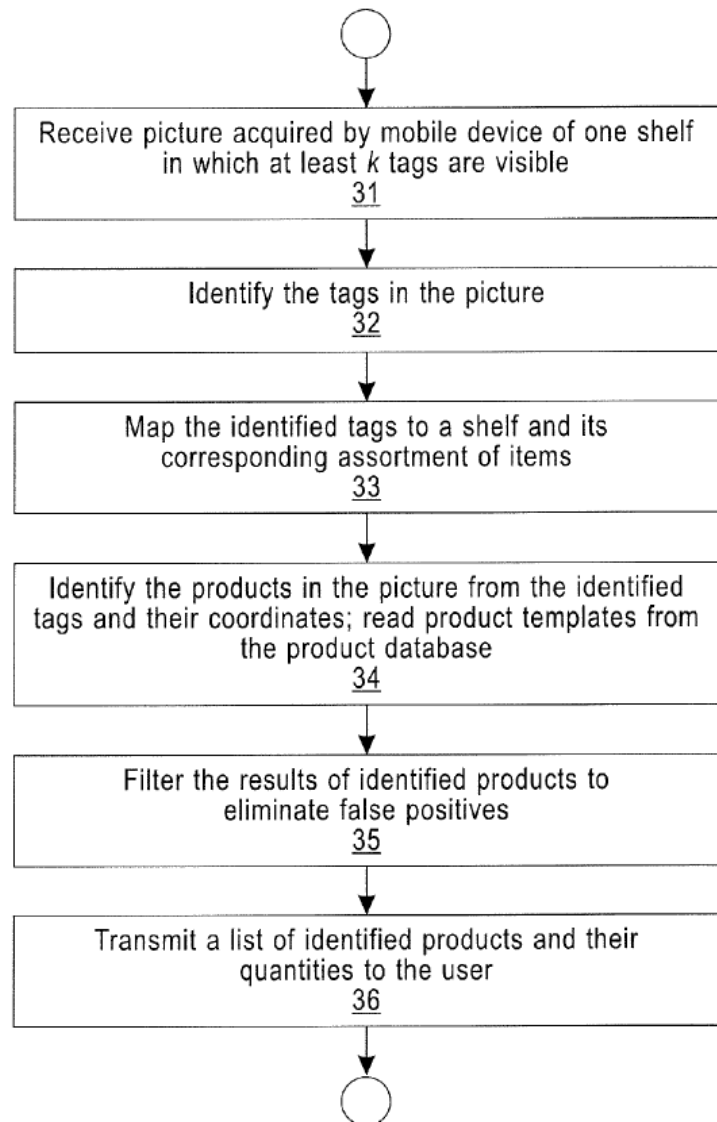


FIG. 3

Id., Fig. 3.

226. An image processing module of a retail store server 12 calls multiple applications in image processing module 17 for capturing and processing images, such as a first image processing application for identifying tags and a second image processing application for identifying products in the pictures. *Id.*, Abstract, Figs. 1-

3, [0019], [0056]. Figure 1 of Avegliano '157 is reproduced below to show server 12 (blue) and image processing module 17 (red).

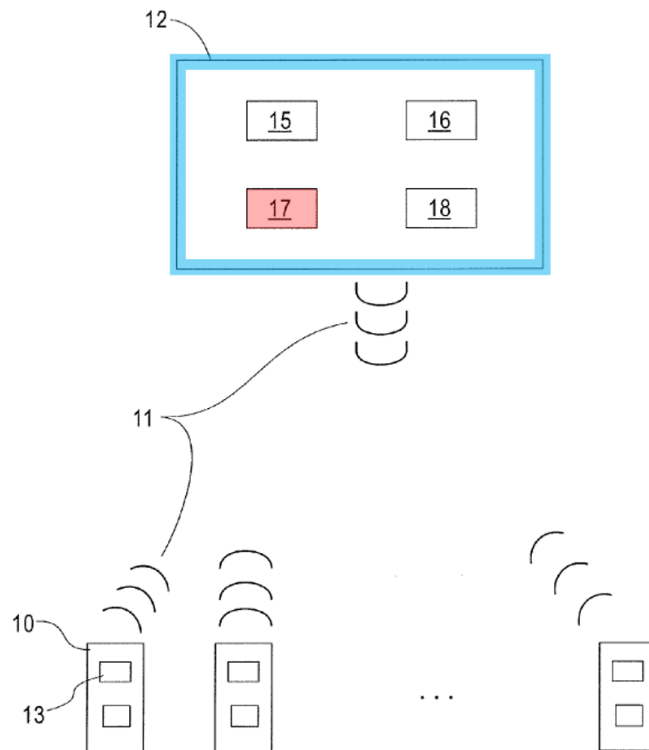


FIG. 1

Id., Fig. 1.

227. Accordingly, it is my opinion that Avegliano '157 discloses or suggests a method comprising steps.

b. Claim 1.1: “capturing one or more images of a portion of a shelf with a camera;”

228. In my opinion, Avegliano '157 discloses or suggests limitation 1.1. Avegliano '157 provides “a method for calculating sales trend of a product at a store shelf” by receiving “availability data of a product measured on a shelf in the retail

store from a portable device.” *Avegliano '157*, [0016]. The availability data is “in the form of **a picture acquired of the product on the shelf.**” *Id.* Referring to Figure 3 below, at step 31 “**a picture** is acquired by the mobile device of **one shelf** in which at least k tags are visible, and is transmitted to the remote server.” *Id.* [0062].

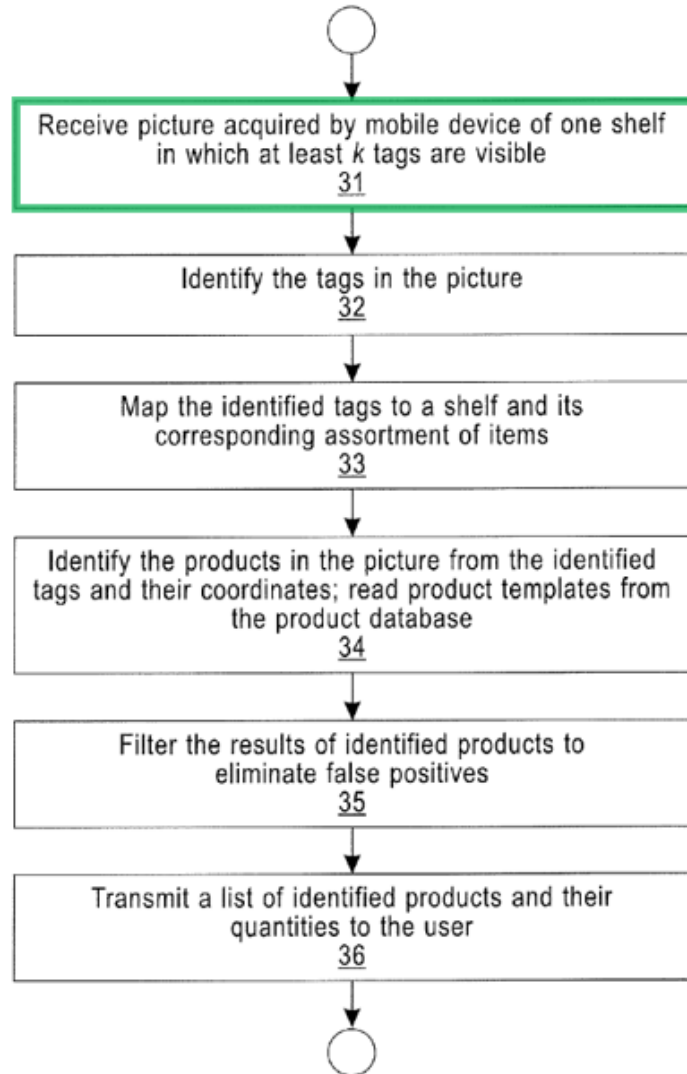


FIG. 3

Id., Fig. 3 (step 31) (annotated). The captured image contains a shelf rack S1 with different types of products A-D and multiple tags T1.1-T1.4 (red) “attached to each store shelf rack.” *Id.* [0056], [0063].

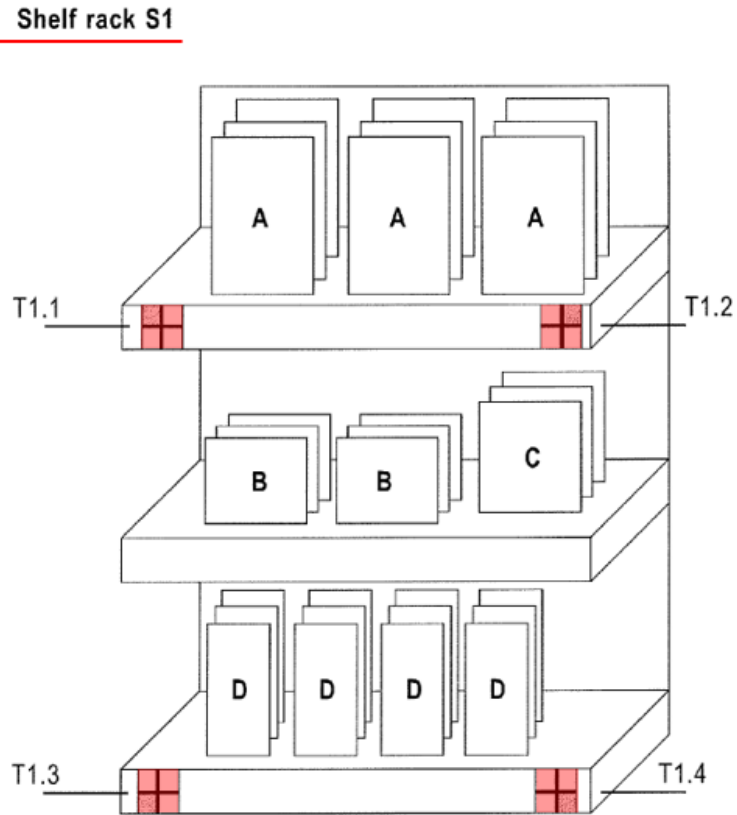


FIG. 4

Id. Fig. 4. The images are captured using “a camera application on the mobile devices that can take pictures of the shelves and transmit them to the server.” *Id.* [0056], [0060].

229. Accordingly, it is my opinion that Avegliano '157 discloses or suggests capturing a picture acquired of the product on the shelf (“one or more images of a portion of a shelf”) using a camera application on a mobile device.

c. Claim 1.2: “detecting and reading content of a shelf label on the portion of the shelf within the one or more images;”

230. In my opinion, Avegliano '157 discloses or suggests limitation 1.2. Avegliano '157 discloses placing sets of tags on supermarket shelves and the tags are “distributed so that whenever a picture of a shelf is taken, at least k tags are displayed in the picture. An exemplary, non-limiting value of k is 3.” *Id.*, [0060]. A variety of tags can be used, such as “low[-]cost paper models, each with a distinctive pattern.” *Avegliano '157*, [0060]. The unique pattern printed on the tag “enables the tag to be **mapped to its location**.” *Id.*, [0063]. Specifically, “[w]hen a user takes a picture of an arbitrary shelf, the first image-processing application identifies the current shelf from the tags.” *Id.*

231. Referring again to Figure 3, at step 32 the remote server calls a first image processing application to identify tags within the images by “us[ing] tag template and distribution information read from the tag database in identifying the tags.” *Avegliano '157*, [0062]. At step 33, the second image processing application uses the identified tags, their coordinates, and the captured image, “to map the tags to a shelf and its corresponding assortment of items.” *Id.* I have annotated and

reproduced Figure 3 of Avegliano '157 below to highlight steps 32 and 33 of the flow chart.

