



(51) International Patent Classification:

G06F 3/0482 (2013.01) G06K 17/00 (2006.01)
G01S 3/786 (2006.01) G06T 7/162 (2017.01)
G06F 16/26 (2019.01) G06V 10/50 (2022.01)
G06F 16/38 (2019.01)

(21) International Application Number:

PCT/US2023/021728

(22) International Filing Date:

10 May 2023 (10.05.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/340,101 10 May 2022 (10.05.2022) US

(72) Inventors; and

(71) Applicants: **MULLER, Michael** [US/US]; 10960 Wilshire Blvd., #1900, Los Angeles, California 90024 (US).
HASSAN, Sharmil [UA/US]; 10960 Wilshire Blvd., #1900, Los Angeles, California 90024 (US).

(74) Agent: **FREYER, Andrew J.** et al.; Brownstein Hyatt Farber Schreck, LLP, 675 15th Street, Suite 2900, Denver, Colorado 80202 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available):

AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: SYSTEMS AND METHODS FOR INGESTING AND PROCESSING ENRICHABLE CONTENT

(57) Abstract: Described herein is a system and a method for ingesting an enrichable content and associated one or more actions. One or more objects are identified from the received enrichable content, and object clusters are generated. Based on each object of the object clusters, an identifier of the object clusters is generated and saved as a database record in a database. The one or more actions corresponding to the enrichable content are stored in one or more databases as associated with an index of a database record representing the identifier of the object clusters.

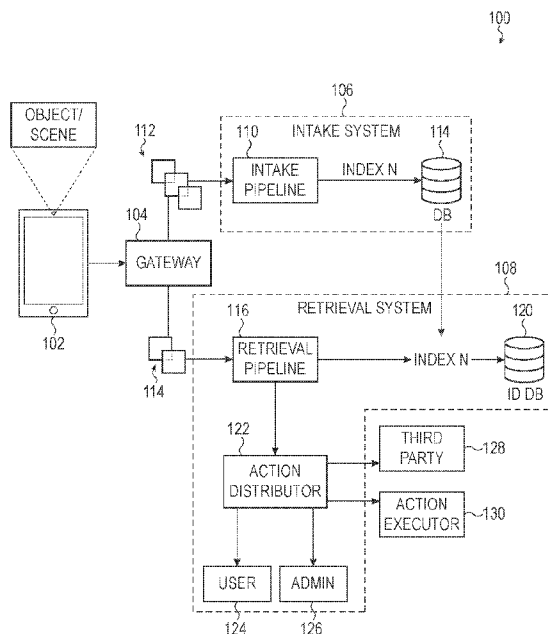


FIG. 1



SYSTEMS AND METHODS FOR INGESTING AND PROCESSING ENRICHABLE CONTENT

CROSS-REFERENCE TO RELATED APPLICATION

5 [0001] This Patent Cooperation Treaty Patent Application claims priority to U.S. Provisional Patent Application No. 63/340,101, filed May 10, 2022, and titled “Systems and Methods for Ingesting and Processing Enrichable Content”, the contents of which are incorporated herein by reference in its entirety.

TECHNICAL FIELD

10 [0002] Embodiments described herein relate to a systems and methods for ingesting and processing enrichable content and delivering enriched content to a client device.

BACKGROUND

[0003] An object or media (e.g., photograph, video, and so on) may be affixed or rendered with a label, barcode, beacon, or tag that encodes information about that object or media, such as
15 an identification number, a uniform resource location (URL), or the like linking to supplemental information about the object or media, and so on. Examples include Quick Response (QR) codes, bar codes, Bluetooth Low Energy (BLE) beacons, Near-Field Communication (NFC) tags, radio frequency identification tags (RFID), and so on. Each of these conventional techniques of encoding information about an object or media have significant limitations.

20 [0004] For example, a barcode such as a QR code is a graphical static matrix encoding information that can be read, and acted upon, by an electronic device with a camera or scanning laser in close proximity of that code. In many cases, such as with televised media, a code may only appear for a brief period of time (e.g., during a commercial), significantly limiting the ability of a viewer to access the content linked to by that code. In other cases, a printed code may
25 be damaged either maliciously by environmental exposure over time.

[0005] Similarly, an electronic tag such as a BLE beacon may be a low-power electronic device that regularly broadcasts encoded (often static) information that can be wirelessly received, and acted upon, by a suitably-capable electronic device within a few meters of the beacon. An NFC tag is typically an unpowered electronic circuit encoding static information that
30 can be read, and acted upon, by an NFC-capable electronic device within a few centimeters of

the tag. Similar limitations are present for RFID tags. Such electronic devices have significantly limited range and may have a limited service life, as battery capacity drains over time.

5 [0006] More generally, each of these and other conventional information encoding techniques exhibit several drawbacks. For example, in many cases, a conventional code, label, tag, or beacon occupies physical space, obscuring a portion of the object or media to which it is affixed. In other cases, conventional codes, labels, tags, and beacons encode static information that cannot be changed or updated remotely without additional intermediate proxies or redirects, many of which are blocked by certain web security policies.

10 [0007] Moreover, conventional codes, labels, tags, and beacons are subject to failure, tampering, and/or damage, often rendering them completely unusable for intended purposes. Further, in almost all contexts, conventional codes, labels, beacons, and tags require specific threshold proximity with an electronic device attempting to read them. More specifically, neither a QR code, nor an NFC tag, nor a BLE beacon can be scanned from a large distance, even if undamaged and operating normally.

15

SUMMARY

[0008] Embodiments described herein take the form of a system for ingesting enrichable content and associating the enrichable content with an action, the system including at least an application server including at least a memory allocation storing executable instructions, and a processor allocation operably coupled with the memory allocation and configured to load from the memory allocation the executable instructions thereby instantiating an instance of a backend application configured to communicably couple to a frontend application instance instantiated over a client device in network communication with the application server, the backend application instance configured to receive a structured data object with a static image of the enrichable content and an attribute identifying an action to associate to the enrichable content, provide the static image as input to a high-accuracy object classifier, receive as output from the high-accuracy object classifier, a set of objects and corresponding bounding boxes identified within the static image, generate an object graph based on the set of corresponding bounding boxes, generate a hash corresponding to object graph, and storing the hash and the action in a database.

[0009] Related and additional embodiments include a configuration in which the client device includes a personal electronic device with a camera.

[0010] Related and additional embodiments include a configuration in which the structured data object includes a set of static images, and the static image may be a member of the set of static images.

[0011] Related and additional embodiments include a configuration in which the set of static images are associated with frames of a captured video.

[0012] Related and additional embodiments include a configuration in which the backend application may be configured to, for each respective static image of the set of static images provide the respective static image as input to the high-accuracy object classifier, receive as output from the high-accuracy object classifier, a respective set of objects and corresponding bounding boxes identified within the respective static image, generate at respective graph based on the respective set of corresponding bounding boxes, and generate a respective hash corresponding to object graph.

[0013] Related and additional embodiments include a configuration in which the backend application may be configured to generate a single hash from each respective hash.

[0014] Related and additional embodiments include a configuration in which the single hash may be associated to the action in the database.

[0015] Related and additional embodiments include a configuration in which the structured data object includes location information of the client device.

[0016] Related and additional embodiments include a configuration in which the hash may be based, at least in part, on the location information.

5 [0017] Related and additional embodiments include a configuration in which the structured data object includes orientation information of the client device.

[0018] Related and additional embodiments include a configuration in which the hash may be based, at least in part, on the orientation information.

10 [0019] Related and additional embodiments include a configuration in which the orientation information may be based at least in part on accelerometer data of the client device, compass data of the client device, or gyroscope information of the client device.

[0020] Related and additional embodiments include a configuration in which the structured data object includes network connectivity information of the client device.

15 [0021] Related and additional embodiments include a configuration in which the hash may be based, at least in part, on the network connectivity information.

[0022] Related and additional embodiments include a configuration in which the network connectivity information corresponds to a Wi-Fi connection, a cellular connection, or a Bluetooth connection.

20 [0023] Related and additional embodiments include a configuration in which the action includes an instruction to cause the client device to load a URL.

[0024] Related and additional embodiments include a configuration in which the action includes an instruction to cause the client device to render a virtual reality scene on a display of the client device.

25 [0025] Related and additional embodiments include a configuration in which the action includes an instruction to cause the client device to render an augmented reality scene on a display of the client device.

30 [0026] Embodiments described herein take the form of a system for identifying enrichable content and causing to be executed at least one action associated with the enrichable content, the system including at least an application server including at least a memory allocation storing executable instructions, and a processor allocation operably coupled with the memory allocation and configured to load from the memory allocation the executable instructions thereby instantiating an instance of a backend application configured to communicably couple to a frontend application instance instantiated over a client device in network communication with the application server, the backend application instance configured to receive a structured data

object with a static image, provide the static image as input to a high-speed object classifier, receive as output from the high-speed object classifier, a set of objects and corresponding bounding boxes identified within the static image, generate an object graph based on the set of corresponding bounding boxes, generate a hash corresponding to object graph, determine
5 whether the hash may be equivalent to a hash previously stored in a database, in response to determining that the hash may be equivalent to a previously-stored hash, retrieve the previously-stored hash from the database, provide the retrieved hash as an input query to an action database, receive from the action database in response to the input query, an action associated with the retrieved hash, provide the retrieved hash to an action distributor to cause the action to be
10 executed by at least one of a third-party system, the backend application, or the client device.

[0027] Related and additional embodiments include a configuration in which the static image may be an image of a scene captured by the client device.

[0028] Related and additional embodiments include a configuration in which the scene includes an active television displaying a broadcast.

15 **[0029]** Related and additional embodiments include a configuration in which the structured data object includes at least one of location information of the client device, or orientation information of the client device.

[0030] Embodiments described herein take the form of a method of operating a server application to ingest media content and associate the media content with one or more actions to
20 be performed by at least one of a client device, a third-party service, or a first-party service, the method including at least receiving a structured data object with a static image of a scene and an attribute identifying an action to associate to the enrichable content, provide the static image as input to a high-accuracy object classifier, receive as output from the high-accuracy object classifier, a set of objects and corresponding bounding boxes identified within the static image,
25 generate an object graph based on the set of corresponding bounding boxes, generate a hash corresponding to object graph, and storing the hash and the action in a database.

[0031] Related and additional embodiments include a configuration in which the static image includes a QR code.

30 **[0032]** Related and additional embodiments include with extracting data encoded by of the QR coded.

[0033] Related and additional embodiments include with associating the action with the extracted data of the QR code.

[0034] Related and additional embodiments include a configuration in which the structured data includes information obtained from an NFC tag, RFID, or Bluetooth tag disposed within the scene.

[0035] Related and additional embodiments include with associating the action with the
5 information.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] Reference will now be made to representative embodiments illustrated in the accompanying figures. It should be understood that the following descriptions are not intended to limit this disclosure to one included embodiment. To the contrary, the disclosure provided herein is intended to cover alternatives, modifications, and equivalents as may be included within the spirit and scope of the described embodiments, and as defined by the appended claims.

[0037] FIG. 1 depicts an example computing network in, or over which embodiments presented in this disclosure may be implemented.

[0038] FIG. 2 depicts an example intake system, as described herein.

[0039] FIG. 3 depicts an example retrieval system, in accordance with some embodiments.

[0040] FIG. 4A depicts an example computing environment corresponding to an intake system, as described herein.

[0041] FIG. 4B depicts an example computing environment corresponding to a retrieval system, as described herein.

[0042] FIGs. 5A-5G depict various example use cases or practical applications of embodiments described herein.

[0043] FIG. 6 depicts an example user interface of a client application executing on a client device, in accordance with some embodiments.

[0044] FIGs. 7A-7B depict an example of object identification and generation of object clusters, in accordance with some embodiments.

[0045] FIG. 8 depicts a flowchart corresponding to example operations of a method being performed by an intake system, in accordance with some embodiments.

[0046] FIG. 9A depicts a flowchart corresponding to example operations of a method being performed by a retrieval system, in accordance with some embodiments.

[0047] FIG. 9B depicts another flowchart corresponding to example operations of a method being performed by the retrieval system, in accordance with some embodiments.

[0048] FIG. 10 depicts a flowchart corresponding to example operations of a method being performed by a retrieval system, in particular, for an occluded QR code and/or an out-of-range NFC tag, RFID tag, and/or a BLE beacon, in accordance with some embodiments.

[0049] The use of the same or similar reference numerals in different figures indicates similar, related, or identical items.

[0050] Additionally, it should be understood that the proportions and dimensions (either relative or absolute) of the various features and elements (and collections and groupings thereof) and the boundaries, separations, and positional relationships presented therebetween, are

provided in the accompanying figures merely to facilitate an understanding of the various embodiments described herein and, accordingly, may not necessarily be presented or illustrated to scale, and are not intended to indicate any preference or requirement for an illustrated embodiment to the exclusion of embodiments described with reference thereto.

DETAILED DESCRIPTION

[0051] Embodiments described herein relate to systems and methods for enriching a user's engagement with physical objects, scenes, and/or media. In particular, embodiments described herein leverage a portable electronic device, such as a cellular phone, that includes a camera to
5 image a scene within a field of view of the camera.

[0052] Frames of the scene, captured by the camera at any suitable frame rate and/or any suitable resolution (in some cases, downscaled and/or segmented to save bandwidth) can be processed by a cloud service as describe herein so as to recognize content within the scene, relative positions between recognized content, and cause the electronic device that imaged the
10 scene to perform one or more actions (thereby "enriching" the scene or object with supplemental content, actions, tasks, media, and so on) based on, or associated with, the recognized content.

[0053] Example enrichments/actions that can be associated with a recognized scene as described herein can include, without limitation: loading a website; launching an application; rendering an augmented reality object or media over a live view of the scene imaged by the
15 portable electronic device; facilitating a purchase; passing sensor data to a remote service or server; creating or passing ownership in a non-fungible token or crypto-asset; creating a digital twin object in a digital environment (e.g., a "metaverse"); issuing a coupon for a particular service or product; generating a promotion code for a particular product or service; rendering a video; playing or causing to be shown selected media; rendering a particular object or media
20 within a virtual reality (VR) environment (e.g., for a user wearing a VR headset) and so on. These examples are not exhaustive; in some cases, a particular scene or object in a scene can be enriched with multiple actions performed by the portable electronic device, a third-party server, or any other suitable electronic device.

[0054] In particular, in many embodiments described herein, one or more frames captured by
25 the camera (along with, in some cases, sensor output from one or more sensors of the electronic device, such as a position sensor, accelerometer, temperature sensor, humidity sensor, proximity sensor, global position sensor, and so on) can be uploaded to a cloud service configured to leverage a trained machine learning model to identify and/or label or classify objects, items, text, or persons in the imaged scene. In many cases, the trained machine learning model can be
30 additionally configured to segment one or more frames into a grid, processing subsets of a complete frame individually (and/or in parallel).

[0055] In addition, the trained machine learning model can be configured to determine relative positions between individual identified objects (in particular, relative positions of bounding boxes, as one example; for example, geometric centers of bounding boxes can be

located in a coordinate space and an object graph can be constructed in which nodes of the graph are associated with bounding boxes and edges of the graph connect nearest neighbor bounding boxes) in the imaged scene.

[0056] These sets of identified/classified objects and associated relative positions

5 therebetween can be collapsed into and/or otherwise represented by a fingerprint, hash, or vector that, in turn, can be used to compare against a database of vector representations of various scenes or objects (e.g., for use with a sparse representation classifier, as one example).

[0057] Once a match is determined, an action database can be queried to access and return one or more actions associated with the recognized scene or object. Thereafter, an action

10 distributor service can cause each action obtained from the database to be performed by an appropriate device or software instance.

[0058] More generally and broadly, embodiments described herein relate to systems and methods for associating arbitrary actions (one or more) with arbitrary real-world scenes, objects, and/or media. As used herein, a scene, person, object, or media that can be uniquely identifiable

15 by methods described herein can be referred to as “enrichable content.” Similarly, a scene, person, object, or media that has been uniquely identified and associated with one or more actions or tasks can be referred to as “enriched content,” “enriched media,” and/or “enriched objects.”

[0059] In view of the foregoing, embodiments described herein generally and broadly relate to systems and methods for ingesting (e.g., identifying and associating actions to) enrichable content and, additionally, to systems and methods for automatically causing to be executed one or more actions in response to subsequent identification of a previously-ingested enrichable content.

[0060] As one example, an enrichable content as described herein may be a social media post that includes a single face, several commercial products in the foreground, several background objects (e.g., picture frames, houseplants, furniture, and so on). An author of the social media post can upload the post to an intake system, as described herein, so that the content of that social media post (e.g., recognized objects, faces, text, and so on) and the unique arrangement of that content (e.g., relative positions of object within clusters, and relative positions of clusters of objects, and so on), can be associated to, and/or otherwise collapsed into, a unique identifier (sometimes referred to herein as vectorization or hashing of one or more object clusters). For example, a list of recognized objects may be captured as a set of bounding boxes surrounding respective recognized/classified/labeled content, a set of confidences, and a graph data structure

30

associating relative positions between bounding boxes. In addition, in many embodiments, clusters of objects (within a threshold distance of one another) can be grouped together.

[0061] For example, the list of objects and clusters may be presented, stored, and/or transmitted as a JavaScript TM Object Notation (JSON) object, such as:

```
5      {
      clusters : [{
      index : 0,
      bounds : [0,0, 5000, 5000],
      nodes : [
10     {
      index : 0,
      bounds: [0, 0, 123, 456],
      object_id : F0E4C2F76C589,
      label : "human_face",
15     label_id : 421C76D77563A,
      label_confidence : 0.81,
      },
      {
      index : 1,
20     bounds: [452, 1233, 667, 988],
      object_id : C834750C0E9B,
      label : "product_COMAPNY Skincare SKU: 12345",
      label_id : 5B062BD022FD,
      label_confidence : 0.95,
25     },
      {
      index : 2,
      bounds: [345, 1233, 667, 988],
      object_id : 73C2E6788FCCA,
30     label : "product_COMAPNY Sunscreen SKU: 45678",
      label_id : FA1914846B010BD1,
      label_confidence : 0.76,
      },
      {
```

```

    index : 3,
    bounds: [15, 678, 27, 2569],
    object_id : A0558FFB854B0,
    label : "furniture_couch",
5    label_id : 64F395BD34C,
    label_confidence : 0.96,
    },
    {
    index : 4,
10   bounds: [250, 260, 35, 37],
    object_id : 31EA478C0934,
    label : "picture_frame",
    label_id : 6EC258F246855,
    label_confidence : 0.82,
15   },
    {
    index : 5,
    bounds: [77, 370, 35, 78],
    object_id : F3C64B80B0A69,
20   label : "plant_dracaena",
    label_id : 0B463070C69FB34,
    label_confidence : 0.91,
    }
  ]
25  }}]
  }

```

[0062] This data object may be parsed by a system as described herein to generate a graph for each cluster that in turn can be hashed into a single output value representing the unique
 30 arrangement of objects in each cluster. For example, a system as described herein may execute an algorithm such as follows:

```

clusters = data.clusters
cluster_graphs = []

```

```
#create graph of nodes within each cluster
for each cluster in clusters:
    cluster_graph = {}
5    for each node in cluster:
        for each other_node in cluster:
            if node == other_node:
                continue
            else:
10            distance = normalize(dist_calc(node.bounds, other_node.bounds))
                angle = normalize(angle_calc(node.bounds, other_node.bounds))

                new_edge = new Edge(node.label_id, other_node.label_id, distance, angle)
                if not new_edge in cluster_graph:
15            cluster_graph.append(new_edge)
                new_cluster_graph = new Graph(cluster.index, cluster_graph, cluster.bounds)
                cluster_graphs.append(new_cluster_graph)

#create graph of clusters
20 clusters_graph = []
    for each cluster in cluster_graphs:
        for each other_cluster in cluster:
            if cluster == other_cluster:
                continue
25            else:
                distance = normalize(dist_calc(cluster.bounds, other_cluster.bounds))
                angle = normalize(angle_calc(cluster.bounds, other_cluster.bounds))

                new_edge = new Edge(cluster.index, other_cluster.index, distance, angle)
30            if not new_edge in clusters_graph:
                clusters_graph.append(new_edge)

#generate output hash(s)
cluster_graphs_outhash = hash_all(cluster_graphs)
```

```

clusters_graph_outhash = hash(clusters_graph)
combined_hash = hash(cluster_graphs_outhash + salt + clusters_graph_outhash)

#returns
5   return (cluster_graphs_outhash, clusters_graph_outhash, combined_hash)

```

[0063] As may be appreciated by a person of skill in the art, the foregoing example algorithm creates one or more graph data structures having edges defined by distances and angles separating individual bounding boxes of objects identified within input data (e.g., the foregoing JSON object). Once an edge is defined (including an angle and distance, which may be based on pixel counts and/or normalized to a standard scale) between each pairing of individual nodes (objects) that edge can be added to a list of edges between labeled objects defining a graph of objects recognized within a particular cluster.

[0064] In addition, positional relationships between clusters (having their own bounding boxes defined to surround objects defining the cluster) can be used to define a higher-order graph data structure defining an arrangement of clusters of objects within a particular scene.

[0065] The graph data structures associated with each cluster, and the overarching graph defining positional relationships between clusters themselves can each be hashed to a unique value. The hash may be an ordered hash function or an unordered hash function. For embodiments in which the hash is ordered, similar arrangements of objects generate similar hash values.

[0066] In some cases, the hashes associated with individual clusters and the hash associated with a graph of clusters within a scene can be concatenated (optionally with one or more salt values) and re-hashed to define a single hash function or identifier representing the arrangement of objects within a particular scene.

[0067] In this manner, the set of identified objects and one or more hashes representing arrangements of those objects can be used to reliably identify a particular object set within a particular scene, such as the social media post of the preceding example.

[0068] Continuing the preceding example, the author of the social media post – after having uploaded the social media post to be processed in the manner described above and/or elsewhere herein – may also associate one or more actions, which may be preselected from a list of actions and/or may be customized (e.g., arbitrary code provided by the author or a third party, which may execute as a lambda function or other serverless function, as one example), to the unique identifier.

[0069] Thereafter, persons viewing the social media post at a later time can cause their respective electronic devices to attempt to identify the social media post by leveraging a system such as described herein.

[0070] For example, the social media post may be uploaded to a content identification system, an action retrieval system, or more simply a “retrieval” system which can perform similar operations to the intake system; in particular, the retrieval system can be configured to identify one or more objects, faces, or text within the social media post, and identify a relative arrangement of those objects and/or clusters thereof.

[0071] With this information, the retrieval system can generate an identifier, fingerprint, or hash which can be compared against (for threshold similarity and/or identity) to a database of unique identifiers of enrichable content previously consumed by the intake system. If a match is determined by the retrieval system, the unique identifier can be used to query an action database, to retrieve one or more actions, some or all of which may have been selected by the author of the social media post. Thereafter, the retrieval system can cause the actions to be performed, executed, or otherwise scheduled. In some cases, an action may be triggered at the respective electronic device of the person viewing the social media post. In other cases, an action may be triggered at a third-party system, such as a content interaction tracking service. These examples are not exhaustive.