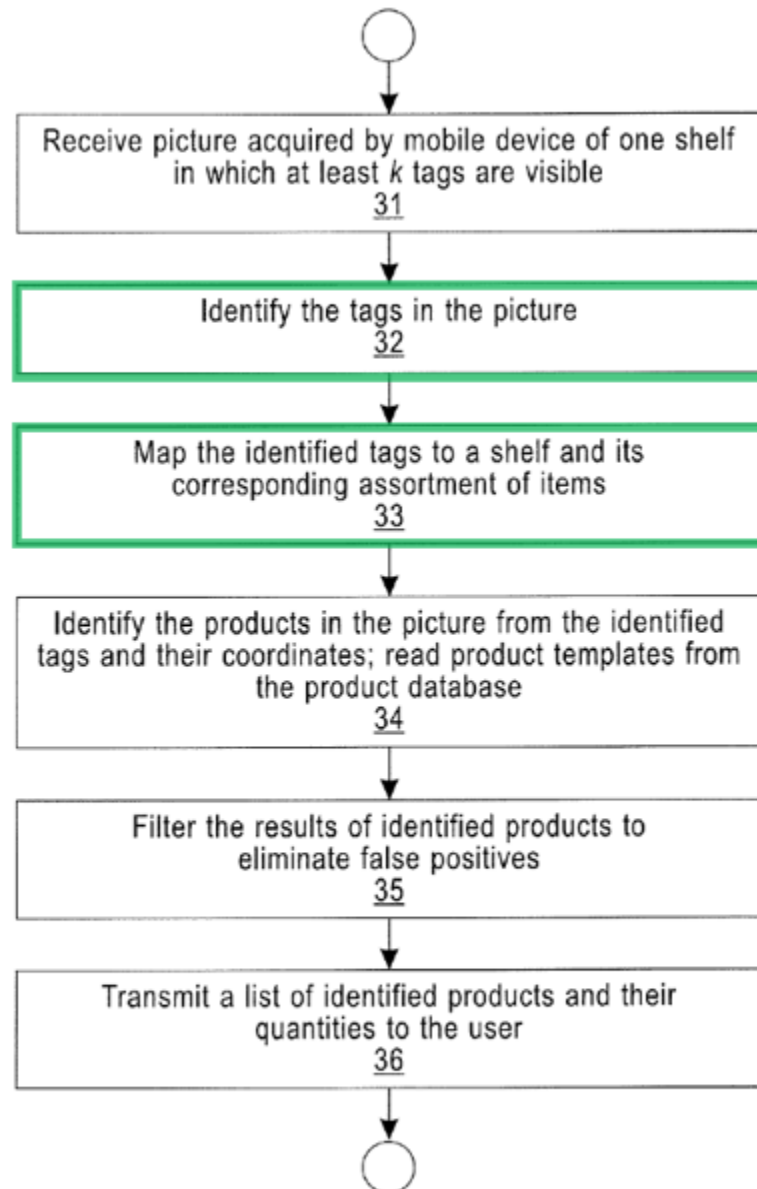


FIG. 3

Id., Fig. 3.

232. Accordingly, it is my opinion that Avegliano '157 discloses or suggests identifying tags based on the tag's unique pattern ("detecting and reading content of a shelf label") from a tagged shelf rack within an image ("portion of the shelf within the one or more images").

d. Claim 1.3: "defining one or more product facing bounding boxes surrounding one or more corresponding facings of a product within the one or more images;"

233. In my opinion, Avegliano '157 discloses or suggests limitation 1.3. As I explained in Section VIII.D.4.b above, an image of tagged supermarket shelves captured by using a mobile device includes a plurality of shelf tags and aisle-facing products. Figure 4 of Avegliano '157 shows a shelf rack S1 with tags (red) and four types of products A, B, C, and D (yellow) placed on three shelves. *Avegliano '157*, [0063]. A POSITA would have recognized that multiple identical products of the same product type are placed on the shelves with the full face of the front-most product visible as viewed from the aisle. I have annotated and reproduced Figure 4 of Avegliano '157 below.

Shelf rack S1

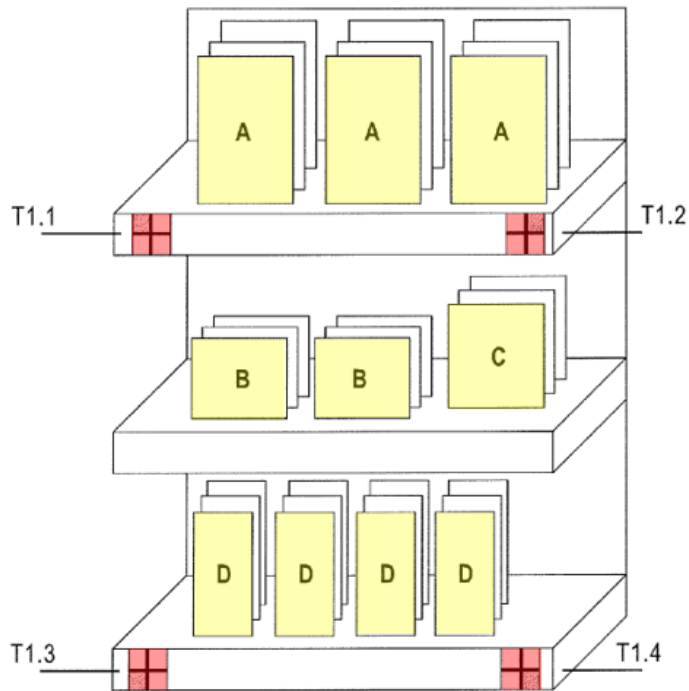


FIG. 4

Id., Fig. 4.

234. After the user “takes a picture of an arbitrary shelf, the first image-processing application identifies the current shelf from the tags, and then retrieves its list of products associated with the current shelf from the product database.” *Avegliano '157*, [0063]. I note that most importantly, information on each product’s width and height is also retrieved. *Id.*, Fig. 3. I have annotated and reproduced Figure 3 of *Avegliano '157* below,

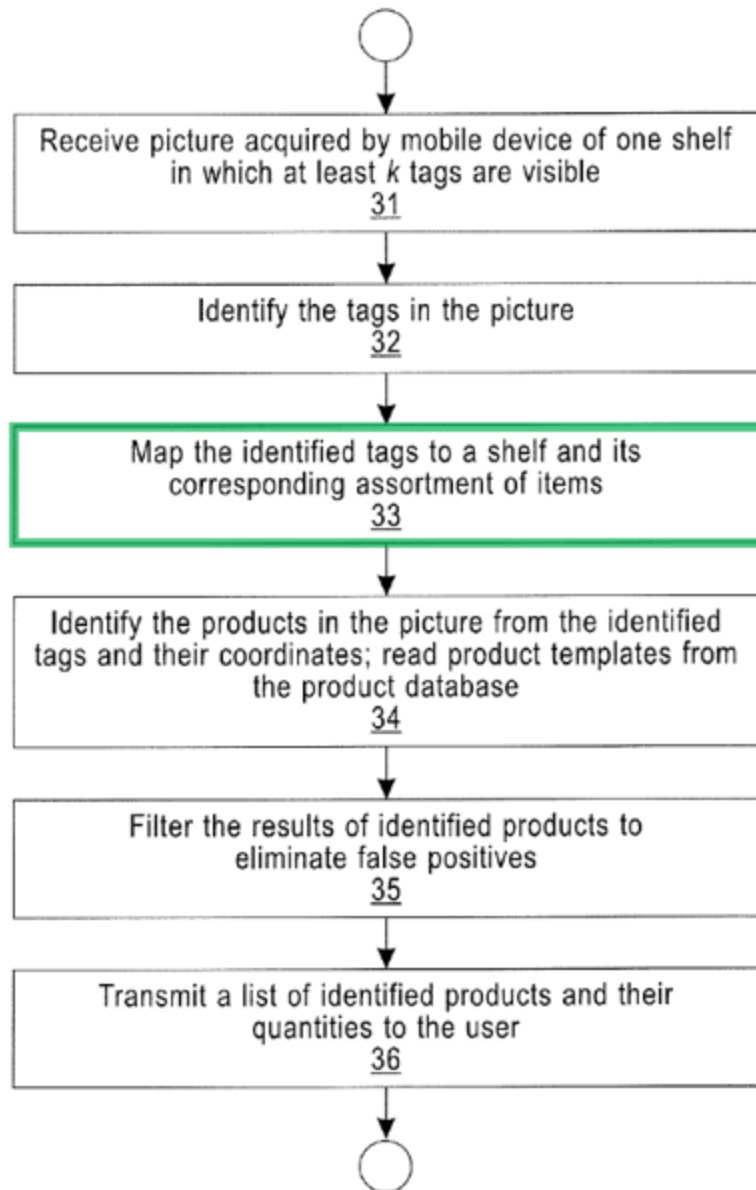
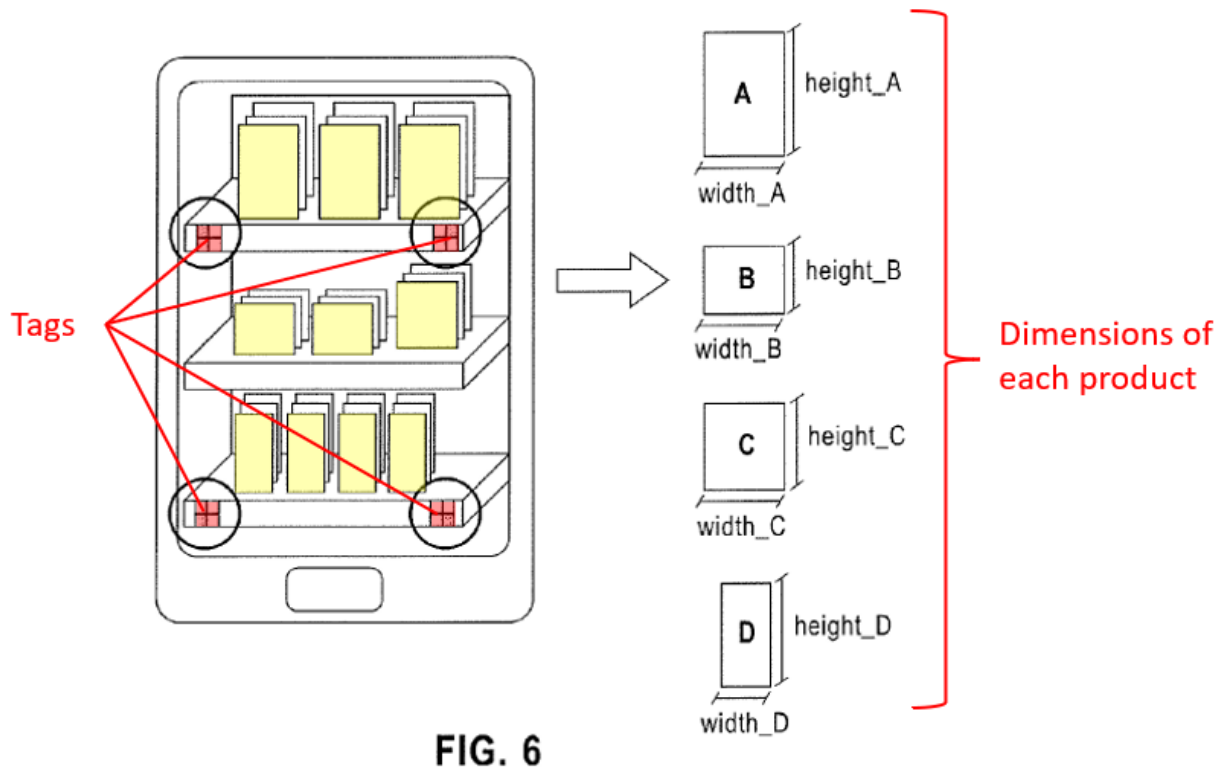


FIG. 3

Id., Fig. 3.

235. Avegliano '157's system depicted in Fig. 1 includes a first database 15 that stores "product templates that include, inter alia, the product name, picture and dimensions, and tag templates and their distribution on the store shelves." *Id.*, [0060].

As shown in Fig. 6 below, the height and width refer to dimensions of the product facing defined by its horizontal and vertical boundaries, which may be different across different types of products. *Id.*, [0040]-[0041], [0056], [0063], Fig. 6. I have annotated and reproduced Figure 6 of Avegliano '157 below. In Figure 6, I note that product A has a height that is greater than products B and C, while product D has a width that is less than the widths of products A-C.



Id., Fig. 6.

236. Using the height and width information for each one of the assortment of products on the shelf, “the second image-processing application attempts to match the template to every item from the list in the image using the height and width

information, as shown in FIG. 7.” *Avegliano '157*, [0063]. According to *Avegliano '157*, using the product’s dimension information to identify products within the image can “increase recognition accuracy and [] filter false positives.” *Id.*, [0061]. In the example shown in Fig. 7, height and width of product A are used to detect product A and coordinates of its product facing within the image. *See id.*, [0062]-[0063], Fig. 7. Figure 7 is annotated and reproduced below.

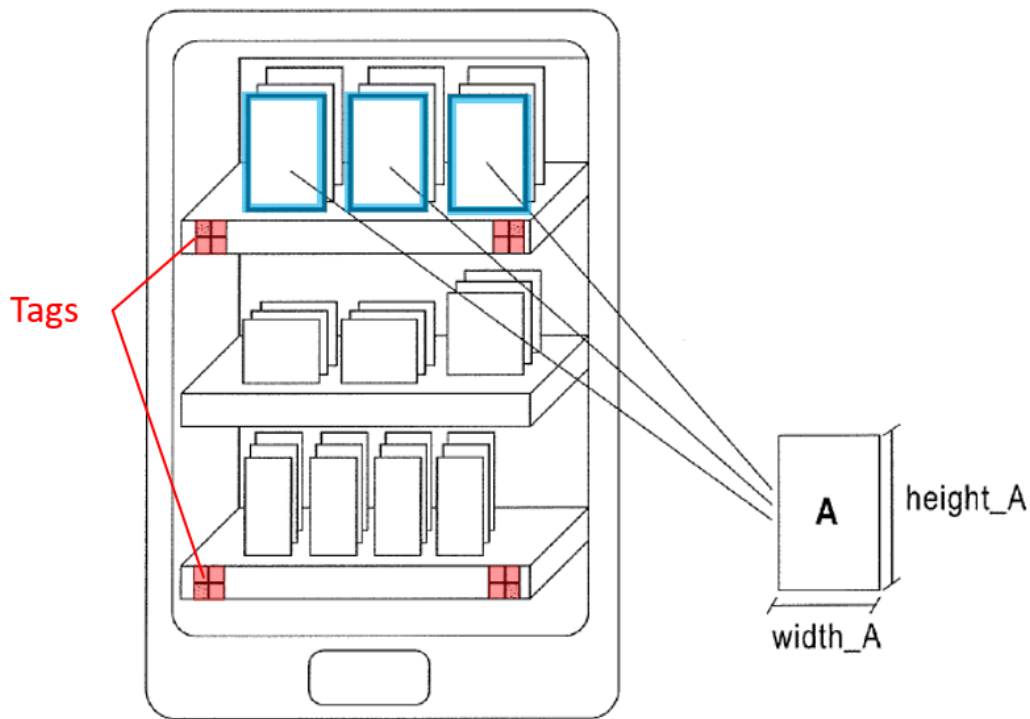


FIG. 7

Id., Fig. 7.