[0072] A person of skill in the art may appreciate that in the foregoing example, the social media post is enriched with additional features and functionality that may not be provided by the social media platform hosting the post. For example, the author of the post may link – via an enrichment as described herein – to a merchandise purchase store, may provide special content (e.g., as an augmented reality (AR) overlay rendered over the social media post), may automatically generate a coupon or promotion code, and so on.

[0073] In other cases, the social media post may be enriched with one or more opaque actions that occur in the background, such as for engagement tracking and/or copyright enforcement. For example, if the social media post is copied to another platform, the content-based hash (including positional relationships of recognized objects, color histograms, facial recognition output, and so on) will remain the same and/or will be within a threshold distance (e.g., cosine distance) of the original content hash. Thus, whenever the copied social media post is propagated to other platforms--even if modified--engagement statistics can be accurately tracked. In addition, copyright enforcement actions (e.g., DMCA takedown notices) can be appropriately initiated, in some cases automatically, if duplication of the post was not authorized by the content owner.

[0074] It may be further appreciated that the content enrichment features described above are not limited to presentations of the social media post on the social media platform. More specifically, because systems described herein leverage machine learning, generative pretrained transformers, AI, and/or computer vision to identify objects, faces, text, and/or other content within arbitrary frames captured by a camera (and/or uploaded as an image file), the enrichment features follow the social media post wherever that post is reproduced.

[0075] For example, if the social media post is printed, a user can direct a camera of their cellular phone to image the printed social media post and the associated enriching actions can be performed (as the retrieval system will still recognize the same set of identified objects and the same relative arrangement thereof). In another example, if a third party reposts the social media post, the original author maintains control over the enrichment features associated therewith.

[0076] In many examples, a retrieval system as described herein can be configured to operate within given tolerances. For example, exact matches object and arrangements thereof may not be required; suitably close matches (which may vary from embodiment to embodiment) can be considered as matches for purposes of providing enriched content, as described herein. In other cases, content hashes can be generated as ordered hashes between which distances may be calculated to infer similarity between different hashed content. In these examples, any appropriate distance measurement technique can be used. In these cases, a threshold can be defined so as to binarize a determination of whether two content hashes represent the same content. In particular, if a distance between two content hashes satisfies the threshold (e.g., is below a threshold distance), then the two underlying contents are defined as the same. Alternatively, if a distance does not satisfy the threshold, the underlying contents are defined as different contents.

[0077] In many cases, distance measurements may be recorded and logged over time to determine and/or verify whether threshold decisions are appropriate; if many near-miss distances are received, threshold values may be increased.

[0078] This foregoing tolerance-based matching approach can provide further advantages to, as one example, the author of the social media post of the preceding example. For example, if another person crops a watermark out of the original post, the original content may still be recognized.

[0079] In yet further examples, a retrieval system as described herein can be configured to segment an input image into a grid (or other subarea of arbitrary shape or size; examples include concentric circles, randomly-distributed rectangular areas, and so on), and each grid element can be independently processed by the retrieval system. As a result of these embodiments, enrichable

content that is within a larger scene can be independently identified and associated actions can be caused to be performed.

[0080] For example, if the social media post of the preceding example is broadcast in a news segment, viewers of the news segment can provide a photo or video of the segment as input to the retrieval system (e.g., by directing a camera of a personal cellular phone to image the news segment), which in turn can optionally segment the image as described above. At least one segment of the segmented input image set contains at least a portion of the enriched content (e.g., the social media post), and can be identified by the retrieval system and the actions associated therewith can be performed.

[0081] In view of the foregoing example, a person of skill in the art can appreciate that systems described herein offer significant improvements over conventional QR codes, NFC tags, RFID tags, BLE beacons, and so on. In particular, as no specific code is required, a QR code (and/or other code or overlay) is not required to obscure or occlude the social media post.

[0082] The foregoing is merely one example. In other cases, other enrichable content can be identified and associated to one or more actions, as described herein. For example, in some cases, a person may upload a photo of their face as an item of enrichable content. In this example, the person may associate a personal website, a contact card, and/or any other suitable information relevant to and/or selected by the person. In this construction, whenever a photograph of the same person's face is uploaded to the retrieval system, the associated actions can be caused to be performed.

[0083] In yet another example, a category of specific objects may be identified by a system as described herein, such as particular apparel or accessories from a particular manufacturer (e.g., shoes, handbags, clothing, jewelry and so on). In these examples, an associated action may be to purchase a similar or identical item, to create a digital twin of the item in a virtual interaction environment (e.g., metaverse or other simulated environment), to update a loyalty points database with a particular manufacturer, and so on. Many examples are possible.

[0084] In yet other examples, embodiments described herein can be leveraged to enrich media content, such as video advertisements, billboard advertisements, television programs, movies, and so on. For example, a commercial may be received by the intake system on a frame-by-frame basis and/or a subset of frames. In these examples, each frame provided as input to the intake system can result in a particular unique identifier, such as described above.

[0085] In yet other examples, virtual objects can be associated to real world geographic locations. In these cases, scanned enrichable content may or may not be associated with the action (i.e., virtual object rendering, in AR as one example) taken in response to scanning that

content. For example, a system as described herein can be used as a virtual geocaching system or a capture the flag game in which searchers scan real-world content to reveal potential virtual-world objects. In other cases, virtual objects associated with geographic locations may be directly associated with the real-world objects or scenes, such as placing a virtual for-sale sign over real property upon taking a photo of the real property location.

[0086] Further embodiments do not rely on a single frame to generate a content hash as described above. In particular, a sequence of frames – and in particular the set of identified objects and relative positions thereof – can be collapsed into a single identifier, in a similar manner as with individual static frames as described above. In other words, content hashes of individual frames can be graphed together with edges corresponding to the duration between the frames that are captured, thereby introducing time-variance of the scene as yet another hashable property.

[0087] In these examples, a retrieval system may be configured to receive as input one or more frames from a portable electronic device imaging a broadcast of a commercial. The frames can be processed by the intake system, and the sequence of processed frames can be collapsed into a vector, hash, or fingerprint and compared against previously-hashed content now stored in the identifier/vector database. In this manner, video media content -- captured by a camera or imaging device -- can be recognized and enriched by a system as described herein, even if the capturing device does not precisely frame the target video media content, does not align precisely in time with the start of the target content, and so on.

[0088] For example, a commercial may be identified by methods described herein and associated with an enrichment that causes a device that scans/images the commercial (e.g., a personal cellular phone) to display a graphical user interface including an option to buy an advertised product, to initiate a trial of an advertised service, to initiate a videocall or telephone call to a company or person, to initiate a crypto-asset transaction, render a digital twin version (in AR or VR) of an advertised product, or any other suitable enrichment action.

[0089] In other cases, an enrichment action can be leveraged as a mechanism for enforcement of copyright. For example, an enrichment action may cause a client device that images a particular enrichable object or scene to upload information (e.g., as a structured data object having one or more attributes, encoding one or more images or video or other media, sensor data such as location information and so on) to a server under the control of (or otherwise accessible to) a content owner. In this manner, the content owner can be made aware of the existence of copies, whether authorized or not, of content owned by the content owner.

[0090] In yet other examples, an enrichment action can be based on and/or selected in view of a particular context in which an enrichable content is imaged as described herein. For example, if a user leverages a personal cell phone to scan a particular storefront (that has been previously imaged and provided as input to an intake system), different actions may be performed depending upon a global position of the user at the time of the scan. More particularly, if a user is standing in front of the store (e.g., a GPS location corresponds to the storefront's GPS location), a menu overlay may be rendered on the user's device.

[0091] Alternatively, if a user scans a picture of the storefront that is available online, an enrichment action may be to direct the user's browser to a reservation page, so that the user may make a reservation at the restaurant.

[0092] These foregoing examples are not exhaustive; generally and broadly it may be appreciated that a content enrichment system as described herein includes an intake portion and a retrieval portion, both of which can be leveraged to identify particular objects, faces, items, and so on and arrangements thereof in a particular scene (static) or in a sequence of scenes (e.g., video images).

[0093] In some cases, either or both an intake system or a retrieval system as described herein can leverage machine learning, artificial intelligence, sensor systems, data aggregators, computer vision, color histogram analysis modules, sound detection and classification systems, or any other suitable software instances or hardware apparatuses. In many constructions, an intake system leverages a higher-performance object classification technique (which may be slower and/or more computationally intensive) than a retrieval system which may leverage a high-speed object classification technique. In other words, in some constructions, intake operations may be more computationally expensive than retrieval operations which may be configured to execute as quickly as possible.

[0094] In some constructions, at least a portion of preprocessing and/or object classification operation of a retrieval and/or an intake system can be performed on a client device. For example, in some cases, object classification operations of a retrieval system can be performed in part by a user's device. For example, the user's device may be configured to de-skew, rotate, color correct, scale, or otherwise modify one or more frames of an imaged scene before transmitting those frames to a remote server configured to execute other operations of a retrieval system as described herein. In other cases, the remote server may include functionality to automatically identify and crop and de-skew content received from a user.

[0095] In still further embodiments, such as noted above, a retrieval system and/or an intake system as described herein can be configured to segment an input image, processing individual

portions of an image as though each was a separate image. In this manner, sub-portions of an image can be processed in parallel, or, in other cases, sub-portions may be processed in a particular pattern or sequence. For example, in some cases, a central tile/segment of a particular image may be processed first, whereas corner or edge segments of the same image may be processed last.

[0096] In still further embodiments, retrieval operations may be staged in multiple stages. For example, a first stage may be configured to perform crude object detection so as to locate a candidate segment of the image to scan first. In other cases, a first stage may be a stage configured to identify a particular trigger symbol or fiducial, such as an icon or watermark.

Based on a known location of the icon or watermark, segmentation of the overall image can be performed reliably.

[0097] For example, in some cases, a crude object detection algorithm may operate to detect a television or other rectangular object within a scene. Thereafter, a second stage can be configured to attempt to identify a watermark in a particular location within a bounding box surrounding the identified television. In some cases, the watermark or icon can be located in a bottom corner of the screen. Upon identifying the icon, content of the television can be accurately cropped, de-skewed and scale, and transmitted to a remote retrieval system for further identification and action triggering.

[0098] In yet further embodiments, an intake system can be configured to scan for barcodes, QR codes, or other optical codes (or plaintext associated with other information, such as a web address or a phone number; any particular text recognized with a suitable data detector or regular expression may be processed) within a particular scene.

[0099] The intake system can thereafter associate the function(s) of the scanned codes with the identified scene. For example, a scene may include a QR code that links to a website. The intake system can record this web address and present the web address as an optional action to users who subsequently image the scene.

[0100] A person of skill in the art should appreciate that this particular example extends functionality of QR codes, barcodes, plaintext, and actionable plaintext (e.g., web addresses, email addresses, and so on). More particularly, because an intake system as described herein reads these codes (and/or OCRs the plaintext, potentially recognizing certain types of text with a data detector) on intake, a user who subsequently scans the same scene does not need to be close enough to actually scan the QR code for the action associated with that QR code to be offered to the user.

[0101] For example, a storefront may post a QR code that links to the store's menu. The storefront itself may be provided as input to an intake system, which can read the QR code and record the URL pointing to the store's menu.

5 **[0102]** At a later time, a potential customer in a vehicle passing the store may capture an image of the storefront. In this example, the potential customer is not only moving, but is likely too far from the QR code itself for even a high-resolution camera to properly resolve or decode the QR code.

[0103] However, by leveraging methods described herein, the storefront itself can be identified (by its unique collection of objects, distributions thereof, global location and/or other
10 identifying sensor input), and the URL encoded by the QR code – that was not possible to be scanned by the potential customer in the passing vehicle – is presented to the potential customer.