237. The determined coordinates for product A define a boundary (blue box, bold in original) within the image that contours and surrounds its product facing. In

the above example, the determined coordinates of product A within the image forms a rectangle box that surrounds the product facing of product A. In addition, a POSITA would have understood that the coordinates in an image form a rectangular boundary surrounding the product because the coordinates are mathematically related to the actual height and width of the product facing. The proportional relationship depends on the distance and angle between the product and the camera. Indeed, Avegliano '157 discloses that “knowing the distance between the tags in the picture and in the actual shelf, can help the product identification method.” *Id.*, [0052]. A POSITA would have understood that the same mathematical relationship between the tag separation depicted in the picture and as measured on the actual shelf also applies to the dimensions of the product in the captured image and the actual dimensions of the product.

238. Accordingly, it is my opinion that Avegliano '157 discloses or suggests determining coordinates of a rectangular boundary (“defining one or more product facing boundary boxes”) that surrounds a product facing within the captured images (“facings of a product within the one or more images.”)

- e. **Claim 1.4: “associating the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;”**

239. In my opinion, Avegliano '157 discloses or suggests limitation 1.4.

240. As I explained in Sections XI.D.2.c, XI.D.2.d, Avegliano '157 discloses identifying shelf tag content to determine the location and products stored on the shelf, receiving product dimension information to determine coordinates that represent a rectangular bounding box surrounding a product facing. *Avegliano '157*, [0020], [0029], [0061]-[0062], Figs. 3-7. Therefore, the shelf tag is associated with the rectangular bounding box that surrounds a product facing, as further explained below.

241. Specifically, Avegliano '157 teaches that “a set of items can be mapped to a shelf, and each set of tags is mapped to a shelf, to define which products are displayed in a rack of shelves.” *Avegliano '157*, [0060]. The method of identifying product on a supermarket shelf includes first identifying shelf tags in the captured images using patterns on the tag. *Id.*, [0063], Fig. 3 (step 32), Fig. 4. I have annotated and reproduced Figure 3 of Avegliano '157 below.

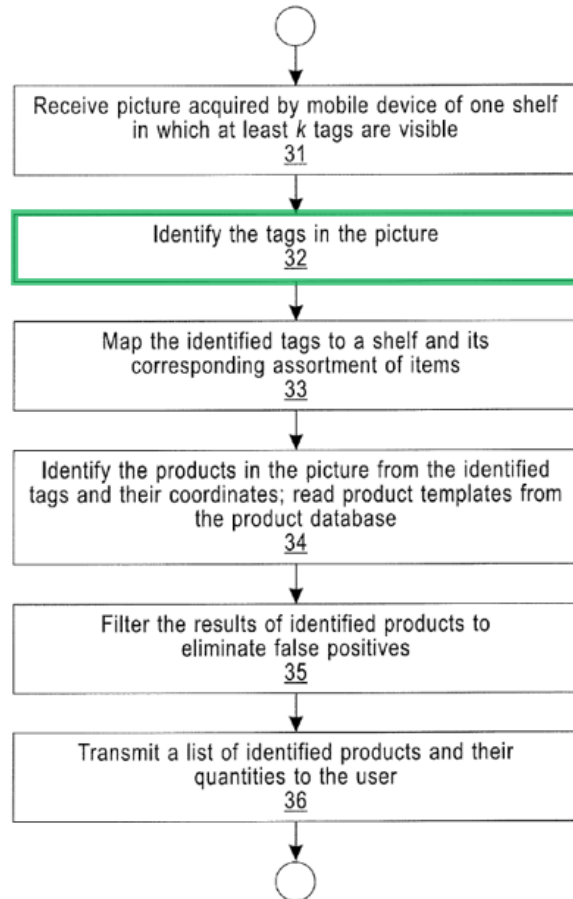


FIG. 3

Id., Fig. 3.

242. The identified shelf tags are then mapped to a certain shelf and its corresponding assortment of items. *Id.*, [0063], Fig. 3 (step 33), Figs. 5-6. In step 34, based on the identified corresponding assortment of items, dimensions of products (e.g., height and width of the product facing) from the assortment of items are used to determine—based on the received dimensions—coordinates that represent a rectangular boundary of the products in the captured image. *Avegliano '157*, [0016]-[0020], [0027],[0029], [0061]-[0063], Figs. 3-7. According to *Avegliano '157*, using

product dimension to identify products can provide the benefit of eliminating false positives. *Id.*, [0020], [0029], [0061]-[0062], Fig. 3 (step 35).

243. Accordingly, it is my opinion that Avegliano '157 discloses or suggests associating the identified shelf tags (“associating the shelf label content”) with the coordinates of the rectangular boundary of the product facing (“product facing bounding boxes that includes and defines each of the one or more corresponding product facings”) identified in an image of the shelf.

- f. Claim 1.5: “cross-correlating the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;”**

244. Avegliano '157 discloses or suggests limitation 1.5. As I explained in Section XI.D.2.e, Avegliano '157 discloses capturing and identifying shelf tags to determine corresponding products, and using product dimensions of the corresponding products to detect coordinates that represent rectangular boundaries of corresponding products.

245. Avegliano '157 discloses or suggests temporal aggregation of retail product and affiliated shelf label information to ensure accurate determination of sales velocity, sales trend, real-time inventory status, and minimizing out-of-the-shelf products. Avegliano '157 recognizes that there is a need for “real time visibility

on consumption and consumption speed information of [MCP⁸]'s products at distribution points and the retailer shelf.” *Avegliano* ’157, [0006]. To address this need, the method and system of Avegliano ’157 can determine product availability data by identifying products on shelves and use the availability data to calculate **sales velocity and sales trends** of products as well as minimize out-of-the-shelf products. *Avegliano* ’157, Abstract, [0016], [0021], [0030], [0033], [0046]-[0048], [0057]-[0059], [0065]. The calculation of sales velocity and trends entails the use of “[a] mobile application, deployed on customers and merchant devices to **periodically** monitor product sales velocity.” *Id.*, [0048]. According to Avegliano ’157, “[t]wo kinds of mobile applications can be used: one, designed for stockists, can be used to collect data regarding the number of products in the shelf and turnover of products **in a period of time**; and another for consumers can be used to report that a product is missing on the shelf.” *Id.*, [0046]. The usage of both mobile applications results in receiving multiple images of product shelves through a period of time.

246. A POSITA would have recognized that in determining sales velocity and sales trend over “a period of time” would entail *cross-correlating* product and

⁸ MCP refers to “manufacturers, consumer goods companies and producers.”

Avegliano ’157, [0006].

affiliated shelf tag information *across* images taken at different points of time. *See Avegliano '157*, [0016], [0021], [0030]-[0033], [0046]-[0048], [0057]-[0059].

247. Avegliano '157 also discloses or suggests spatial aggregation of retail product and affiliated shelf label information to ensure accurate determination of sales velocity, sales trend, and real-time inventory status across multiple agents on the retail sales platform. According to Avegliano '157, its method's "[o]ptimization and visualization have the potential to further improve the quality of a platform." *Avegliano '157*, [0050]. Specifically, Avegliano '157's method of determining sales velocity and trends also benefits "scenarios with **several retailers and several producers** [that] may have distribution plans properly modeled as Multi-commodity Flow problems." *Id.* As an example, "techniques dedicated to the Vehicle Routing Problem can employ knowledge about **sales velocity** to generate replenishment for merchants." *Id.*

248. A POSITA would have recognized that in Avegliano '157's "vendor-managed inventory (VMI) business model," achieving improved efficiency in a multi-retailer scenario would entail *cross-correlating* product and affiliated shelf tag information *across* multiple retailers, store types, and time because this enables a more accurate determination and evaluation of sales trends and sales velocity across multiple stores in a geographical region. *See Avegliano '157*, [0011]-[0013], [0056].

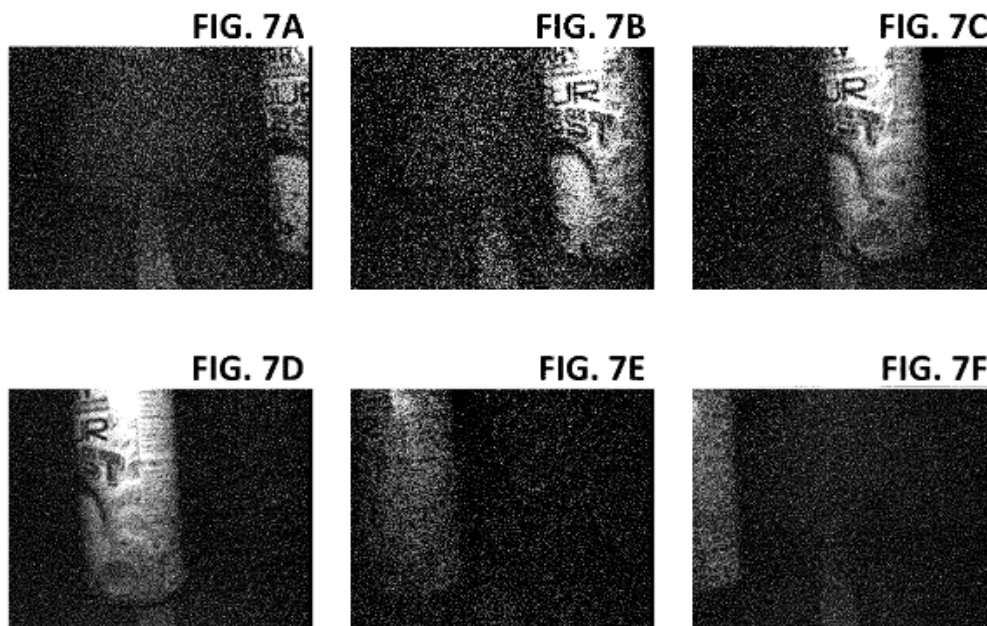
249. A POSITA would have understood that retailer chains would have desired inventory optimization across their subsidiary retail locations to best track and inform global manufacturer and supplier distributors while optimizing regional sales and local inventory auditing processes, all of which the combination of Avegliano '157 and Rodriguez's method would have enabled.

250. Further, a POSITA would have understood that the same product is often positioned in multiple locations of a single store. For example, candy bars may be included in an aisle, as well as near the check-out register. When images of those candy bars are collected, it would have been obvious to a POSITA to keep all of the images of the same product together. The combination of Avegliano '157 and Rodriguez, which associates product images with shelf tag content, would provide such cross-correlation.

251. To the extent that Patent Owner argues Avegliano '157 does not teach this limitation, Rodriguez does, and it would be present in the combination of Avegliano '157 and Rodriguez.

252. Rodriguez teaches that it is advantageous to provide a machine learning system as many training images as possible to improve the "classification accuracy." *Id.*, [0052], [0087], [0159]. Indeed, Rodriguez discloses capturing many images of products over time. For example, the images can be "**a sequence of images** captured by a point-of-sale product scanner during a checkout procedure." *Id.*, [0160]; *see*

also id., [0082] (images [are] captured at “30 frames per second”). Figs. 7A-7F are examples of a series of images obtained when a “can is swept through a supermarket checkout area.” *Rodriguez*, [0089]; Figs. 7A-7F. I note that although the images below are “degraded imagery,” a goal of Rodriguez “is to improve object recognition from degraded memory.” *Id.*, [0078]. I have reproduced Figures 7A-7F from Rodriguez below.



Id., Figs. 7A-F.

253. The captured images contain “machine-readable identifier” such as “a barcode.” *Id.*, [0159]. “Identification data associated with the product” is derived from the images using “a machine recognition technology.” *Id.*, [0160]. Both the captured images and the derived identification data such as bar code data (label data)

are provided to a learning system which learns “**to recognize—and label—the item** from image data.” *Id.*, [0160]; *see also id.*, [0161]-[0162].

254. As such, when “a checkout clerk passes a box of Cheerios cereal in front of such a system, a series of up to a dozen images may be captured,” and after the barcode is identified, **these images are grouped together** because “all of these images are known to depict a box of Cheerios cereal.” *Rodriguez*, [0082]-[0083]. Afterwards, “[e]ach of the images can be tagged with label information indicating that it depicts a box of Cheerios cereal, and can be provided to a learning system as a training image.” *Id.*, [0083]. Rodriguez discloses that these images can be captured at a rate of “30 frames per second.” *Id.*, [0082]. Therefore, a POSITA would have understood that these images are temporally separated. In the example of capturing images of Cheerios cereal boxes, each image of the “series of up to a dozen images” is captured at 1/30 second after the previous image. *Id.* Rodriguez also discloses that multiple individual products (e.g., Cheerios cereal boxes and Sour Blast fruit drinks) are being scanned at Rodriguez’s point-of-sale checkout stations and their images are captured for identification. *Id.*, [0082]-[0084], [0089]-[0091].

255. Therefore, Rodriguez’s learning system is trained to recognize, label, and group (“cross-correlating”) at least a first image depicting a product and tagged with label information (“one or more product facing bounding boxes” and “shelf label content”) and a second image—taken before or after the first image—depicting

a product and tagged with label information (“one or more other product facing bounding boxes across one or more stores” and “other shelf label content”). *Id.*, [0082]-[0084].

256. Rodriguez does not limit applying image grouping to only temporally separated images. Rodriguez also discloses or suggests grouping images captured at different locations. For example, images captured in “**aisles** of a supermarket” of Rodriguez are “processed using barcode... or other techniques [] to identify one or more objects.” *Rodriguez*, [0186]. Additionally, a POSITA would have recognized that Rodriguez’s system captured images at different point-of-sale checkout stations, which are grouped together when customers bring the same product (e.g., Cheerios) to different checkout stations.

257. As I explained in Section XI.D.1, a POSITA would have been motivated to implement Rodriguez’s method of machine-learning based image processing techniques to identify and recognize objects efficiently and reliably. *Rodriguez*, Abstract, [0019], [0159]. A POSITA would have recognized that Rodriguez’s method of capturing many images of the same product over time (e.g., series of images) and/or space (e.g., different checkout lanes or aisles) and providing those images to its learning system can be readily implemented into Avegliano ’157’s method of gathering images over space (taken in aisles). For example, both Avegliano ’157 and Rodriguez disclose or suggest capturing images

in aisles and use label information (e.g., barcodes and shelf tags) to identify objects in the images. Like Avegliano '157, Rodriguez discloses that images of products captured in “aisles of a supermarket” are “processed using barcode... or other techniques [] to identify one or more objects.” *Rodriguez*, [0186]. These images captured in “aisles of a supermarket” also similarly “provided to a system” as training images. *Id.*

258. A POSITA would also have known that the same product is often displayed in multiple locations, such as in aisles, aisle endcaps, near the point-of-sale checkout counter, and/or in different stores. A POSITA would have understood that a person walking through the supermarket, as disclosed by Avegliano '157, would have captured images of the same product displayed at different locations. Capturing images from multiple locations would be beneficial to meeting Rodriguez's goal of capturing as many training images as possible to improve the “classification accuracy.” *Id.*, [0052]. Further, a POSITA would have understood that capturing more training images in an environment similar to that for which the classifier will be used for prediction (i.e., store shelves) will provide more accurate results. Merler, 4 (discussing the benefits of data that “represents the typical low quality of a real[-]world image”).

259. Accordingly, it is my opinion that Avegliano '157 alone or the combination of Avegliano '157 and Rodriguez discloses or suggests spatial-

temporal aggregation of retail product and affiliated shelf tag information to determine sales velocity and trends (“cross-correlating...product facing bounding boxes...across one or more stores by matching shelf label content”).

g. Claim 1.6⁹: “building training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes;”

260. It is my opinion that the Avegliano ’157 and Rodriguez combination discloses or suggests limitation 1.6. As discussed in Sections XI.D.2.e–XI.D.2.f, Avegliano ’157 discloses associating shelf tag content with cross-correlated product facing bounding boxes. Although Avegliano ’157 does not explicitly teach building training data, Rodriguez discloses this, as I explain below.

261. According to Rodriguez, “imagery gathered during normal operation of a retail store is used as **training imagery** for a machine-learning system.” *Rodriguez*, [0081]. In the operation of the training system, “one of these captured images may be selected as the best, and is processed to identify the product (e.g., by decoding the product's barcode information).” *Id.*, [0083]. In addition, each of the images that depict the same item “can be tagged with label information... and can be provided to a learning system as a **training image**.” *Id.* According to Rodriguez, “[a]ll such information can serve as ‘ground truth’ data, to label the product depicted in the

⁹ See *supra* note 5.

sequence of imagery captured by the point-of-sale system, permitting such imagery to serve as **training data** for a supervised learning system.” *Id.*, [0085].

262. A POSITA would have understood that the combination of Avegliano ’157 and Rodriguez uses the association of the product label data or shelf tag as the ground truth for the identification of the products in the captured images to train a machine learning system, to recognize other images of the same product. *Rodriguez*, [0160]-[0163]. Mitchell corroborates that when using this type of training data neural networks are particularly helpful because they are tuned for “training data [that] contains noisy, complex sensor data” such as “inputs from cameras.” *Mitchell*, 83; *see also* Section VII.A.1. A POSITA would have easily recognized that the same utility neural networks that provide for, e.g., “face recognition” using input facial images would apply equally and advantageously to product recognition using input product images. *Mitchell*, 112; *see also* Section VII.A.1.

263. A POSITA would have understood that the building training data process of Rodriguez was a suitable method of accomplishing Avegliano ’157’s goals of accurately and timely identifying products on a retail store shelf.

264. At least because of the reasons described in Section XI.D.1, a POSITA would have been motivated to modify or supplement Avegliano ’157’s object identification system using Rodriguez’s machine-learning based object

identification system to perform object identification with improved accuracy. In particular, the combination of Avegliano '157 and Rodriguez would use the shelf label provided by Avegliano '157 in place of the barcode on the product used by Rodriguez for product identification. The association of product identifier (from the shelf label), as described in Section XI.D.2.d, and the images collected from the bounding boxes, as described in Section XI.D.2.e, provides the training data for the product classifier taught by Rodriguez.

265. Accordingly, it is my opinion that the Avegliano '157 and Rodriguez combination teaches “building training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes.

h. Claim 1.7¹⁰: “using the training data to build a product classifier; and”

266. In my opinion, the Avegliano '157 and Rodriguez combination discloses or suggests limitation 1.7.

267. As I explained in Section VIII.D.4.g, images captured in a retail store “is used as **training imagery** for a machine-learning system.” *Rodriguez*, [0081]. The training system select “one of these captured images” as the best and processes it to identify the product by decoding the barcode. *Id.*, [0083]. Each image that

¹⁰ See *supra* note 5.

depicts the same item “can be tagged with label information... and can be provided to a learning system as a **training image**.” *Id.* “[A]ll such information can serve as ‘ground truth’ data, to label the product depicted in the sequence of imagery captured by the point-of-sale system, permitting such imagery to serve as **training data** for a supervised learning system.” *Id.*, [0085]. Once the object’s identity is known, Rodriguez’s system uses such information “in **training a classifier** (e.g., a machine learning system) to recognize the object from others of the captured images.” *Rodriguez*, Abstract, [0127], [0131], [0134]-[0142], [0182]. Rodriguez’s classifier system includes “a processor and a memory” that “contains software instructions that cause the system to classify input image stimulus into two or more classes.” *Id.*, [0182]. The processors and memory devices in Avegliano ’157 and Rodriguez perform the claimed functions, classifying an item as a product [claim 1]; performing an inventory management activity related to a product [claim 10]; and performing one or more of: (1) an inventory management activity related to a product or (2) classifying an item as a product [claim 18]. The disclosed structure in the Avegliano ’157 and Rodriguez combination is equivalent to the structure disclosed in the ’272 patent, i.e., a robot, set of robots, or cloud server implementing a convolutional, feedforward, or recurrent neural network.

268. Accordingly, it is my opinion that the combination of Avegliano '157 and Rodriguez discloses or suggests training a classifier with the training data (“using the training data to build a product classifier”)

i. Claim 1.8: “using the product classifier to classify an item as the product.”

269. It is my opinion that the combination of Avegliano '157 and Rodriguez discloses or suggests limitation 1.8.

270. As I explained in Section XI.D.2.h, Rodriguez discloses that the trained classifier (e.g., a machine learning system) “to recognize the object from others of the captured images.” *Rodriguez*, Abstract, [0127], [0131], [0134]-[0142], [0182]. The recognized object is a product. *See id.*, [0088]-[0092].

271. Accordingly, the combination of Avegliano '157 and Rodriguez discloses or suggests using a trained classifier (“product classifier”) to recognize the object from others of the captured images (“classify an item as the product”).

3. Independent Claim 10

272. It is my opinion that the combination of Avegliano '157 and Rodriguez discloses or suggests claim 10 for substantially the same reasons as claim 1, as I explained in Section XI.D.2, because the claims are substantively identical. *Compare* claim limitations 10.0-10.8, *with* claim limitations 1.0-1.8.

273. The table below correlates the reasons why the combination of Avegliano '157 and Rodriguez renders obvious claim limitations 10.0-10.8 for the reasons rendering claim 1 obvious.

Claim 10		
Limitation	Substantially Similar Limitation(s)	Section Reference
10.0	1.0	XI.D.2.a
10.1	1.1	XI.D.2.b
10.2	1.2	XI.D.2.c
10.3	1.3	XI.D.2.d
10.4	1.4	XI.D.2.e
10.5	1.5	XI.D.2.f
10.6	1.6	XI.D.2.g
10.7	1.7	XI.D.2.h
10.8	1.8	XI.D.2.i

274. To the extent the discussion of limitation 1.8 in Section XI.D.2.i does not explicitly address “using the product classifier to perform an inventory management activity related to the product” as in 10.8, it is my opinion that is disclosed by the combination of Avegliano '157 and Rodriguez. For example, Avegliano '157 discloses “method steps for managing inventory and sales trends in a retail environment using mobile devices.” *Avegliano '157*, [0030]; *see also id.*, Title, Abstract, [0045], [0056], [0074]. A POSITA would have understood that the product identification and classification process of the Avegliano '157 and

Rodriguez combination would achieve Avegliano '157's goal of "near real time" inventory management and minimizing "out-of-the-shelf conditions." *Avegliano '157*, [0005]-[0009], [0022], [0043], [0047]-[0049], [0055].

4. Independent Claim 18

275. It is my opinion that the combination of Avegliano '157 and Rodriguez discloses or suggests claim 18 for substantially the same reasons as claim 1, as I explained in Section XI.D.2, because the claims are substantively identical. To the extent the discussion of limitation 1.8 in Section XI.D.2.i does not explicitly address "using the product classifier to perform an inventory management activity related to the product" as in 18.10, it is my opinion that is disclosed by the combination of Avegliano '157 and Rodriguez teach it for the reasons immediately above.

276. By exception, claim limitations 18.0-18.2 recite a "system comprising: a processor" and a "system memory coupled to the processor...storing instructions configured to cause the processor" to execute the same steps as recited in claim 1. *Compare* claim limitations 18.3-18.10, *with* claim limitations 1.1-1.8.

277. Avegliano '157 discloses or suggests "a non-transitory program storage device readable by a computer" that when executed by the computer, "perform[s] the method steps for managing inventory and sales trends in a retail environment using mobile devices." *Avegliano '157*, [0030], [0040], [0066]-[0076]. Fig. 9 of Avegliano '157 reproduced below illustrates "an exemplary computer system for

implementing a method for managing inventory and sales trends in a retail environment using mobile devices,” the computer system 91 includes “a central processing unit (CPU) 92” connected via bus 94 to memory 93 that stores Avegliano ’157’s CPU-executable routine and further includes “micro instruction code.” *Id.*, [0074]-[0075], Fig. 9. I have annotated and reproduced Figure 9 of Avegliano ’157 to show the computer system 91 in blue and the CPU 92 in red. Therefore, the combination of Avegliano ’157 and Rodriguez discloses or suggests a computer system 91 (“system”) including a CPU 92 (“processor”).

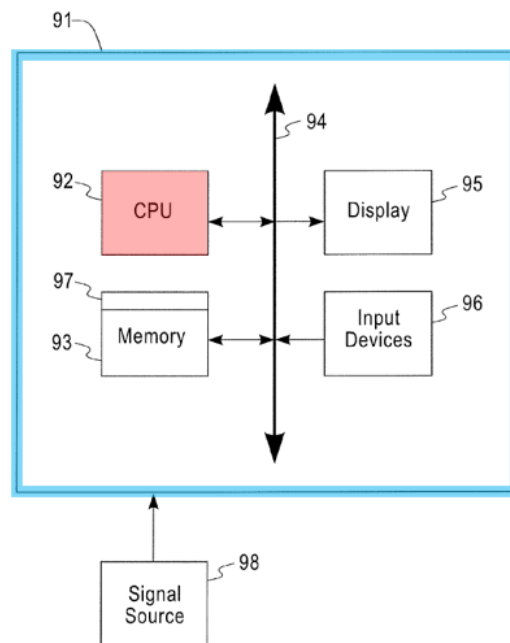


FIG. 9

Avegliano ’157, Fig. 9 (91, 92) (annotated), [0074]-[0075].