[0104] More generally and broadly, a system as described herein can effectively enable functionality of distant QR codes otherwise impossible to scan. In addition, a system as described herein can effectively enable functionality of an occluded QR code that is otherwise
15 not visible. In yet other examples, a system as described herein can effectively enable functionality of an irreparably damaged QR code that is so damaged that in-built error correction and redundancy fails.

[0105] In addition, a system as described herein can present a URL to a user that is written in plain text but is too far away to be read by the user.

20 **[0106]** In yet further embodiments, an intake system as described herein can record NFC tags, RFID tags, data from BLE beacons, in order to enable functionality thereof in much the same manner as described above with respect to QR codes, barcodes, plaintext.

[0107] For example, a system as described herein can effectively enable functionality of distant NFC tags, RFID tags, and BLE beacons otherwise impossible to scan. In some cases, a
25 system as described herein can enable functionality of NFC tags and BLE beacons for personal electronic devices not capable to scan NFC tags or BLE beacons. In addition, as with QR codes, a system as described herein can effectively enable functionality of an occluded NFC tag or BLE beacon that is otherwise not scannable due to the occlusion. In yet other examples, a system as described herein can effectively enable functionality of an irreparably damaged NFC tag or
30 damage or unpowered BLE beacon that is so damaged that in-built error correction and redundancy fails.

[0108] These forgoing examples are not exhaustive. For example, in some cases, techniques described herein can be leveraged in real time by an augmented or virtual reality headset to scan

an environment nearby a wearer of the headset and to present graphics via the AR/VR display that are selected and/or based on recognized objects.

[0109] In other cases, identification operations as described herein can be assisted with other sensor inputs, such as GPS input. For example, GPS information retrieved from a user's device can be used to filter a possible set of enriched objects to only those objects known to be within the geographic region in which the scan takes place. For example, many fast food franchise storefronts may include similar distributions of physical objects or classifiable objects, but each occupies a different physical location; filtering by physical location can improve recognition accuracy of a particular object over other similar objects.

[0110] In some embodiments, an intake system can be configured by an administrator or content uploader to ignore particular objects or classes of objects. For example, a storefront owner may instruct an intake system to ignore human persons. In other cases, a social media content uploader may instruct the intake system to ignore certain background objects.

[0111] Similarly, in some cases, a retrieval system can be configured to ignore particular object classes when determining content of a particular imaged scene. For example, in some cases, the retrieval system may be configured to movable or moving objects such as persons, animals, vehicles, and the like. In other cases, different instances or different threads of the retrieval system may be configured to ignore different sets of objects. In these examples, a single content item or object can be associated with -- and may be matched by comparing to -- many different hashes, each of which may be based on different segments, portions, object classes or types, or other combinations thereof.

[0112] In view of the foregoing, it may be appreciated that generally and broadly a system as described herein relates to systems and methods for uniquely identifying enrichable content and delivering enrichments/actions to a user on a client device.

[0113] In one example, enrichable content may include one or more images, and/or one or more video frames uploaded by a user to an application server of an intake system. The user may associate one or more actions with the uploaded enrichable content, and the content and the actions may be stored in a database that is communicatively coupled to the application server. Enrichable content associated with the one or more corresponding user actions may generate content that is enriched by one or more corresponding actions.

[0114] Generally and broadly, an application server of an intake system may ingest the uploaded enrichable content and process the uploaded enrichable content using an artificial intelligence algorithm, a machine learning algorithm, facial attribute classifier, generative pretrained transformer, and/or a computer vision algorithm, which may be referred herein

collectively as “AI/ML/CV” or “machine learning algorithm” or “trained classifier” or “predictive model” or “transformers.”

[0115] In particular, the enrichable content uploaded by a user may be processed to identify one or more objects in the uploaded enrichable content, a spatial relationship and/or a temporal relationship (e.g., how objects and/or clusters of objects move between sequences of frames) between the one or more objects identified using the machine learning algorithm.

[0116] In some embodiments, at least one instance of a server application may maintain a library or database based on the enrichable content and associate one or more actions with the enrichable content. In one example, the enrichable content may include an image or other media, and the library may include multiple different images generated based on the image uploaded by the user. For example, multiple images may be generated by varying properties of the uploaded image, such as brightness, contrast, temperature, tint, hue, gamma, color, blur, color tint, rotation, scale, an aspect ratio, and so on.

[0117] In some embodiments, a machine learning algorithm may be used to create a list of one or more objects within in an image (or, more generally, a scene) and generate an object descriptor or label corresponding to each object. In some cases, the machine learning algorithm may be a supervised machine learning algorithm. In other cases, other types of machine learning algorithms may be used in place of and/or with a supervised algorithm. For example, a generative pretrained transformer or other artificial neural networks may be configured to determine object labels and/or object sets based, in part, on language models encoding object label co-occurrences in a single scene. More simply, a transformer may be leveraged as a label filter to prevent nonsensical identification of different objects in the same scene. For example, the system may be configured to determine that a low-confidence identification of a shark from a naïve CV classifier is incorrect if the same scene includes a high-confidence identification of a farmhouse.

[0118] As described above, the object descriptor may include a list of objects identified in the image. In one example, the list of objects identified in the image may be accompanied by and/or may include a bounding box surrounding each identified object or content item. A bounding box may also describe a spatial location of an object in an image relative to a coordinate system particular to an anchor point of the image, such as a corner thereof or a geometric center thereof. Accordingly, as noted above, spatial relationships between two more objects identified in an image may also be determined.

[0119] In some embodiments, each bounding box may be further divided into a number of sections, for example, 16 sections (4x4 sections) or 64 sections (8x8 sections) or an arbitrary

non-square number of sections. A number of sections for dividing each bounding box may be configurable. In some embodiments, for example, each bounding box may be further divided into a number of sections, each of a fixed size. Within each section of an identified object, an object attribute classifier may be executed to further characterize the identified object.

5 [0120] In some embodiments, a vector, hash, or other fingerprint, may be generated based on the number of sections representing each bounding box identifying each object.

[0121] Further, a dictionary of vectors may be leveraged by a machine learning algorithm (e.g., sparse representation classification algorithm) as described herein to identify a candidate scene as a scene already processed by the intake system. Each set of vectors in the dictionary
10 may correspond to object lists and relative positions within an uploaded image (and/or other images derived from that image, such as images of different scale or color content), as described herein. In some embodiments, each vector of the dictionary may be normalized to a particular uniform length.

[0122] Each vector corresponding to each scene or object identified as a result of operation of
15 the intake system can be associated with a particular index number, each of which may be stored in an index database (e.g., an index database, communicatively coupled with the application server of the intake system), which can in turn be associated with actions to be performed whenever the scene corresponding to a particular index is scanned.

[0123] Similarly, a retrieval system may include an application server executing an instance
20 of a server application configured to receive an enrichable content, such as an image or a video, from a client device over a user interface, as described herein. As described herein, the user interface may present, to a user, options regarding a particular usage intended by the user. Accordingly, when the user indicates the user is uploading the image or video to retrieve one or more user actions associated with the image or video being uploaded by the user, the instance of
25 the application server may process the uploaded image or video to identify one or more user actions that may be associated with the image or video uploaded by the user.

[0124] As with the intake system, the instance of the application server may leverage a machine learning algorithm to identify one or more objects captured in an image uploaded by a user, or one or more objects captured in a video (e.g., a series of images). As described herein,
30 one or more bounding boxes surrounding the one or more objects identified in the image or the series of images may be used to generate a vector, as described herein, with reference to the intake system. The vector may be then compared with multiple vectors stored in the vector table to identify the closest matching vector with a predetermined or preconfigured threshold matching error, such as described above.

[0125] A database record index associated with the closest matching vector in the vector table may then be searched in the vector set table to identify a vector set and corresponding database record index corresponding to the matched vector(s). The database record index of the vector set may then be used to query an action database to identify one or more user actions that are
5 associated with the image or video uploaded by the user, some of which may be performed by the user's device, some of which may be performed by a third-party system, some of which may be performed by the retrieval system itself. Many configurations are possible.

[0126] In some cases, a particular image uploaded by a user to a retrieval system may have been associated with an action that may allow the user to purchase or lease specific content for
10 use in metaverse. For example, when a user takes and uploads an image of some apparel previously scanned by the intake system, the retrieval system may retrieve an action to generate a temporary digital twin/copy of the apparel item in a digital environment which may be associated to a user account, may be offered for purchase to the user, or may be offered for lease to the user.

[0127] In yet another example, a user-uploaded image may include an image of purchased
15 item. For example, the user may have purchased a new toaster, which may be still in the original packaging. A user may capture an image including the purchased item in its original packaging. When the user uploads the taken image, based on identification of the purchased item by the retrieval system, the user may be offered an option whether the user would like to complete
20 registration for the purchased item.

[0128] In some embodiments, if the user exercises the option to complete registration, many of the details for the online registration may be automatically filed based on metadata associated with the uploaded image and/or information corresponding to a client device used for uploading the image (e.g., such as a user account) as a structured data object with multiple attributes, each
25 of which may correspond to an image captured by the client device and/or sensor information of sensors of the client device such as GPS information, accelerometer information, gyroscope information, compass information, network connection information (e.g., Wi-Fi, Bluetooth, UWB, cellular connections and so on).

[0129] For example, location data included in the metadata of the uploaded image and/or a
30 GPS location of the client device may be used to complete residence/business address information. Similarly, a username associated with an account authorized on the client device may also be used to complete the online registration for the product.

[0130] In some embodiments, based on identifying a particular product in the uploaded image, a search may be performed if the particular product is registered at an address that may be

identified as described above or associated with the user that is identified as described above. If the particular product is found as not have been registered at the address or associated with the user, an option for online registration may be presented to the user. As described above, many of the details for the online registration may be automatically filed based on metadata associated with the uploaded image and/or information corresponding to a client device used for uploading the image.

[0131] In other cases, a user may leverage a system as described herein to quickly and efficiently complete an inventory of items purchased and owned by the user and placed within the user's home, for example, for homeowners or renters' insurance underwriting or claims purposes. In other examples, a user may leverage a system as described herein to create digital twins of the user's real-world possessions in a virtual environment.

[0132] In some embodiments, a user may upload the image on the user's one or more social network accounts. Based on the number of other visits to the particular uploaded image, the user may be awarded reward points by a third party or by the social media platform.

[0133] In some embodiments, a user may be provided a suggestion that there are other features or actions associated with a specific object that may be captured using a camera and uploaded. For example, a logo design or a symbol may be placed on the object or rendered over a media item for broadcast (e.g., watermark). The user, upon seeing the particular logo or symbol may leverage a nearby camera to capture an image of the scene or media or object displaying the logo or symbol and upload it to a retrieval system as described herein.

[0134] These foregoing and other embodiments are discussed below with reference to FIGs. 1 - 10. However, those skilled in the art will readily appreciate that the detailed description given herein with respect to these figures is for explanation only and should not be construed as limiting.

[0135] FIG. 1 depicts an example computing network in, or over which embodiments as described herein may be implemented. As shown in FIG. 1, an example computing network implemented as a system 100. The system 100 includes a client device 102 communicatively coupled via a network to a gateway 104. The gateway 104 may further provide coupling between the client device 102 and an intake system 106 and/or a retrieval system 108. The system 100 can be leveraged by the client device 102 in an intake mode or a retrieval mode, both of which are described in greater detail below. Generally, and broadly, when operated in an intake mode, the client device 102 is configured to upload one or more images of an object or scene to the intake system 106. With this input, the intake system 106 is configured to identify enrichable

content within the scene and to associate one or more selected actions with that enrichable content.

5 [0136] When operated in a retrieval mode, the client device 102 is configured to upload one or more images of an object or scene to the retrieval system 108. With this input, the retrieval system 108 is configured to leverage previously-identified content (e.g., content previously uploaded to the intake system 106) and to cause to be executed one or more actions associated with the identified content. Each of these operations are described in greater detail below, however it may be appreciated that in many embodiments a client device operating the system 100 in an intake mode may be different from a client device operating the system 100 in a retrieval mode; the client device 102 is depicted as operable in both modes only for simplicity of description and illustration.

10 [0137] In some embodiments, the client device 102 may be a phone, a tablet, a smartwatch, a laptop, a computer, the internet-of-Things (IoT), and electronic device that has at least one imaging system (e.g., camera) and a transceiver to communicate with the gateway 104 via the network.

15 [0138] In an intake mode, a user of the client device 102 may cause the device to transmit an image or video (and/or sequence of images) captured with a camera system of the client device 102 to either the intake system 106 via the gateway 104 for further processing.

20 [0139] As noted above, the subject(s) of an image or video (and/or sequence of still frames) may be referenced herein as enrichable content as one or more actions that may be associated with the image or video, and the content may be enriched with additional information, such as one or more actions that may be triggered when the same or another user captures an image or a video that has similarity to the image or the video that is enriched with the one or more associated actions.

25 [0140] In some embodiments, the image or video may be captured using an application executing on the client device 102, or using a web interface providing a web connection to the gateway 104. Accordingly, in the intake mode, the application executing on the client device 102 or the web interface providing the web connection to the gateway 104 may present to the user options whether the user would like to associate one or more actions with the enrichable content or the user would like to retrieve any actions that may be associated with the image or video.

30 [0141] In some embodiments, based on the user's selection of one or more options (presented by the application, for example in a graphical user interface thereof) executing on the client device or the web interface, the captured images or video may be transmitted to the gateway 104 for forwarding to the intake system 106. To route these transmissions, the gateway 104 can

leverage a flag, attribute, or header indicating that the images are intended for the intake system 106.

[0142] As a result of this construction, the intake system 106 may receive from the client device 102 via the gateway 104 over the network one or more images and/or one or more frames of a video (e.g., the frames 112) at an intake pipeline 110.

[0143] The intake pipeline 110 may comprise one or more servers, applications, libraries, functions, modules, algorithms, and/or machine learning models/algorithms, and so on, for processing of the frames 112 and providing as output an index identifying particular content within the object/scene and associating that content to one or more actions/enrichments. The index(es) generated by the pipeline can be stored in a database 114.

[0144] In the retrieval mode, the retrieval system 108 may receive from the client device 102 via the gateway 104 over the network one or more images and/or one or more frames of a video (e.g., the frames 118) at a retrieval pipeline 116. As with the intake pipeline 110, the retrieval pipeline 116 may comprise one or more servers, applications, libraries, functions, modules, algorithms, and/or machine learning models/algorithms, and so on, for processing of the frames 118, which may have a different resolution and/or may be presented at a different frame rate than the frames 112. Processing of the frames 118 by the retrieval pipeline 116 is described in detail below.

[0145] As with the intake pipeline 110, the retrieval pipeline 116 may determine an index of a database record corresponding to the enriched content that is matching with enrichable content stored at the index. The retrieval pipeline 116 may then identify one or more actions stored in a database 120, which may be a part of the retrieval system 108 or the intake system 106; in either construction, the database 120 may be communicably coupled to either or both the intake system 106 and the retrieval system 108.

[0146] The retrieval pipeline 116 may then communicate the retrieved one or more actions to an action distributor 122. The action distributor 122 may analyze the one or more actions – and/or metadata or descriptive data thereof – and transmit the one or more actions to appropriate destinations for execution and/or other handling. For example, the action distributor 122 may be configured to identify at least one action that should be executed by the client device 102; in other words, certain actions may be intended to benefit, or receive input from, or provide output to a user 124, which may be a user of the client device 102. In these examples, the action distributor 122 may generate one or more messages or notifications or instructions to be received by the client device 102, which in response may modify a graphical user interface, may retrieve sensor output (GPS, accelerometer, compass, and so on) and forward that sensor output to

another service, may generate a notification and so on. Many potential actions may be performed by the client device 102.

[0147] In some embodiments, the action distributor 122 may also communicate with an administrator or an admin console 126 for various purposes such as logging, recording, and/or troubleshooting, and so on.

[0148] The action distributor 122 may also communicate with a third-party system 128 for example to retrieval additional actions from the third-party system 128 and/or to update the third-party system of the particular enriched content uploaded by a user.

[0149] The action distributor 122 may also communicate with an action executor 130. The action executor 130 cause performing of the one or more actions based on the user selection.

[0150] FIG. 2 depicts an example intake system, as described herein. The intake system 200 may include one or more application servers executing instances of one or more algorithms, e.g., machine-learning algorithms, artificial intelligence-based algorithms, computer-vision based algorithms, which may be referenced as a machine-learning algorithm in the present disclosure.

Collectively, these are represented in the figure as the high-performance machine learning algorithm 206.

[0151] The one or more algorithms may be configured to process one or more images and/or one or more frames 204 captured of an object or scene uploaded by a client device 202. The one or more images and/or one or more frames of a video may be received at an intake system via a gateway, such as the gateway 104 of FIG. 1.

[0152] As with many embodiments described herein, the high-performance machine learning algorithm 206 may analyze the received enrichable content, and identify one or more objects in the scene, and relative positions thereof.

[0153] In some embodiments, the high-performance machine learning algorithm 206 may be used to identify one or more objects included in the scene and provide an object list to an object cluster identifier 208.

[0154] The object cluster identifier 208 be configured to identify groupings of objects identified by the high-performance machine learning algorithm 206. In one example, the list of objects identified in the scene may be created using a bounding box surrounding each object in the enrichable content that can be identified.

[0155] A bounding box may also describe a spatial location of an object in the scene, in two or three dimensions. Accordingly, spatial relationships between two or more objects identified in the enrichable content may also be determined. In some embodiments, when the scene contains

an object (e.g., an enrichable content item) that is movable, a temporal relationship between objects identified in the enrichable content may also be determined.

[0156] In some embodiments, a single identifier can be used to reference the one or more clusters of objects 210 output by the object cluster identifier 208. For example, a vectorizer 212 can be configured to consume a set of clusters and to provide as output a single vector 214, which in turn can be identified by a single index value.

[0157] In this manner, each recognizable enrichable object or scene can have its content summarized by a single vector, which in turn can be indexed. As described above, in some cases, more than one vector can be used to describe a single scene. For example, some vectors may be created based on different image manipulations of input images (e.g., color adjustments, scale changes, and so on) and some vectors may be based on different sets of identifiable objects. For example, a first vector describing a scene may reject or ignore human persons whereas another vector describing the same scene may not reject or ignore labels of human persons. In this manner, recognition accuracy by a retrieval system such as described herein can be improved.

[0158] In some embodiments, each vector 214 corresponding to the one or more clusters of objects 210 may be normalized in accordance with criteria including one or more of: a data size of a vector, a particular physical dimension of a vector, and so on.

[0159] The classifier 216 may be executing on the one or more application servers of the intake system 200 and may store the set of vectors into an index database 218. An index corresponding to a database record storing the set of vectors in the index database 218 may be retrieved by the classifier 216 and may be further associated with the set of actions stored in an action database 220. Accordingly, the enrichable content uploaded by a user of the client device and other generated enrichable content having different properties, as described herein, may be associated with the set of actions 222 as provided by the user of the client device, and an acknowledgement or confirmation may be sent back to the user for displaying on the client device 202.

[0160] As described herein, the intake system 200, receives enrichable content with associated one or more actions, which may be then retrieved using a retrieval system described using FIG. 3.

[0161] FIG. 3 depicts an example retrieval system, as described herein. A retrieval system 300 may include one or more application servers executing a scene or a boundary detector 306 and algorithms 308, e.g., machine-learning algorithms, artificial intelligence-based algorithms, computer-vision based algorithms, which may be referenced as a machine-learning algorithm 308 in the present disclosure.

[0162] The scene or boundary detector 306 may receive frames 304 from a client device 302 via a gateway, such as the gateway 104, as described herein. The frames 304 may correspond contain enrichable content previously uploaded to an intake system by the same or another user.

[0163] The scene or boundary detector 306 may identify one or more objects in the frames 304 using techniques described herein. Each object identified in the content may has a particular spatial relationship with each other classified object. If the uploaded content is a sequence of frames, then objects identified in the content may also have temporal relationship with each other.

[0164] An output of the scene or boundary detector 306, which may include one or more objects with its corresponding bounding box may then then be processed though the machine-learning algorithm 308. The machine-learning algorithm 308 may be high-speed algorithm that prioritizes speed over accuracy or precision.

[0165] The machine-learning algorithm 308 may operate with an object cluster identifier 310 to generate object clusters, which may be represented as structured data and/or as the object clusters 312.

[0166] In some embodiments, a classifier 318 may compare the vector 316 with one or more vectors stored in an index database 320 (which may be the index database 218 of FIG.).

[0167] The classifier 318 may identify a vector stored in the index database 320 that matches with the vector 316 according to particular criteria. The particular criteria may include at least a specific number of objects in the vector 316 and the matching vector stored in the index database 320, spatial relationship between two or more objects that match above a particular threshold value, and so on. In some cases, sparse representation classification can be used to identify which among the set of vectors is a closest match to the vector 316. Once a match is determined within a tolerance or threshold, an index corresponding to that match can be returned.

[0168] The index corresponding to the vector stored in the index database 320 that matches the vector 316 may be used to search an action database 322. The action database 322 may be the action database 220. In this manner, and as a result of the depicted architecture, an index of records storing one or more actions associated with the frames 304 may be used to retrieve an actions list 324.

[0169] In some embodiments, the retrieved actions list 324 may be then sent to an action distributor 326 for distributing to various stakeholders, such as a third-party 328, an administrator 332, and/or a user of the client device 302. In some embodiments, the retrieved actions list 324 may be reported to the third-party 328, for example, for informing the third-party 328 about a location of the user of the client device 302. The third-party may also display or

push information about their services and/or sales to the client device 302 through the action distributor 326.

[0170] The third-party 328 may be any party, person, and/or an entity which may be interested to know the user of the client device 302 has uploaded frames 304 that is associated with one or more actions identified in the actions list 324. The administrator 332 may receive information of the frames 304 and the corresponding one or more actions of the actions list 324 for logging, monitoring, and/or troubleshooting, and so on. In some embodiments, the retrieved actions list 324 may be displayed on the client device 302 of the user, and may be performed by an action executor 330 in accordance with selection of a particular action of the action list displayed on the client device.

[0171] These foregoing embodiments depicted in FIGs. 1-3 and the various alternatives thereof and variations thereto are presented, generally, for purposes of explanation, and to facilitate an understanding of various configurations and constructions of a system, such as described herein. However, it will be apparent to one skilled in the art that some of the specific details presented herein may not be required in order to practice a particular described embodiment, or an equivalent thereof.

[0172] Thus, it is understood that the foregoing and following descriptions of specific embodiments are presented for the limited purposes of illustration and description. These descriptions are not targeted to be exhaustive or to limit the disclosure to the precise forms recited herein. To the contrary, it will be apparent to one of ordinary skill in the art that many modifications and variations are possible in view of the above teachings.

[0173] FIG. 4A depicts an example computing environment corresponding to an intake system in, or over which, embodiments as described herein may be implemented. As shown in a computing environment 400a, a client device 402 may include a processor 404, a memory 406, a display 408, an input system 410, a camera 412 (which may be an event driven/neuromorphic camera, a CMOS camera, CCD camera or any other suitable camera operating in visible bands, IR bands, UV bands or multiple bands thereof), and a sensor 414.