278. If “processor” is construed as MPF, the combination of Avegliano ’157 and Rodriguez discloses or suggests the structure discussed in Section X. As I

explained above, Avegliano '157 discloses computer system 91 and CPU 92 for executing a “program of instructions” to “perform the method steps for managing inventory and sales trends in a retail environment using mobile devices.” *Avegliano '157*, [0030], [0040], [0066]-[0076], Fig. 9 (91, 92). Avegliano '157's computer system 91 including CPU 92 along with Input Devices 96 perform the claimed function of “capture images with a camera; detect and read content of a shelf label; associate shelf label content with product facing bounding boxes; cross-correlate product facing bounding boxes and other product facing bounding boxes across stores; build training data, use the training data to build a product classifier, and use the product classifier to perform one or more of an inventory management activity related to the product or classify an item as the product.” The disclosed structures in the prior art are no more than insubstantially different from the disclosed structures in the '272 patent, i.e., processing modules for computing, a server.

279. Additionally, the table below correlates the reasons why the combination of Avegliano '157 and Rodriguez renders obvious claim limitations 18.3-18.10 for the reasons rendering claim 1 obvious.

Claim 18		
Limitation	Substantially Similar Limitation(s)	Section Reference
18.3	1.1	XI.D.2.b

18.4	1.2	XI.D.2.c
18.5	1.3	XI.D.2.d
18.6	1.4	XI.D.2.e
18.7	1.5	XI.D.2.f
18.8	1.6	XI.D.2.g
18.9	1.7	XI.D.2.h
18.10	1.8, 10.8	XI.D.2.i and XI.D.3

5. Claims 3, 15

- a. **Claim 3: “The method of claim 1, further comprising using manual input to identify potential inventory designated in the one or more product facing bounding boxes.”**
- b. **Claim 15: “The method of claim 10, further comprising using manual input to identify potential inventory designated in the one or more bounding boxes.”**

280. It is my opinion that the combination of Avegliano ’157 and Rodriguez renders obvious claims 1 and 10. *See* Sections XI.D.2 and XI.D.3. As I explained in Sections XI.D.2 and XI.D.3, Avegliano ’157’s the method and system can determine product availability data by identifying products on shelves. *See Avegliano ’157*, Abstract, [0016], [0021], [0030], [0033], [0046]-[0048], [0057]-[0059], [0065]. Product identification entails the use of images taken by “[a] mobile application, deployed on customers and merchant devices.” *Id.*, [0048]. An example of mobile application is one designed “for consumers [] to report that a product is missing on the shelf.” *Id.*, [0046]. As such, Avegliano ’157 discloses that missing product information extracted from captured images is used to determine sales velocity and

trends, which in turn can be used to manage store inventory. *Avegliano* '157, Abstract, [0016], [0021], [0030], [0033], [0046]-[0048], [0057]-[0059], [0065]. Alternatively, *Avegliano* '157 discloses that “information regarding the description, quantity, and location of missing items on each shelf of the store” “can be entered **manually** by the employee.” *Avegliano* '157, [0064].

281. Accordingly, in my opinion the combination of *Avegliano* '157 and *Rodriguez* discloses or suggests manually entering information of missing product from store shelves (“using manual input to identify potential inventory”).

6. Claim 4: “The method of claim 1, wherein using the training data to build a product classifier comprises integrating training to: (1) learn product image descriptors and (2) classify items into a neural network.”

282. It is my opinion that the combination of *Avegliano* '157 and *Rodriguez* renders obvious claim 1. *See* Section XI.D.2.

283. *Rodriguez* uses “[v]arious spatial filtering techniques” when deriving “spectral image[-]based **classifiers**.” *Rodriguez*, [0127]-[0131]. Specifically, “to detect particular structures for **classifying an object**,” “filters that compare each pixel with one or more neighboring pixels” can utilize a “**texture descriptor model**” that “parameterizes the relationship between a pixel of interest and its neighbor or group of neighbors.” *Rodriguez*, [0131]; *see also id.*, [0136].

284. Recognizing that “machine learning and neural network technology are re-emerging [] as promising solutions to a great variety of problems,” Rodriguez discloses that “**Neural Networks** and associated learning methods (e.g., RNN, Refractory neural nets and vision) may also be applied to **design an object classifier.**” *Rodriguez*, [0019], [0124].

285. Accordingly, in my opinion the combination of Avegliano ’157 and Rodriguez discloses or suggests (1) classifying an object using a text descriptor model (“learn product image descriptors”), and (2) applying neural networks in designing an object classifier for classifying products (“classify items into a neural network”).

7. Claim 5: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises manually associating the shelf label content with each of the one or more product facing bounding boxes.”

286. It is my opinion that the combination of Avegliano ’157 and Rodriguez renders obvious claim 1. *See* Section XI.D.2.

287. Rodriguez discloses that “[t]he purpose of a classifier is to take an input stimulus and produce, as an output, one of a discrete number of class identifiers.” *Rodriguez*, [0004]. In the classification method and system discussed in §XI.D.2, identification data associated with the product is derived from “image frames captured in a retail store.” *Rodriguez*, [0160]-[0171]. The derivation is achieved using “a machine recognition technology, e.g., digital watermark decoding,”

“barcode decoding,” or image fingerprint matching. *Id.*, [0163]. “All such information can serve as ‘ground truth’ data, to label the product depicted in the sequence of imagery captured by the point-of-sale system, permitting such imagery to serve as training data for a **supervised learning system**.” *Id.*, [0083]-[0085]. Further, “[i]n the case of supervised learning, **the classifier designer controls labeling of training signals** as being in our [sic] out of a desired class.” *Id.*, [0049]. This enables the “classifier designer” to train the system’s response based on “whether [the training signal] has the features of the desired classification or not.” *Id.* A POSITA would have understood that the classifier designer is a person that can control labeling of the product and manually associate the label content with potential product candidates based on whether the images contain the features of the desired classification.

288. Accordingly, it is my opinion that the combination of Avegliano ’157 and Rodriguez discloses or suggests a classifier designer who controls label contents and manually associates label content (“shelf label content”) with potential product candidates (“product facing bounding boxes”).

8. Claims 6, 17

- a. **Claim 6: “The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”**
- b. **Claim 17: “The method of claim 10, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.”**

289. It is my opinion that the combination of Avegliano '157 and Rodriguez renders obvious claims 1 and 10. *See* Sections XI.D.2 and XI.D.3. The combination of Avegliano '157 and Rodriguez render obvious claims 6 and 17 for the same reasons as discussed in claim limitation 1.4, showing association between shelf tag content and product facing bounding boxes without human intervention, above. *See* Section XI.D.2.e.

290. Specifically, Avegliano '157's solution is “a fully **automatic** procedure” that obtains “an identification with high accuracy and speed by taking a picture of a shelf rack.” *Avegliano '157*, [0015]; *see also id.*, [0051], [0064]. The list of product can also “be retrieved **automatically** from the tagged shelves.” *Id.*

291. Accordingly, in my opinion, the combination of Avegliano '157 and Rodriguez renders obvious claims 6 and 17.

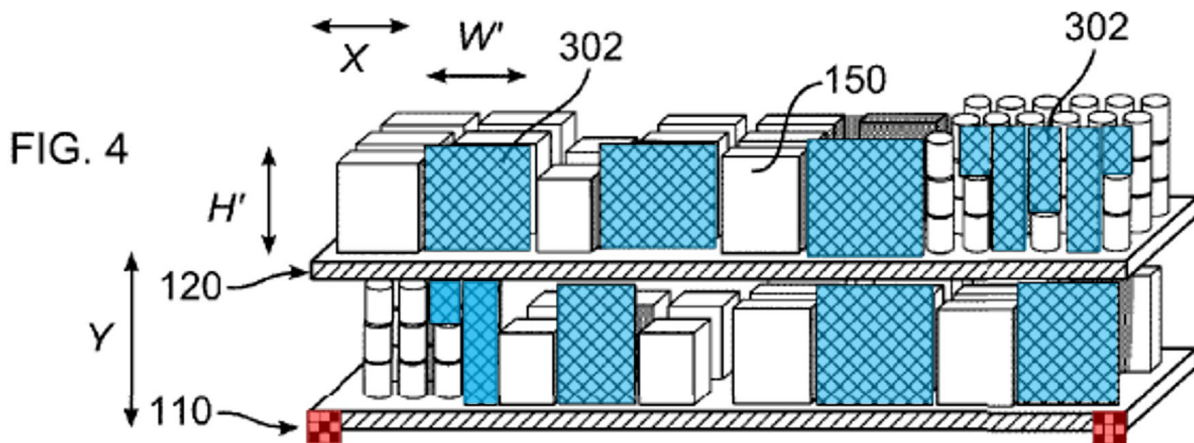
9. Claim 16: “The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.”

292. It is my opinion that the combination of Avegliano '157 and Rodriguez renders obvious claim 10. *See* Section XI.D.3. Avegliano '157 discloses or suggests the additional limitations of claim 16.

293. As I explained in Section XI.D.2.f, the method and system of Avegliano '157 can determine product availability data by identifying products on shelves. *See Avegliano '157*, Abstract, [0016], [0021], [0030], [0033], [0046]-[0048], [0057]-[0059], [0065]. The product identification entails the use of images taken by “[a] mobile application, deployed on customers and merchant devices” “for consumers [] to report that **a product is missing on the shelf.**” *Id.*, [0046], [0048]. Specifically, data of “missing product items on a shelf acquired by the mobile application is in the form of a picture” and “[t]he method includes identifying products on the shelf using tags attached to the shelves.” *Id.*, [0026]. As such, Avegliano '157 discloses that missing product information extracted from captured images is used to determine sales velocity and trends, which in turn can be used to manage store inventory. *Avegliano '157*, Abstract, [0016], [0021], [0024], [0026] [0030], [0033], [0046]-[0048], [0057]-[0059], [0065].

294. Although Avegliano '157 does not expressly disclose that “missing product items” on a shelf is represented by a gap between adjacent inventory

products, it was well known to a POSITA that missing products from retail store shelf leave behind a vacant shelf area where the missing products are supposed to be displayed. For example, in the same field of endeavor, Avegliano '430 discloses “an image obtained of shelf 120, together with marker 110, includes sufficient information to provide a scale, whereby it is possible to determine dimensions of other objects within the image, **including a size of vacant areas upon a shelf 120.**” Avegliano '430, [0018]. Avegliano '430 further discloses that a vacant area is less than fully stocked condition results in vacant areas on the store shelves. Specifically, “system 100 determines which areas of the shelves do not have product extending to the forward region of the shelf, which is indicative of a **less than fully stocked condition**. This condition results in **vacant areas 302**, shown in hatching in FIG. 4.” Figure 4 of Avegliano '430 is reproduced below to show shelves 120, markers 110 (e.g., shelf labels) (red), and vacant areas 302 (blue).



Id., Fig. 4.

295. Subsequently, “[o]nce it is determined which display row has missing product, the size of the vacant area can be determined with reference to objects of known dimension within the image.” *Id.*, [0028].

296. It would also have been obvious to a POSITA to identify the “missing product items” on shelves using the method for identifying products that are present, such as matching at least the product’s height and width from the product template to the height and width of vacant space on the identified shelf and determine the coordinates of an area on the shelf where the product is supposed to be. *See Avegliano ’157*, [0051]-[0053], [0061]-[0063]. For example, *Avegliano ’430* discloses that “an identification of the missing product can be further validated, for example if the vacant area is about the size of a particular product, or a multiple of the size of a particular product.” *Avegliano ’430*, [0028]-[0032], [0040]-[0053], Fig. 4.

297. *Avegliano ’157* identifies missing products in the same manner as the process of identifying “gaps between products” to “identify or locate missing products” that is described in the ’272 patent. For example, *Avegliano ’157* discloses or suggests using the width of a vacant shelf area (e.g., shelf area without product facing present) to determine the type of missing product. *Avegliano ’157*, [0051]-[0053], [0061]-[0063]. In the same way, the ’272 patent identifies gaps between products for identifying “missing/absent inventory.” According to the ’272 patent,

“[g]aps between products are useful for identifying shelf spacings, product separation, or missing/absent inventory.” ’272*pat*, 8:16-17.

298. Accordingly, it is my opinion that the combination of Avegliano ’157 and Rodriguez discloses or suggests determining coordinates of a vacant shelf area between adjacent products (“bounding boxes surround gaps between potential inventory”).

10. Claim 20: “The system of claim 18, wherein instructions configured to capture one or more further images comprise instructions configured to capture the one or more further images subsequent to capturing the one or more images.”

299. It is my opinion that the combination of Avegliano ’157 and Rodriguez renders obvious claim 18. *See* Section XI.D.4. The combination further discloses or suggests the additional limitations of claim 20.

300. As I explained in Sections XI.D.2.f and XI.D.2.i, Avegliano ’157 discloses that the calculation of sales velocity and trends entails the use of “[a] mobile application, deployed on customers and merchant devices to **periodically** monitor product sales velocity.” *Id.*, [0046]-[0048]. To periodically monitor sales velocity (e.g., the change of sale volume over time), the inventory should be examined at different times to determine how many products were sold between the beginning and the end of the monitored time. Therefore, to periodically monitor product sales velocity, multiple images of product shelves should be captured through a period of time, such as images captured by the mobile applications.