[0174] A single instance of a processor, a memory, a display, an input system, a camera, and/or a sensor are shown in FIG. 4A, but it is appreciated that there may be more than one instance of the processor 404, the memory 406, the display 408, the input system 410, the camera 412, and the sensor 414 present in the client device 402.

[0175] In these examples the client device 402 can leverage the processor 404 to access an executable asset from the memory 406 to instantiate an instance of software configured to access an intake system and/or a retrieval system, such as described herein. The instance of software

can be referred to as a client application, a frontend application, a browser application, a native application, or by another name. In some examples, the instance of software may be a portion of a kernel or operating system of the client device 402.

[0176] An intake system, as described herein, may include an application server 416 may include one or more resource allocations 418, which can include processing resources and memory resources. The application server 416 (also referred to as a host server) may be connected to an index database 420 and an action database 422.

[0177] In these examples the application server 416 can leverage the processor allocation to access an executable asset from the memory allocation to instantiate an instance of software configured to operate as at least a portion of an intake system and/or a retrieval system, such as described herein. The instance of software can be referred to as a server application, a backend application, a host service, or by another name. In some examples, the instance of software may be a portion of a kernel or operating system of the application server 416.

[0178] In FIG. 4A, the index database 420 and the action database 422 are shown separate, but this is not required. There may be only one database having tables corresponding to the index database 420 and the action database 422. As described herein the index database 420 may store identifiers of enrichable content as database records, and the action database 422 may store one or more actions associated with enrichable content as cross referenced with one or more indexes of the index database 420.

[0179] The client device 402 may be communicatively coupled with the application server 416 of the intake system via a gateway, such as the gateway 104 over a network, although this is not required. In some embodiments, the network may include the open internet.

[0180] In some embodiments, the client device 402 may be a phone, a computer, a laptop, a tablet, a smartwatch, and/or another suitable device, whether portable or stationary. The processor 404 of the client device 402 may be any suitable computing device or logical circuit configured to execute one or more instructions to perform or coordinate one or more operations on or to digital data. In many embodiments, the processor or processors of the client device 402 may be a physical processor, although this is not required of all embodiments; virtual components may be suitable in some implementations. Similarly, the memory 406 of the client device 402 may be configured and/or implemented in a number of suitable ways and may be partially or completely virtualized.

[0181] As noted above, the processor 404 of the client device 402 is configured to access at least one executable asset from the memory 406 of the client device 402. More particularly, the processor 404 of the client device 402 may be configured to access a data store portion of the

memory to load, into a working portion of the memory, at least one executable asset or executable program instruction. In response to loading the instruction or executable asset into working memory, the processor 404 of the client device 402 may instantiate an instance of software referred to herein as a client application. The client application and its corresponding user interface is described using, for example, FIG. 6 below.

5 **[0182]** Using the input system 410, a user of the client device 402 may perform various functions, including but not limited to, launch a client application, capture one or more images and/or video of an object and/or a scene using the camera 412 of the client device 402.

10 **[0183]** The sensor 414 may be a GPS sensor which may be used to determine a current location of the client device 402; this data can be received by the intake system and may be associated to a particular enriched content. In addition, there may be other type of sensors as well such as, an inertial measurement unit sensor for detecting speed and movement-related information of the client device 402 according to the movement of a user of the client device 402.

15 **[0184]** As described herein the client device may communicate with the application server 416 one or more images (data), one or more frames of video, and/or one or more actions, and receive acknowledgement or confirmation according to an outcome of associating the one or more actions with the one or more images and/or the one or more frames of video as specified by the user of the client device 402.

20 **[0185]** In some embodiments, the one or more resource allocation functions/modules 418 may allocate resources, including but not limited to, a processor or a computational resource, a memory, network usage or bandwidth, and so on. The application server 416 may be executing a server application. In some cases, there may be more than one instance of the server application executing on the application server 416.

25 **[0186]** FIG. 4B depicts an example computing environment corresponding to a retrieval system in, or over which, embodiments as described herein may be implemented. As shown in a computing environment 400b, a client device 424 may include a processor 426, a memory 428, a display 430, an input system 432, a camera 434, and a sensor 436.

30 **[0187]** Even though a single instance of a processor, a memory, a display, an input system, a camera, and/or a sensor are shown in FIG. 4B, there may be more than one instance of the processor 426, the memory 428, the display 430, the input system 432, the camera 434, and the sensor 436 present in the client device 424. A retrieval system may include an application server 438 may include one or more resource allocation functions/modules 440. The application server 438 may be connected to an index database 442 and an action database 444.

[0188] Even though, in FIG. 4B, the index database 442 and the action database 444 are shown separate, there may be only one database storing tables corresponding to the index database 442 and the action database 444. As described herein the index database 442 may store identifiers of the one or more objects of the enrichable content as database records, and the
5 action database 444 may store one or more actions associated with the enrichable content as cross referenced with one or more indexes of the index database 420 corresponding to the enrichable content and other content having different properties generated from the enrichable content.

[0189] The client device 424 (which may be the same client device as shown in FIG. 4A or
10 may be a different client device) may be communicatively coupled with the application server 438 of the retrieval system via a gateway, such as the gateway 104 over a network, although this is not required. In some embodiments, the network includes the open internet.

[0190] In some embodiments, the client device 424 may be a phone, a computer, a laptop, a tablet, a smartwatch, and/or another suitable client device. The processor 426 of the client device
15 424 may be any suitable computing device or logical circuit configured to execute one or more instructions to perform or coordinate one or more operations on or to digital data. In many embodiments, the processor or processors of the client device 424 may be a physical processor, although this is not required of all embodiments; virtual components may be suitable in some implementations. Similarly, the memory 428 of the client device 424 may be configured and/or
20 implemented in a number of suitable ways and may be partially or completely virtualized.

[0191] In some embodiments, the processor 426 of the client device 424 is configured to access at least one executable asset from the memory 428 of the client device 424. More particularly, the processor 426 of the client device 424 may be configured to access a data store
25 portion of the memory to load, into a working portion of the memory, at least one executable asset or executable program instruction. In response to loading the instruction or executable asset into working memory, the processor 426 of the client device 424 may instantiate an instance of software referred to herein as a client application. The client application and its corresponding user interface is described with reference to FIG. 6 below.

[0192] Using the input system 432, a user of the client device 424 may perform various
30 functions, including but not limited to, launch a client application, capture one or more images and/or video of an object and/or a scene using the camera 434 of the client device 424. The sensor 436 may be a GPS sensor which may be used to determine a current location of the client device 424. In addition, there may be other type of sensors as well such as, an inertial

measurement unit sensor for detecting speed and movement related information of the client device 424 according to the movement of a user of the client device 424.

[0193] As described herein the client device 424 may communicate with the application server 438 one or more images and/or one or more frames of video and receive one or more actions corresponding to the one or more images and/or frames of video according to an outcome of processing of the one or more images and/or frame of video by the application server 438.

[0194] In some embodiments, the one or more resource allocation functions/modules 440 may allocate resources, including but not limited to, a processor or a computational resource, a memory, network usage or bandwidth, and so on. The application server 438 may be executing a server application. In some cases, there may be more than one instance of the server application executing on the application server 438.

[0195] Further, it may be appreciated that although referred to as a singular “application server” in FIG. 4A and/or FIG. 4B, a host server supporting the backend may be a cluster of different computing resources, which may be geographically separated from one another. In this manner, because specific implementations may vary, the application server and the client device may be referred to, simply, as “computing resources” configured to execute purpose-configured software (e.g., the client application and the backend application).

[0196] As used herein, the general term “computing resource” (along with other similar terms and phrases, including, but not limited to, “computing device” and “computing network”) may be used to refer to any physical and/or virtual electronic device or machine component, or set or group of interconnected and/or communicably coupled physical and/or virtual electronic devices or machine components, suitable to execute or cause to be executed one or more arithmetic or logical operations on digital data.

[0197] Example computing resources contemplated herein include, but are not limited to: single or multi-core processors; single or multi-thread processors; purpose-configured co-processors (e.g., graphics processing units, motion processing units, sensor processing units, and the like); volatile or non-volatile memory; application-specific integrated circuits; field-programmable gate arrays; input/output devices and systems and components thereof (e.g., keyboards, mice, trackpads, generic human interface devices, video cameras, microphones, speakers, and the like); networking appliances and systems and components thereof (e.g., routers, switches, firewalls, packet shapers, content filters, network interface controllers or cards, access points, modems, and the like); embedded devices and systems and components thereof (e.g., system(s)-on-chip, Internet-of-Things devices, and the like); industrial control or automation devices and systems and components thereof (e.g., programmable logic controllers,

programmable relays, supervisory control and data acquisition controllers, discrete controllers, and the like); vehicle or aeronautical control devices systems and components thereof (e.g., navigation devices, safety devices or controllers, security devices, and the like); corporate or business infrastructure devices or appliances (e.g., private branch exchange devices, voice-over internet protocol hosts and controllers, end-user terminals, and the like); personal electronic devices and systems and components thereof (e.g., cellular phones, tablet computers, desktop computers, laptop computers, wearable devices); personal electronic devices and accessories thereof (e.g., peripheral input devices, wearable devices, implantable devices, medical devices and so on); and so on. It may be appreciated that the foregoing examples are not exhaustive.

5
10 **[0198]** These foregoing embodiments depicted in FIGs. 4A-4B and the various alternatives thereof and variations thereto are presented, generally, for purposes of explanation, and to facilitate an understanding of various configurations and constructions of a system, such as described herein. However, it will be apparent to one skilled in the art that some of the specific details presented herein may not be required in order to practice a particular described
15 embodiment, or an equivalent thereof.

[0199] Thus, it is understood that the foregoing and following descriptions of specific embodiments are presented for the limited purposes of illustration and description. These descriptions are not targeted to be exhaustive or to limit the disclosure to the precise forms recited herein. To the contrary, it will be apparent to one of ordinary skill in the art that many
20 modifications and variations are possible in view of the above teachings.

[0200] FIGs. 5A-5G depict various example use cases or practical applications of embodiments described herein. FIG. 5A depicts an example operation of a retrieval system 500a as described herein.

[0201] The figure includes a client device 502, such as a smartphone is shown in front of a
25 television 504. A program with a current scene 508 may have a particular logo or symbol 506 displayed on a display screen of the television 504. The particular logo or symbol may suggest a user present in the room that there may be one or more actions associated with a particular content being displayed on the television 504. The logo may be animated or static, may be positioned suitably anywhere within the television active display area, may be added over the
30 media content by a broadcaster and/or may be encoded into the media itself, and so on. the logo may be rendered by a set top box. In some cases, the logo may be colored to maximize contrast with surrounding pixels.

[0202] A user may use an input system of the client device 502 to launch an application that may leverage a camera (CMOS, CCD, neuromorphic, and so on) of the client device 502 to

capture one or more images and/or one or more frames of the program as a video (“content”).

The application executing on the client device may then communicate the content to a retrieval system, as described herein. The retrieval system may then process the content to identify one or more actions associated with the content received from the client device 502, and transmit to the client device for the user to select and take an action.

[0203] As shown in FIG. 5A, user may not be required to focus or zoom the camera of the client device 502. More generally, the camera of the client device 502 may capture an image including the current scene 508 being displayed on the display screen of the television 504, a television frame, and other objects, such as a portion of a television console, and so on. The particular logo or symbol 506 may not (in some embodiments) carry any information; in other

[0204] As described herein one or more actions may be associated with a particular object within a scene rendered on the television. When a user uploads one or more images taken using a client application executing on the client device 502, one or more objects may be identified from the uploaded one or more images by a retrieval system as described herein.

[0205] The retrieval system may identify one or more actions that may be associated with an object recognized within the active display area of the television. The one or more actions retrieved from a database may be communicated to the client device 502 for the user to select and execute.

[0206] As shown in FIG. 5B, for a use case 500B, a user 510 using a client device 512, such as a smartphone may capture one or more images or videos of particular program being displayed on a display screen of a television 514. As shown in FIG. 5B, a currently being displayed scene may include a particular logo or symbol 518, which may suggest the user 510 that there are one or more actions associated with the program being displayed on the television.

The user may retrieve one or more actions associated with the program by taking one or more images or videos using a camera of the client device 512 via a client application executing on the client device 512. As shown in FIG. 5B, the user may capture image and/or video from any angle and/or under any lighting condition. However, relevant objects from the one or more images and/or video frames may be identified, and one or more actions associated with the relevant objects may be identified. As described herein one or more actions may have been associated with enrichable content uploaded by a user, and also content that may be generated having different properties, such as brightness, contrast, temperature, tint, hue, gamma, color, blur, color tint, an angle at which an image may be taken, and/or an aspect ratio, and so on. Accordingly, the

content uploaded by the user 510 is not required to be exactly identical with content used to associated one or more actions with the content.

[0207] In FIG. 5C, a view 500c illustrates a client application view of a client application executing on a client device 520. As shown in FIG. 5C, the client application executing on the client device may use a camera of the client device 520 for taking one or more images and/or videos of a program currently being displayed on a television as a view 522.

[0208] While the view 522 is being displayed and/or captured using the camera, a notification 524 or other visual or haptic indication of detecting enrichable content may be displayed or triggered. The notification may be displayed based on detection of the particular logo or symbol 518 in a known location. For example, the notification may be displayed when a camera may be used by the client application to capture one or more images and/or videos whether or not one or more actions are associated with the particular content.

[0209] In FIG. 5D, a view 500d illustrates one or more actions retrieved by a retrieval system corresponding to the one or more images and/or videos uploaded to the retrieval system by the user 510. As shown in FIG. 5D and for example, one or more actions displayed on a client device 526 may include a movie trailer shown as a picture-in-picture or an overlay, buy movie tickets 530, buy merchandise 532, use digital content as digital twin 534, and/or other actions 536.

[0210] The retrieved actions, in the present example, may be associated with, for example, an upcoming movie, and the user may have taken one or more images and/or videos of particular movie trailer or advertisement being displayed on the television 514. Accordingly, any number of actions and/or any type of actions may be associated with the enrichable content for retrieval by a user.

[0211] FIG. 5E illustrates a use case as 500e in which functionality of a QR code can be enabled despite that the QR code itself is unreadable for a particular user's client device. However, as noted above, for successful scanning of the QR code, the user is required to be within a threshold distance of the QR code.

[0212] For example, a user may capture an image of a building having an entrance door 540 with a QR code 542 displayed. The features and objects of the entrance door 540, along with optionally supplemental information such as GPS information, can be used by an intake system to uniquely identify the entrance door 540 and/or the building.

[0213] In this manner, when another user captures an image of the building with a camera of a client device 538 (through a client application executing on the client device 538), and transmits to a retrieval system, the retrieval system may identify that the recognized building is

associated with a QR code. As a result, the user may be notified with a graphical user interface element 544 that a distant QR code has been detected and a user may execute an associated action by selecting a second user interface element 546 to trigger an action associated with the QR code 542.

5 **[0214]** FIG. 5F illustrates a use case 500f, in which a distant QR code is occluded, hidden, or blocked. In some cases, the distant QR code may have been damaged. Accordingly, when a user captures an image, using a camera of a client device 548 through a client application executing on the client device 548, of the building in which the entrance door 540 is seen having the QR code 542 blocked, for example, by a person standing in front of the QR code 542.

10 **[0215]** As described herein a retrieval system may identify based on the image transmitted to it from the client device 548, that the user has transmitted an image of the building, which has a QR code, and a corresponding action associated with the QR code. As a result, a notification may be displayed to the user that an occluded QR code has been detected, and the user can select a user interface element to trigger an action associated with the QR code 542.

15 **[0216]** FIG. 5G illustrates yet another use case 500g, in which an out-of-range NFC and/or a Bluetooth low energy beacon may be detected as within an image of a particular already-scanned scene. As shown in FIG. 5G, similar to the example described above using FIG. 5E and/or FIG. 5F, a user may associate one or more actions that may be accessible only when a client device within a specific distance range supported by a near-field communication and/or a Bluetooth low energy (BLE) range. However, a user may upload enrichable content to an intake system and associate it with actions available when a client device is with NFC and/or BLE range.

20 **[0217]** For example, an NFC tag 562 may be present on a table 560, and a user may trigger one or more actions by enabling an NFC mode on a client device. However, for a user having a client device 558 that is not within the NFC range (or having a device incapable of NFC reading), may capture an image of the table 560, for example, and transmit the image to a retrieval system.

25 **[0218]** As described herein the retrieval system may identify one or more objects in the image uploaded from the client device 558 and identify that the user has transmitted an image which has one or more actions available through NFC and/or BLE. Accordingly, a notification 564 may be displayed on the client device 558 indicating an out-of-range NFC and/or BLE is detected, so that the user may select and execute functions available using NFC and/or BLE but without being in proximity from the NFC tag and/or BLE of a particular NFC and/or BLE range.

30 **[0219]** In the above use cases or practical applications, the use case or application is described using an image. However, the embodiments described herein may also be relevant when a video

including one or more video frames are uploaded by a user to an intake system and/or a retrieval system.

[0220] These foregoing embodiments depicted in FIGs. 5A-5G and the various alternatives thereof and variations thereto are presented, generally, for purposes of explanation, and to facilitate an understanding of various configurations and constructions of a system, such as described herein. However, it will be apparent to one skilled in the art that some of the specific details presented herein may not be required in order to practice a particular described embodiment, or an equivalent thereof.

[0221] Thus, it is understood that the foregoing and following descriptions of specific embodiments are presented for the limited purposes of illustration and description. These descriptions are not targeted to be exhaustive or to limit the disclosure to the precise forms recited herein. To the contrary, it will be apparent to one of ordinary skill in the art that many modifications and variations are possible in view of the above teachings.

[0222] FIG. 6 depicts an example user interface of a client application executing on a client device, in accordance with some embodiments. As shown in a view 600, a client application executing on a client device 602 may provide one or more menu options to a user while communicating with an intake system described herein in accordance with some embodiments. As described herein, a user may upload enrichable content including one or more images and/or video frames of a video and associate one or more actions to the uploaded enrichable content. Accordingly, a client application executing on the client device 602 may present to a user of the client device options, such as upload or scan content 604, set action(s) 606, add action(s) 608, update action(s) 610, delete action(s) 612, and/or delete content 614. The user may save or transmit the enrichable content and/or actions by selecting a save option 616.

[0223] In some embodiments, the upload/scan content may allow the user to access a camera of the client device to capture an image and/or a video of a particular scene and/or object. The user may then associate one or more actions using set action(s) 606 options. As described herein, the set of actions may include, for example, purchase ticket, play a trailer, purchase merchandize, purchase/lease digital content for use in metaverse, and so on. The set of actions may also include an action associated with a specific QR code, an NFC tag, and/or a BLE.

[0224] The user may add additional actions using add action(s) 608 options and/or update or delete actions using update action(s) 610 or delete action(s) 612 when the user has previously associated one or more actions with the particular enrichable content. Similarly, the user may delete enrichable content using delete content 614 option, and thereby, could also remove all actions that may have been associated with the deleted enrichable content. The save option 616

may allow the user to transmit the enrichable content and corresponding user action associated with the enrichable content to an intake system described herein in accordance with some embodiments.

5 [0225] These foregoing embodiments depicted in FIG. 6 and the various alternatives thereof and variations thereto are presented, generally, for purposes of explanation, and to facilitate an understanding of various configurations and constructions of a system, such as described herein. However, it will be apparent to one skilled in the art that some of the specific details presented herein may not be required in order to practice a particular described embodiment, or an equivalent thereof.

10 [0226] Thus, it is understood that the foregoing and following descriptions of specific embodiments are presented for the limited purposes of illustration and description. These descriptions are not targeted to be exhaustive or to limit the disclosure to the precise forms recited herein. To the contrary, it will be apparent to one of ordinary skill in the art that many modifications and variations are possible in view of the above teachings.

15 [0227] FIGs. 7A-7B depict an example of object identification and generation of object clusters, in accordance with some embodiments. As described herein, and as shown in FIG. 7A as a view 700A, a user may be located anywhere in a room in which a television 704 is present and showing a program having a scene as shown in the view 700a. Based on a particular location of a user, and based on a specific zoom range set by a user for a camera of a client device 702, a user may have many objects, such as a television 704, a television console, and so on, in an image. In these examples, raw imagery captured by the client device 702 can be subdivided into multiple sections, and each section can be provided to a retrieval system as described herein and/or to a crude object detector, such as described herein. In some cases, segments of an image can be sequentially scanned, starting centrally and working outward.

20 [0228] As shown in FIG. 7B, as a view 700B, however when the image having many objects is transmitted to a retrieval system, as described herein, individual segments of the image may be identified as separately-enriched objects. In some embodiments, a central portion 706 of the image may only be analyzed for identifying one or more objects in the image. Accordingly, extraneous information included in content may be excluded from processing by the intake system or retrieval system. The central portion 706 of the image may be further divided into a number of sections 706A-706I for identifying one or more objects in the central portion 706, and spatial relationship between the one or more objects identified in the central portion 706.

30

[0229] Based on the one or more objects identified based on the number of sections 706A-706I, object clusters may be generated as described herein by the intake system and/or retrieval system.

[0230] These foregoing embodiments depicted in FIGs. 7A-7B and the various alternatives thereof and variations thereto are presented, generally, for purposes of explanation, and to facilitate an understanding of various configurations and constructions of a system, such as described herein. However, it will be apparent to one skilled in the art that some of the specific details presented herein may not be required in order to practice a particular described embodiment, or an equivalent thereof.

[0231] Thus, it is understood that the foregoing and following descriptions of specific embodiments are presented for the limited purposes of illustration and description. These descriptions are not targeted to be exhaustive or to limit the disclosure to the precise forms recited herein. To the contrary, it will be apparent to one of ordinary skill in the art that many modifications and variations are possible in view of the above teachings.

[0232] FIG. 8 depicts a flowchart corresponding to example operations of a method being performed by an intake system, in accordance with some embodiments. The method 800 includes the operation 802 at which enrichable content and a set of actions may be received at an intake system, as shown herein in accordance with some embodiments in FIG. 1, FIG. 2, and/or FIG. 4A.

[0233] In some embodiments, the enrichable content received at the intake system may include one or more images and/or videos taken using a camera of a client device, such as the client device 102, 202, and/or 402. A user of the client device 102, 202, and/or 402 may launch an application and/or a web interface to take the one or more images and/or video using the camera of the client device. In some embodiments, a user may transmit the enrichable content, and one or more actions associated with the enrichable content using an interface shown in FIG. 6. The enrichable content may be received at the intake system via a gateway over a network.

[0234] At the operation 804, object clusters may be generated in which one or more objects in the received enrichable content may be identified. For example, a predetermined area of the enrichable content may be analyzed for determining one or more objects. The predetermined area may be a central portion of an image. In some embodiments, the predetermined area may be selected based on blurring of an image. Accordingly, a picture area without blurring may be selected for analysis to generate object clusters as described herein.

[0235] At the operation 806, an identifier of a number of objects in object clusters generated based on the enrichable content received at the operation 802 may be generated. In some

embodiments, one or more objects of the object clusters may be converted into a single vector representing the one or more objects and/or object clusters. In some embodiments, the generated vector may be normalized. The generated vector may then be stored as a database record in a database, for example, an index database.