301. A POSITA would have recognized that images are captured across time and across stores to create a relative sequence of captured images. For example, to determine “sales velocity” or “sales trend,” the mobile application would capture a first set of images depicting a product on the shelf, and after “a period of time,” it would capture another set of images of the same shelf to determine if that product is missing (e.g., sold out).

302. Accordingly, it is my opinion that the combination of Avegliano ’157 and Rodriguez discloses or suggests this limitation.

11. Claim 21: “The system of claim 18, wherein instructions configured to capture one or more further images of another portion of a shelf comprise instructions configured to capture the one or more further images.”

303. The combination of Avegliano ’157 Rodriguez renders obvious claim 18. *See* Section XI.D.4. The combination further discloses or suggests the additional limitations of claims 21. As I explained in claim 20, Avegliano ’157 receives multiple images of product shelves through a period of time to determine “sales velocity” and “sales trend.” *Id.*, [0046]-[0048].

304. A POSITA would have recognized, therefore, that the method and system of the combination of Avegliano ’157 and Rodriguez captures a plurality of images (“further images”) across a period of time. Accordingly, in my opinion, the combination of Avegliano ’157 and Rodriguez discloses this limitation.

E. Ground 4: Avegliano '157, Rodriguez, and Avegliano '430 Renders Obvious Claims 2, 7-9, 11-14, 16, and 19

1. Motivation to Combine Avegliano '157 and Rodriguez with Avegliano '430

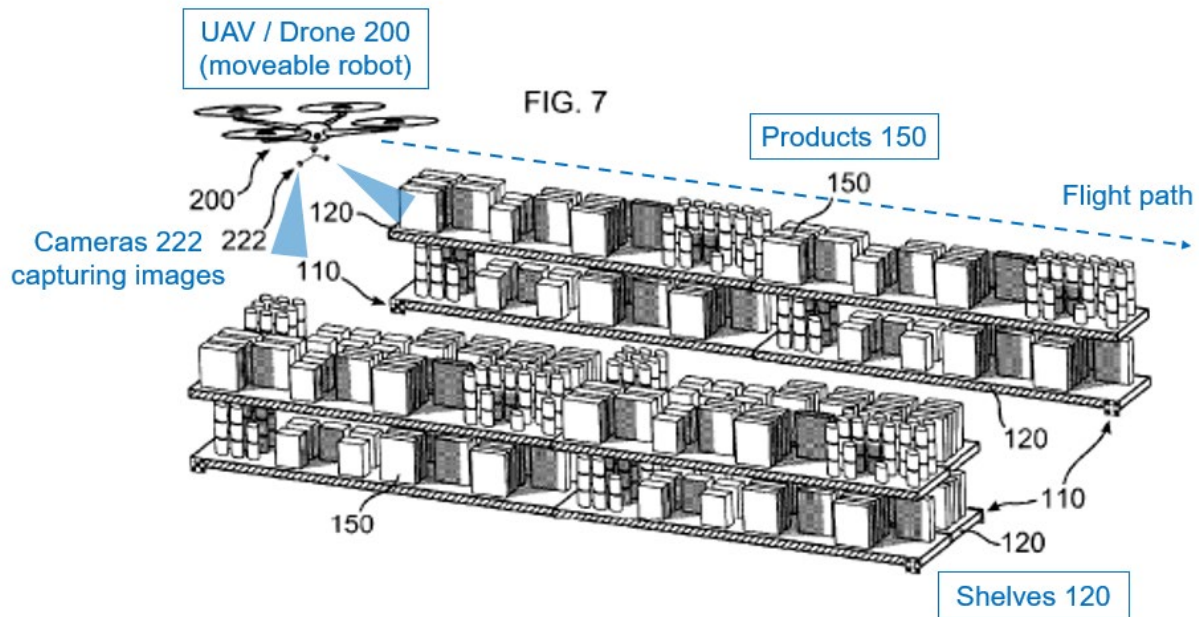
305. A POSITA would have been motivated to further modify the combination of Avegliano '157 and Rodriguez with Avegliano '430.

306. The combination of Avegliano '157 and Rodriguez discloses an object identification solution that begins with capturing images in real-world retail stores, the method continues with reading contents of shelf tags within the images to identify products on the shelf and build a product classifier based on the image and the shelf tag's content. *Avegliano '157*, Abstract, [0016], [0018]-[0019], Figs. 3-4; *Rodriguez*, Abstract, [0009], [0081], [0086], [0166], [0186]. Avegliano '157 describes “systems and methods for managing inventory and analyzing sales trends in a retail environment using mobile devices” to “minimize the time of out-of-the shelf situations.” *Avegliano '157*, [0045], [0047] Fig. 4. Avegliano '157 relies on an employee or a customer walking around the aisles taking photos of the shelves and products using mobile devices. *Avegliano '157*, [0048], [0056] However, as Avegliano '157 acknowledges, “shelf monitoring in supermarkets is necessary” to avoid “items going out of stock,” but “[m]onitoring products in supermarket is an intensive work typically performed by the employees.” *Id.*, [0010].

307. A POSITA would have looked to Avegliano '430 for help with this problem because it discloses flying an unmanned aerial vehicle (UAV) that carries cameras and moves along rows of displayed products to gather image data of the shelves and products. *Avegliano '430*, [0034]-[0039]. The drone is fully autonomous and enjoys the advantage of imaging adjacent rows with a single pass. Specifically, drone 200 contains “[t]wo cameras 220 and/or two depth measuring devices 222” and “with a single pass,” the “adjacent rows can be imaged simultaneously.” *Avegliano '430*, [0039]; *see also id.*, [0035]. In addition, “[d]rone 200 can pass along all rows of a store during a single flight session, or can fly along given rows at intervals, where certain more active rows have more frequent overflights.” *Id.*

308. Supplementing the combination of Avegliano '157 and Rodriguez with drones described in Avegliano '430 can provide the benefit of obtaining “near or real-time information relating to a current stocking level of display shelves” without “human intervention.” *Id.*, [0018], [0035], [0039], [0050]. This modification is no more than applying a known technique to a known device ready for improvement to yield predictable results. For example, this improvement applies the known method of automatically taking images of store aisles to the known device (mobile device). Indeed, doing so would not only relieve employees from the labor-intensive and tedious work of walking around aisles and capturing images of shelves, but also assist in acquiring and processing images outside of business hours (e.g., final

inventory check after end of day) and across multiple store locations. *Avegliano* '430, [0022], [0039].

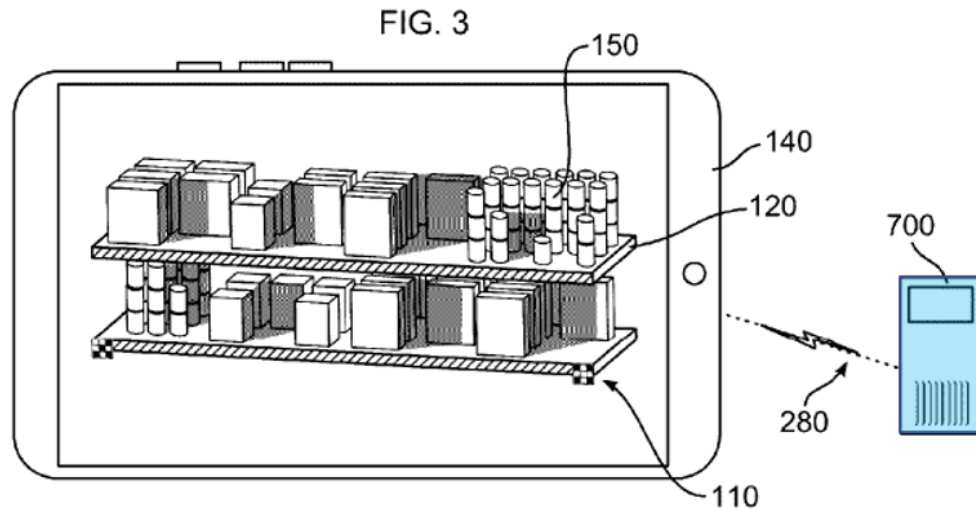


Id., Fig. 7.

309. It would have been obvious to a POSITA to substitute the camera in the combination of *Avegliano* '157 and Rodriguez for the drone cameras of *Avegliano* '430 because *Avegliano* '430 discloses capturing images by the drone and/or the employee. *Id.*, [0022] (“an image of a shelf section has been captured using a digital camera”), [0034] (“FIGS 5-7 illustrate an alternative to having an employee carry out such image gathering and/or depth readings”) (“a drone 200 carrying one or more cameras 220...can fly along rows of displayed product, gathering image/video data.”)

310. Also, like the combination of Avegliano '157 and Rodriguez, Avegliano '430 discloses capturing images of products in retail stores and reading tags, which are referred to as markers, to identify the products. *Avegliano '157*, [0016]; *Rodriguez*, [0009]; *Avegliano '430*, [0003]-[0005].

311. In addition, it would have been obvious to modify the combination of Avegliano '157 and Rodriguez using Avegliano '430's computer system that supports capturing images across stores and sending the images to a remote server for analysis. As I explained in Section XI.C.1, like Avegliano '157 and Rodriguez, Avegliano '430 discloses capturing images of store shelves to identify products and determine “what objects are missing from the vacant areas.” *Avegliano '430*, [0003], [0026]. Avegliano '430 further discloses that “it can be advantageous to obtain and process the image data at a remote server” to “leverage information for **multiple store locations**.” *Avegliano '430*, [0022]. Specifically, “shelf images can be identified as to **a store location**, time of day, shelf section, or any other criteria” and sent to “a remote server 700.” *Id.* I have annotated and reproduced Figure 3 of Avegliano '430 to show remote server 700 in blue.



Id., Fig. 3.

312. Data derived from images from multiple stores can be processed to determine “restocking needs,” “sale velocity,” and “comparison of sales using **different placement locations.**” *Avegliano '430*, [0040]. Therefore, *Avegliano '430* discloses or suggest capturing images from multiple stores and sending the captured images to a remote server 700 for object identification and determination of sales trend across multiple stores. A POSITA would have been motivated to do so to track sales velocity and trends across multiple stores regionally or nationwide. This is one of the traditional reasons for taking inventory.

313. A POSITA would have had a reasonable expectation of success of implementing (1) the drone image capture system and (2) the computer system that supports capturing and analyzing images across stores in the combination of *Avegliano '157* and *Rodriguez* combination because this modification represents

nothing more than the combination of known elements used according to their ordinary and expected operations.

314. In addition, Avegliano '157 and Avegliano '430 share the same first inventors, Applicants, content, and field of use. *Avegliano '157*, cover, Abstract, [0002]-[0034]; *Avegliano '430*, cover, Abstract, [0002]-[0005].

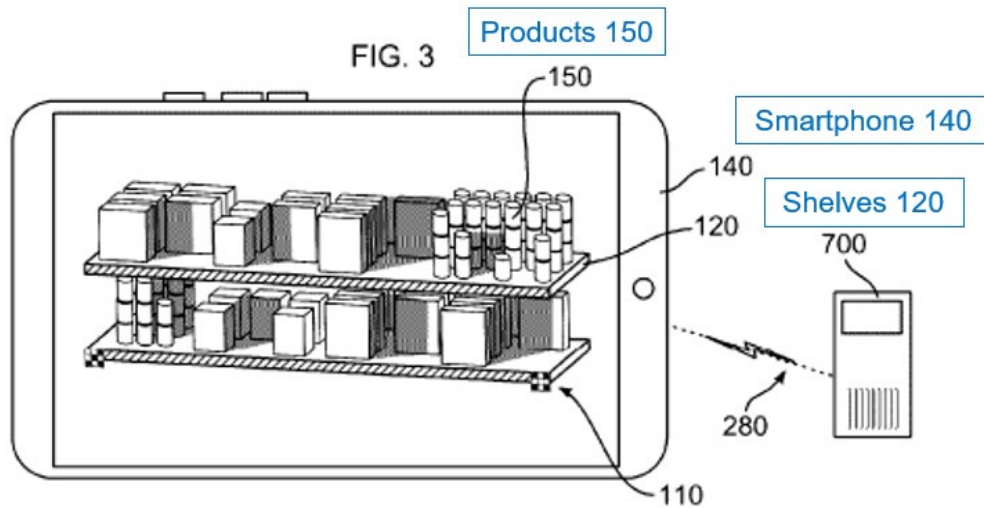
315. The modifications were well within the skills of a POSITA and would have required nothing more than the application of known, conventional techniques (utilizing a drone to capture images of retail products on shelves and capturing and analyzing images across stores).

2. Claims 2, 14

- a. Claim 2: “The method of claim 1, wherein capturing one or more images comprises capturing the one or more images with a camera attached to a moveable robot.”**
- b. Claim 14: “The method of claim 10, wherein capturing one or more images of a portion of a shelf with a camera comprises capturing the one or more images with a camera attached to a moveable robot.”**

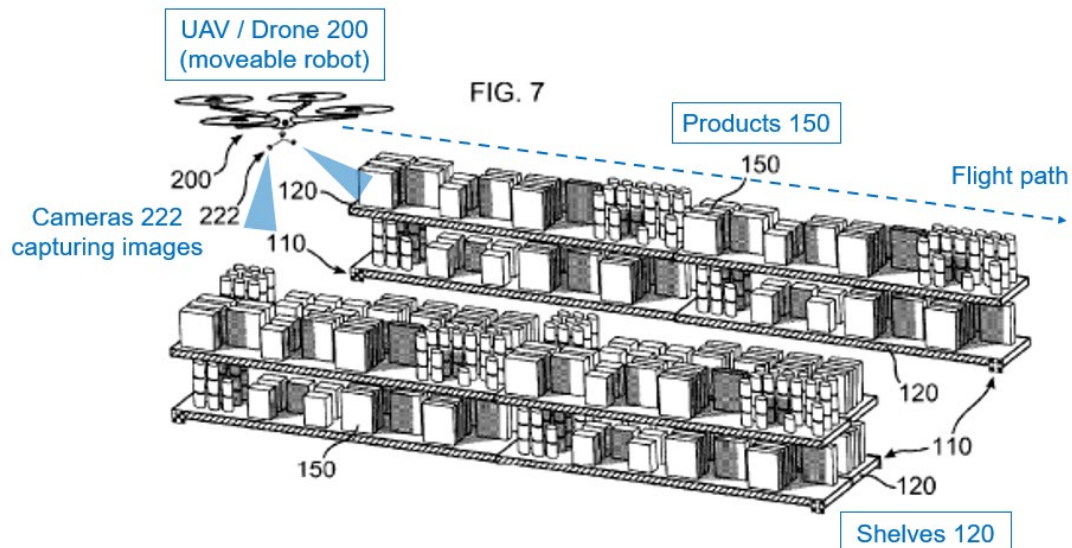
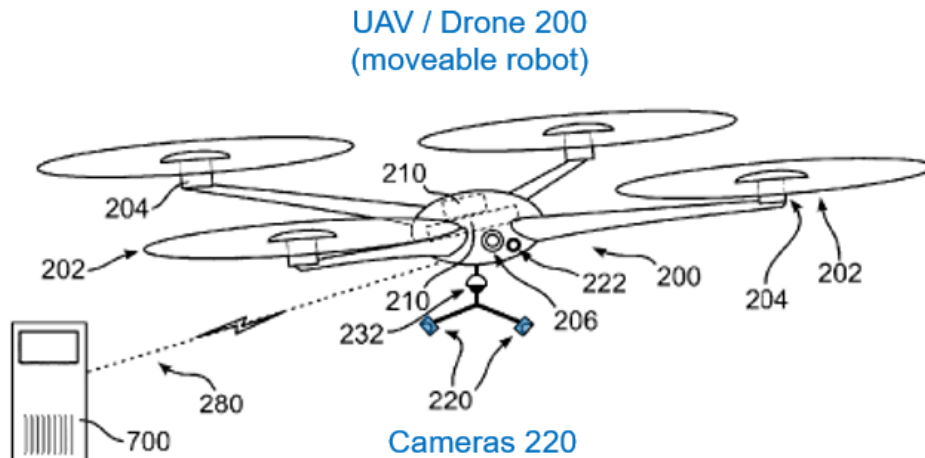
316. It is my opinion that the combination of Avegliano '157 and Rodriguez renders obvious independent claims 1 and 10 as discussed above. *See* Sections XI.D.2, XI.D.3. In my opinion, Avegliano '430 discloses or suggests claims 2, 14. For example, Avegliano '430 discloses capturing an image of a shelf section using a smartphone or other portable computing device camera as shown below.

Avegliano '430, [0022]. I have annotated and reproduced Figure 3 of *Avegliano '430* below to show products 150 being placed on shelves 120 and smartphone 140.



Id., Fig. 3.

317. An alternative embodiment can mitigate the need for an “employee [to] carry out such image gathering” by flying an unmanned aerial vehicle (UAV) or drone 200 carrying cameras 220 and depth measuring devices 222 along rows of displayed product and gathering image data. *Id.*, [0034]-[0039]. Figure 5 below shows the detailed components of drone 200 and Figure 7 depicts drone 200 capturing images while moving along store aisles.



Id., Figs. 5, 7.

318. Avegliano '430's UAV/drone includes the control, navigation, and communication circuitry, cameras, sensors, electronics, and circuitry to operate remotely, semi-autonomously, or fully autonomously. *Id.*, [0035], [0037], [0039].

Specifically, the UAV/drone “can conduct flights either autonomously or under human control.” *Id.*, [0039].

319. Accordingly, it is my opinion that Avegliano ’430 discloses or suggests gathering images with cameras attached to a UAV or drone (“camera attached to a moveable robot”). It would have been obvious to combine Avegliano ’430 with Avegliano ’157 and Rodriguez for the reasons described in Section XI.E.1.

3. Claim 7

320. I explained that the combination of Avegliano ’157 and Rodriguez renders obvious independent claim 1. *See* Section XI.D.2. As I also explained in Sections XI.D.2.f–XI.D.2.i, Avegliano ’157 discloses or suggests spatial aggregation that determines sales velocity, sales trend, and real-time inventory status to benefit “scenarios with several retailers and several producers.” *Avegliano ’157*, [0050].

321. To the extent Patent Owner argues that the combination of Avegliano ’157 and Rodriguez does not teach capturing images across stores and perform object identification in images captured across stores, it is my opinion that Avegliano ’430 discloses or suggests this feature, as explained below.

- a. **Claim 7.0: “The method of claim 1, wherein capturing one or more images of a portion of a shelf comprises capturing one or more images of a portion of a shelf in a first store; and further comprising:”**

322. If limiting, it is my opinion that the combination of Avegliano ’157 and Rodriguez renders obvious limitation 7.0 for the same reasons discussed in claim limitations 1.1 (“capturing one or more images of a portion of a shelf with a camera”) and 1.5 (“cross correlating...product facing bounding boxes...*across one or more stores*”). See Sections XI.D.2.b, XI.D.2.f. Specifically, it is my opinion that the combination of Avegliano ’157 and Rodriguez discloses capturing of images and identifying objects in at least one store.

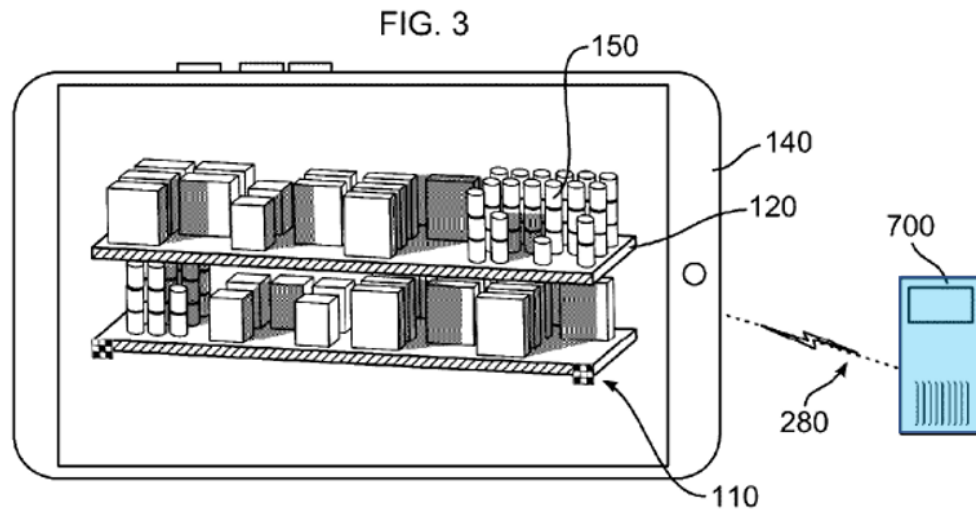
- b. **Claim 7.1.1: “capturing one or more further images of portion of another shelf in a second store;”**

323. It is my opinion that the combination of Avegliano ’157, Rodriguez, and Avegliano ’430 renders obvious limitation 7.1.1.

324. As I explained in Sections XI.C.1, XI.E.1 that like Avegliano ’157 and Rodriguez, Avegliano ’430 discloses capturing images of store shelves to identify products and determine “what objects are missing from the vacant areas.” *Avegliano ’430*, [0003], [0026].

325. Avegliano ’430 further discloses that “it can be advantageous to obtain and process the image data at a remote server” to “leverage information for **multiple store locations**.” *Avegliano ’430*, [0022]. Specifically, “shelf images can be

identified as to **a store location**, time of day, shelf section, or any other criteria” and sent to “a remote server 700.” *Id.* I have annotated Figure 3 of Avegliano ’430 to highlight server 700 (blue).



Id., Fig. 3.

326. Data derived from images from multiple stores can be processed to determine “restocking needs,” “sale velocity,” and “comparison of sales using **different placement locations**.” *Avegliano ’430*, [0040]. Therefore, Avegliano ’430 discloses or suggests capturing images from multiple stores and sending the captured images to a remote server 700 for object identification and determination of sales trend across multiple stores. It would have been obvious to a POSITA that comparison of sales velocity and sales trends between a first store and a second store could include capturing images from the first and second stores and comparing their sale velocity and trends by analyzing the captured images. This is one of the traditional reasons for taking inventory.

327. Accordingly, it is my opinion that the combination of Avegliano '157, Rodriguez, and Avegliano '430 discloses or suggests this limitation.

c. Claim 7.1.2: “reading content of another shelf label on the portion of the other shelf within the one or more further images;”

328. It is my opinion that the combination of Avegliano '157, Rodriguez, and Avegliano '430 renders obvious limitation 7.1.2 for the same reasons as claim limitations 1.2 and 7.0 and additionally because Avegliano '430 discloses executing this step at multiple store locations. *See* Sections XI.D.2.c, XI.E.3.a.

d. Claim 7.1.3: “defining one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and”

329. It is my opinion that the combination of Avegliano '157, Rodriguez, and Avegliano '430 renders obvious limitation 7.1.3 for the same reasons as claim limitations 1.3 and 7.0 and additionally because Avegliano '430 discloses executing this step at multiple store locations. *See* Sections XI.D.2.d XI.E.3.a.

e. Claim 7.1.4: “associating the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;”

330. It is my opinion that the combination of Avegliano '157, Rodriguez, and Avegliano '430 renders obvious limitation 7.1.4 for the same reasons as claim limitations 1.4 and 7.0 and additionally because Avegliano '430 discloses executing this step at multiple store locations. *See* Sections XI.D.2.e, XI.E.3.a.

- f. **Claim 7.2: “wherein cross-correlating comprises cross-correlating the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the second store by matching the shelf label content with the other shelf label content; and”**

331. It is my opinion that the combination of Avegliano ’157, Rodriguez, and Avegliano ’430 renders obvious limitation 7.2 for the same reasons as claim limitations 1.5 and 7.0 and additionally because Avegliano ’430 discloses executing this step at multiple store locations. *See* Sections XI.D.2.f, XI.E.3.a.

- g. **Claim 7.3: “wherein building training data comprises building the training data based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store.”**

332. It is my opinion that the combination of Avegliano ’157, Rodriguez, and Avegliano ’430 renders obvious limitation 7.3 for the same reasons as claim limitations 1.6 and 7.0 and additionally because Avegliano ’430 discloses executing this step at multiple store locations. *See* Sections XI.D.2.g, XI.E.3.a.

4. Claim 11

333. It is my opinion that the combination of Avegliano ’157, Rodriguez, and Avegliano ’430 renders obvious claim 11 for substantially the same reasons as claim 7, as I explained in Section XI.E.3 because the claims are substantively identical. *Compare* claim limitations 11.0-11.3, *with* claim limitations 7.0-7.3.

334. The table below correlates the reasons why the combination of Avegliano '157, Rodriguez, and Avegliano '430 renders obvious claim limitations 11.0-11.3 for the reasons rendering claim 7 obvious.

Claim 11		
Limitation	Substantially Similar Limitation(s)	Section Reference
11.0	7.0	XI.E.3.a
11.1.1	7.1.1	XI.E.3.b
11.1.2	7.1.2	XI.E.3.c
11.1.3	7.1.3	XI.E.3.d
11.1.4	7.1.4	XI.E.3.e
11.2	7.2	XI.E.3.f
11.3	7.3	XI.E.3.g

5. Claim 19

335. The combination of Avegliano '157, Rodriguez, and Avegliano '430 renders obvious claim 19 for substantially the same reasons as claims 7 and 11, as I explained in Sections XI.E.3, XI.E.4 because the claims are substantively identical. *Compare* claim limitations 19.0-19.3, *with* claim limitations 7.0-7.3.

336. The table below correlates the reasons why the combination of Avegliano '157, Rodriguez, and Avegliano '430 render obvious claim limitations 19.0-19.3 for the reasons rendering claim 7 obvious.

Claim 19		
Limitation	Substantially Similar Limitation(s)	Section Reference
19.0	7.0	XI.E.3.a
19.1.1	7.1.1	XI.E.3.b
19.1.2	7.1.2	XI.E.3.c
19.1.3	7.1.3	XI.E.3.d
19.1.4	7.1.4	XI.E.3.e
19.2	7.2	XI.E.3.f
19.3	7.3	XI.E.3.g

6. Claims 8, 12

- a. **Claim 8: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images subsequent to capturing the one or more images.”**
- b. **Claim 12: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further subsequent to capturing the one or more images.”**

337. The combination of Avegliano '157, Rodriguez, and Avegliano '430 renders obvious claims 8 and 12 for the same reasons as claims 1, 10, 18, and 20 and additionally because Avegliano '430 discloses executing this step at multiple store locations. *See* Sections XI.D.2.a, XI.D.3, XI.D.4, XI.D.10. In addition, a POSITA would have understood that there are finite options of the temporal relationship between the capture times of two images. Specifically, an image can be captured before, after, or at the same time another image is captured. It would have been obvious for a POSITA to try any or all of those three options with a reasonable

expectation of success. For example, the sale velocity and sales trend can be determined by calculating the difference of inventory using images taken at different times, i.e., one image is captured before or after another image is captured. In addition, a POSITA would have been motivated to sequentially take the shelf images and send the images to remote server 700 to avoid overwhelming the computation power of server 700 when the server 700 receives and analyzes multiple images taken and transmitted at the same time.

7. Claims 9, 13

- a. Claim 9: “The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”**
- b. Claim 13: “The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.”**

338. It is my opinion that Avegliano ’157, Rodriguez, and Avegliano ’430 renders obvious claims 7 and 11. *See* Sections XI.E.3, XI.E.4. The combination also discloses or suggests the additional limitations of claims 9 and 13 for the reasons and image capture timing (“prior to”) discussed in claims 8 and 12. *See* Section XI.E.6.

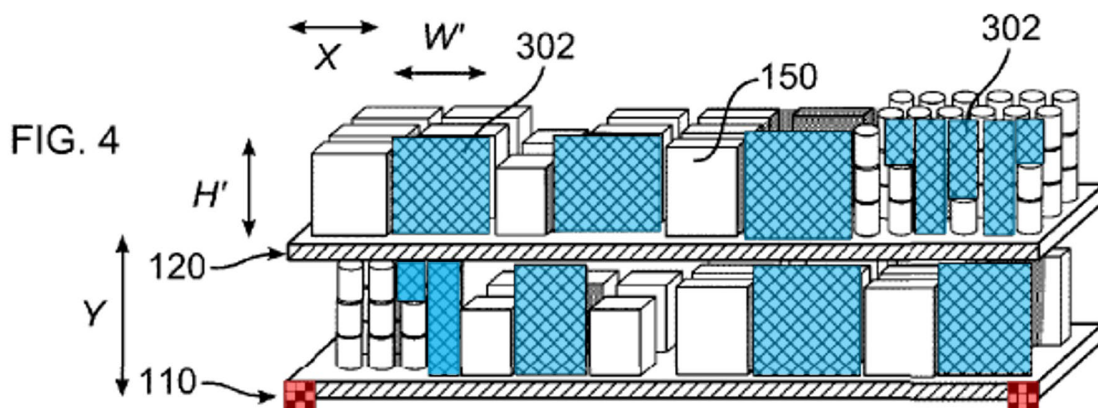
8. Claim 16: “The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.”

339. It is my opinion that Avegliano ’157, Rodriguez, and Avegliano ’430 renders obvious claim 16. *See* Section XI.D.3.

340. As I explained in Sections XI.D.2.f and XI.D.9, Avegliano '157 determines product availability data by identifying products on shelves. *Avegliano '157*, Abstract, [0016], [0021], [0030], [0033], [0046]-[0048], [0057]-[0059], [0065]. The product identification uses images taken by “[a] mobile application, deployed on customers and merchant devices” “for consumers [] to report that a product is missing on the shelf.” *Id.*, [0046], [0048]. Specifically, data of “missing product items on a shelf acquired by the mobile application is in the form of a picture” and “[t]he method includes identifying products on the shelf using tags attached to the shelves.” *Id.*, [0026]. As such, Avegliano '157 uses missing product information extracted from captured images to determine sales velocity and trends, which in turn can be used to manage store inventory. *Avegliano '157*, Abstract, [0016], [0021], [0024], [0026] [0030], [0033], [0046]-[0048], [0057]-[0059], [0065].

341. Although Avegliano '157 does not expressly disclose “missing product items” on a shelf represented by a gap between adjacent inventory products, Avegliano '430 discloses this feature. For example, in the same field of endeavor, Avegliano '430 discloses “an image obtained of shelf 120, together with marker 110” are used to determine “a size of vacant areas upon a shelf 120.” *Avegliano '430*, [0018]. Vacant areas are “areas of the shelves [that] do not have product extending to the forward region of the shelf, which is indicative of a less than fully stocked

condition.” *Id.*, [0024]. I have annotated Fig. 4 of Avegliano ’430 to show shelves 120, markers 110 in red, and vacant areas 302 in blue.



Id., Fig. 4.

342. Avegliano ’430 further discloses that “[o]nce it is determined which display row has missing product, the size of the vacant area can be determined with reference to objects of known dimension within the image.” *Id.*, [0028]. Avegliano ’430 further uses product’s “known height and width” to identify the missing product. *Avegliano ’430*, [0029]; *id.*, Figs. 3-4, 7 [0028]-[0032], [0040]-[0053]. It would also have been obvious to a POSITA to identify Avegliano ’157’s “missing product items” on shelves using Avegliano ’430’s method for identifying missing products because doing so amounts to no more than applying a known technique (Avegliano ’430’s methods for identifying missing products) to a piece of

prior art (Avegliano '157's object identification method) ready for improvement (accurately determine identification of missing products).

343. In addition, it is my opinion that the processes of identifying missing products described in the combination of Avegliano '157, Rodriguez, and Avegliano '430 operate in the same manner as the processes of identifying “gaps between products” to “identify or locate missing products” described in the '272 patent. For example, both Avegliano '157 and Avegliano '430 use the width of a vacant shelf area (e.g., shelf area without product facing present) to determine the type of missing product. *Avegliano '157*, [0051]-[0053], [0061]-[0063]; *Avegliano '430*, [0028]-[0032], [0040]-[0053]. In the same way, the '272 patent identifies gaps between products for identifying “missing/absent inventory.” '272*Pat*, 8:16-17.

XII. SECONDARY CONSIDERATIONS

344. At this stage of these proceedings, it is my understanding that Petitioner has no burden to identify and rebut secondary considerations. Rather, it is my understanding that Patent Owner must first present a prima facie case for such consideration, which Petitioners should then have the chance to rebut. That said, I have considered evidence of secondary considerations that I am aware of at this time, and I am currently unaware of any evidence of secondary considerations that would support a finding of non-obviousness.

XIII. CONCLUSION

I, Andrew Wolfe, 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.

Dated: June 20, 2024

Respectfully submitted,

A handwritten signature in black ink, appearing to read 'Andrew Wolfe', is written above a horizontal line.

Andrew Wolfe, Ph.D.

LISTING OF CHALLENGED CLAIMS

Reference	Claim Limitation
Claim 1	
1.0	A method comprising:
1.1	capturing one or more images of a portion of a shelf with a camera;
1.2	detecting and reading content of a shelf label on the portion of the shelf within the one or more images;
1.3	defining one or more product facing bounding boxes surrounding one or more corresponding facings of a product within the one or more images;
1.4	associating the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;
1.5	cross-correlating the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;
1.6	building training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes;
1.7	using the training data to build a product classifier; and
1.8	using the product classifier to classify an item as the product.
Claim 2	
2	The method of claim 1, wherein capturing one or more images comprises capturing the one or more images with a camera attached to a moveable robot.
Claim 3	
3	The method of claim 1, further comprising using manual input to identify potential inventory designated in the one or more product facing bounding boxes.
Claim 4	

Reference	Claim Limitation
4	The method of claim 1, wherein using the training data to build a product classifier comprises integrating training to: (1) learn product image descriptors and (2) classify items into a neural network.
Claim 5	
5	The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises manually associating the shelf label content with each of the one or more product facing bounding boxes.
Claim 6	
6	The method of claim 1, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.
Claim 7	
7.0	The method of claim 1, wherein capturing one or more images of a portion of a shelf comprises capturing one or more images of a portion of a shelf in a first store; and further comprising:
7.1.1	capturing one or more further images of portion of another shelf in a second store;
7.1.2	reading content of another shelf label on the portion of the other shelf within the one or more further images;
7.1.3	defining one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and
7.1.4	associating the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;
7.2	wherein cross-correlating comprises cross-correlating the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the

Reference	Claim Limitation
	second store by matching the shelf label content with the other shelf label content; and
7.3	wherein building training data comprises building the training data based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store.
Claim 8	
8	The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images subsequent to capturing the one or more images.
Claim 9	
9	The method of claim 7, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.
Claim 10	
10.0	A method comprising:
10.1	capturing one or more images of a portion of a shelf with a camera;
10.2	detecting and reading content of a shelf label on the portion of the shelf within the one or more images;
10.3	defining one or more product facing bounding boxes surrounding one or more facings of a product within the one or more images;
10.4	associating the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;
10.5	cross-correlating the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;
10.6	building training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes;

Reference	Claim Limitation
10.7	using the training data to build a product classifier; and
10.8	using the product classifier to perform an inventory management activity related to the product.
Claim 11	
11.0	The method of claim 10, wherein capturing one or more images of a portion of a shelf comprises capturing one or more images of a portion of a shelf in a first store; and further comprising:
11.1.1	capturing one or more further images of a portion of another shelf in a second store;
11.1.2	reading content of another shelf label on the portion of the other shelf within the one or more further images;
11.1.3	defining one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and
11.1.4	associating the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;
11.2	wherein cross-correlating comprises cross-correlating the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the second store by matching the shelf label content with the other shelf label content; and
11.3	wherein building a training data comprises building the training data set based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store.
Claim 12	
12	The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further subsequent to capturing the one or more images.

Reference	Claim Limitation
Claim 13	
13	The method of claim 11, wherein capturing one or more further images comprises capturing the one or more further images prior to capturing the one or more images.
Claim 14	
14	The method of claim 10, wherein capturing one or more images of a portion of a shelf with a camera comprises capturing the one or more images with a camera attached to a moveable robot.
Claim 15	
15	The method of claim 10, further comprising using manual input to identify potential inventory designated in the one or more bounding boxes.
Claim 16	
16	The method of claim 10, wherein the one or more bounding boxes surround gaps between potential inventory.
Claim 17	
17	The method of claim 10, wherein associating the shelf label content with each of the one or more product facing bounding boxes comprises automatically associating the shelf label content with each of the one or more product facing bounding boxes.
Claim 18	
18.0	A system comprising:
18.1	a processor;
18.2	system memory coupled to the processor and storing instructions configured to cause the processor to:
18.3	capture one or more images of a portion of a shelf with a camera;
18.4	detect and reading content of a shelf label on the portion of the shelf within the one or more images;
18.5	define one or more product facing bounding boxes that surround an surrounding one or more corresponding facings of a product within the one or more images;

Reference	Claim Limitation
18.6	associate the shelf label content with each of the one or more product facing bounding boxes that includes and defines each of the one or more corresponding product facings;
18.7	cross-correlate the one or more product facing bounding boxes and one or more other product facing bounding boxes across one or more stores by matching the shelf label content with other shelf label content associated with the one or more other product facing bounding boxes;
18.8	build training data based on the association of the shelf label content with the cross-correlated product facing bounding boxes;
18.9	use the training data to build a product classifier; and
18.10	use the product classifier to perform one or more of: (1) an inventory management activity related to the product or (2) classifying an item as the product.
Claim 19	
19.0	The system of claim 18, wherein instructions configured to capture one or more images of a portion of a shelf comprise instructions configured to capture one or more images of a portion of a shelf in a first store; and further comprising instructions configured to:
19.1.1	capture one or more further images of a portion of another shelf in a second store;
19.1.2	read content of another shelf label on a portion of the other shelf within the one or more further images;
19.1.3	define one or more other product facing bounding boxes surrounding one or more other corresponding facings of another product within the one or more further images; and
19.1.4	associate the other shelf label content with each of the one or more other product facing bounding boxes that includes and defines each of the one or more other corresponding product facings;
19.2	wherein instructions configured to cross-correlate the one or more product facing bounding boxes and the one or more other product facing bounding boxes across the first store and the

Reference	Claim Limitation
	second store by matching the shelf label content with the other shelf label content; and
19.3	wherein instructions configured to build a training data comprising building the training data sent based on the association of the shelf label content and the other shelf label content with the cross-correlated product facing bounding boxes correlated across the first store and the second store.
Claim 20	
20	The system of claim 18, wherein instructions configured to capture one or more further images comprise instructions configured to capture the one or more further images subsequent to capturing the one or more images.
Claim 21	
21	The system of claim 18, wherein instructions configured to capture one or more further images of another portion of a shelf comprise instructions configured to capture the one or more further images.