5 [0236] At the operation 808, a set of actions received at the operation 802 from a client device using a user interface shown in FIG. 6 may be associated with a vector generated at the operation 806 above, and stored as a database record in a database, such as an action database, along with an index of a database record storing the generated vector. As described herein, in some embodiments, the action database and the index database may be a single database.

10 [0237] FIG. 9A depicts a flowchart corresponding to example operations of a method being performed by a retrieval system, in accordance with some embodiments. As shown in a method 900a, at the operation 902, enrichable content may be received at a retrieval system, as shown herein in accordance with some embodiments in FIG. 1, FIG. 3, and/or FIG. 4B.

[0238] In some embodiments, the enrichable content received at the retrieval system may include one or more images and/or videos taken using a camera of a client device, such as the client device 102, 302, and/or 424. A user of the client device 102, 302, and/or 424 may launch an application and/or a web interface to take the one or more images and/or video using the camera of the client device. In some embodiments, a user may transmit the enrichable content to the retrieval system via a gateway over a network.

20 [0239] At the operation 904, object clusters may be generated in which one or more objects in the received scene image may be identified. For example, a predetermined area of the enrichable content may be analyzed for determining one or more objects. The predetermined area may be a central portion of an image. In some embodiments, the predetermined area may be selected based on blurring of an image. Accordingly, a picture area without blurring may be selected for analysis to generate object clusters as described herein.

[0240] At the operation 906, an identifier of a number of objects in object clusters generated based on the enrichable content received at the operation 902 may be generated. In some embodiments, one or more objects of the object clusters may be converted into a single column representing the one or more objects of the object clusters. In some embodiments, the generated vector may be normalized, as described herein in accordance with criteria including one or more of: a data size of a vector, a particular physical dimension of a vector, and so on. The generated vector may then be compared with one or more stored vectors in a database, for example, an index database, at 806.

30

[0241] If one or more vectors stored in the index database are found to be matching above a specific threshold, then it may be checked whether each of the one or more vectors stored in the index database correspond to the same enrichable content and/or content generated having different properties based on the received enrichable content. If it is determined that the one or more vectors stored in the index database and found matching with a vector generated at the operation 906 all correspond to the same enrichable content and/or content generated having different properties based on the received enrichable content, then an index corresponding to any of the one or more vectors may be identified for further processing at the operation 908.

[0242] However, if it is determined that the one or more vectors stored in index database and found matching with a vector generated at the operation 906 do not correspond to the same enrichable content and/or content generated having different properties based on the received enrichable content, then an index corresponding to a vector having the best match may be identified for further processing at the operation 908.

[0243] At the operation the operation 908, a set of actions associated with an index of the vector stored in index database, and as determined at the operation 906 above, may be retrieved from an actions database. At the operation 910, the retrieved set of actions may be returned to the user for displaying on the user's client device. The user may then select one or more actions for execution.

[0244] FIG. 9B depicts another flowchart corresponding to example operations of a method being performed by a retrieval system, in accordance with some embodiments. As shown in a method 900b, at the operation 912, the set of actions retrieved at the operation 908 may include an action identifying whether the enrichable content is available for use in metaverse. For example, a user may have uploaded an image of particular clothing, which may be available not only as a physical purchase but may only be available as a digital purchase or lease.

Accordingly, based on the determination made at the operation 912, at the operation 914, one or more options related to purchase and/or lease of the digital content in metaverse may be presented to a user.

[0245] FIG. 10 depicts a flowchart corresponding to example operations of a method being performed by a retrieval system, in particular, for an occluded QR code and/or an out-of-range NFC and/or BLE beacon, in accordance with some embodiments.

[0246] As shown in a method 1000, at the operation 1002, enrichable content may be received at a retrieval system, as shown herein in accordance with some embodiments in FIG. 1, FIG. 3, and/or FIG. 4B. In some embodiments, the enrichable content received at the retrieval system may include one or more images and/or videos taken using a camera of a client device, such as

the client device 102, 302, and/or 424. A user of the client device 102, 302, and/or 424 may launch an application and/or a web interface to take the one or more images and/or video using the camera of the client device. In some embodiments, a user may transmit the enrichable content to the retrieval system via a gateway over a network.

5 [0247] At the operation 1004, object clusters may be generated in which one or more objects in the received enrichable content may be identified. For example, a predetermined area of the enrichable content may be analyzed for determining one or more objects. The predetermined area may be a central portion of an image. In some embodiments, the predetermined area may be selected based on blurring of an image. Accordingly, a picture area without blurring may be
10 selected for analysis to generate object clusters as described herein.

[0248] At the operation 1006, an identifier of a number of objects in object clusters generated based on the enrichable content received at the operation 1002 may be generated. In some
15 embodiments, one or more objects of the object clusters may be converted into a single column representing the one or more objects of the object clusters. In some embodiments, the generated vector may be normalized, as described herein in accordance with criteria including one or more of: a data size of a vector, a particular physical dimension of a vector, and so on. The generated vector may then be compared with one or more stored vectors in a database, for example, an index database, at the operation 806.

[0249] If one or more vectors stored in the index database are found to be matching above a
20 specific threshold, then it may be checked whether each of the one or more vectors stored in the index database correspond to the same enrichable content and/or content generated having different properties based on the received enrichable content. If it is determined that the one or more vectors stored in the index database and found matching with a vector generated at the operation 1006 all correspond to the same enrichable content and/or content generated having
25 different properties based on the received enrichable content, then an index corresponding to any of the one or more vectors may be identified for further processing at the operation 1008.

[0250] However, if it is determined that the one or more vectors stored in index database and found matching with a vector generated at the operation 1006 do not correspond to the same
30 enrichable content and/or content generated having different properties based on the received enrichable content, then an index corresponding to a vector having the best match may be identified for further processing at the operation 1008.

[0251] At the operation 1008, a set of actions associated with an index of the vector stored in index database, and as determined at the operation 1006 above, may be retrieved from an actions database. At the operation 1008, the retrieved set of actions may be analyzed, and it may be

determined that the enrichable content received at the operation 1002 corresponds with a QR code, an NFC tag, and/or a BLE. The user may be within an NFC and/or BLE range or may be outside of the NFC and/or BLE beacon range. Similarly, the QR code may be scannable or may be occluded, hidden, and/or damaged. In response to determining that the enrichable content received at 1002 above includes an occluded QR code and/or an out-of-range NFC tag and/or BLE beacon, a user may be notified of the occluded QR code and/or the out-of-range NFC tag and/or BLE beacon.

[0252] At the operation 1010, a user input may be received in which the user may have selected an option to execute an action associated with the occluded QR code and/or the out-of-range NFC tag and/or BLE beacon. Accordingly, a user can execute functions associated with an out-of-range NFC tag and/or BLE beacon even when the user is not within an NFC and/or BLE beacon range. Similarly, a user can execute functions associated with a QR code even when the QR code is unscannable, occluded, hidden, and/or damaged.

[0001] The foregoing examples and description of instances of purpose-configured software, whether accessible via API as a request-response service, an event-driven service, or whether configured as a self-contained data processing service are understood as not exhaustive. In other words, a person of skill in the art may appreciate that the various functions and operations of a system such as described herein can be implemented in a number of suitable ways, developed leveraging any number of suitable libraries, frameworks, first or third-party APIs, local or remote databases (whether relational, NoSQL, or other architectures, or a combination thereof), programming languages, software design techniques (e.g., procedural, asynchronous, event-driven, and so on or any combination thereof), and so on. The various functions described herein can be implemented in the same manner (as one example, leveraging a common language and/or design), or in different ways. In many embodiments, functions of a system described herein are implemented as discrete microservices, which may be containerized or executed/instantiated leveraging a discrete virtual machine, that are only responsive to authenticated API requests from other microservices of the same system. Similarly, each microservice may be configured to provide data output and receive data input across an encrypted data channel. In some cases, each microservice may be configured to store its own data in a dedicated encrypted database; in others, microservices can store encrypted data in a common database; whether such data is stored in tables shared by multiple microservices or whether microservices may leverage independent and separate tables/schemas can vary from embodiment to embodiment. As a result of these described and other equivalent architectures, it may be appreciated that a system such as described herein can be implemented in a number of suitable ways. For simplicity of description,

many embodiments that follow are described in reference an implementation in which discrete functions of the system are implemented as discrete microservices. It is appreciated that this is merely one possible implementation.

5 [0002] In addition, it is understood that organizations and/or entities responsible for the access, aggregation, validation, analysis, disclosure, transfer, storage, or other use of private data such as described herein will preferably comply with published and industry-established privacy, data, and network security policies and practices. For example, it is understood that data and/or information obtained from remote or local data sources, only on informed consent of the subject of that data and/or information, should be accessed aggregated only for legitimate, agreed-upon, 10 and reasonable uses.

[0003] As used herein, the phrase “at least one of” preceding a series of items, with the term “and” or “or” to separate any of the items, modifies the list as a whole, rather than each member of the list. The phrase “at least one of” does not require selection of at least one of each item listed; rather, the phrase allows a meaning that includes at a minimum one of any of the items, 15 and/or at a minimum one of any combination of the items, and/or at a minimum one of each of the items. By way of example, the phrases “at least one of A, B, and C” or “at least one of A, B, or C” each refer to only A, only B, or only C; any combination of A, B, and C; and/or one or more of each of A, B, and C. Similarly, it may be appreciated that an order of elements presented for a conjunctive or disjunctive list provided herein should not be construed as limiting the 20 disclosure to only that order provided.

[0004] One may appreciate that although many embodiments are disclosed above, that the operations and steps presented with respect to methods and techniques described herein are meant as exemplary and accordingly are not exhaustive. One may further appreciate that alternate step order or fewer or additional operations may be required or desired for particular 25 embodiments.

[0005] Although the disclosure above is described in terms of various exemplary embodiments and implementations, it should be understood that the various features, aspects and functionality described in one or more of the individual embodiments are not limited in their applicability to the particular embodiment with which they are described, but instead can be 30 applied, alone or in various combinations, to one or more of the some embodiments of the invention, whether or not such embodiments are described and whether or not such features are presented as being a part of a described embodiment. Thus, the breadth and scope of the present invention should not be limited by any of the above-described exemplary embodiments but is instead defined by the claims herein presented.

CLAIMS

What is claimed is:

1. A system for ingesting enrichable content and associating the enrichable content with an
5 action, the system comprising:
an application server comprising:
a memory allocation storing executable instructions; and
a processor allocation operably coupled with the memory allocation and
configured to load from the memory allocation the executable instructions thereby instantiating
10 an instance of a backend application configured to communicably couple to a frontend
application instance instantiated over a client device in network communication with the
application server, the backend application instance configured to:

receive a structured data object comprising a static image of the enrichable
content and an attribute identifying an action to associate to the enrichable content;
15 provide the static image as input to a high-accuracy object classifier;
receive as output from the high-accuracy object classifier, a set of objects
and corresponding bounding boxes identified within the static image;
generate an object graph based on the set of corresponding bounding
boxes;
20 generate a hash corresponding to object graph; and
storing the hash and the action in a database.
2. The system of claim 1, wherein the client device comprises a personal electronic device
comprising a camera.
25
3. The system of claim 1, wherein the structured data object comprises a set of static images
and the static image is a member of the set of static images.
4. The system of claim 3, wherein the set of static images are associated with frames of a
30 captured video.
5. The system of claim 3, wherein the backend application is configured to, for each
respective static image of the set of static images:

provide the respective static image as input to the high-accuracy object classifier;
receive as output from the high-accuracy object classifier, a respective set of objects and
corresponding bounding boxes identified within the respective static image;

5 generate at respective graph based on the respective set of corresponding bounding
boxes; and
generate a respective hash corresponding to object graph.

6. The system of claim 5, wherein the backend application is configured to generate a single
hash from each respective hash.

10 7. The system of claim 6, wherein the single hash is associated to the action in the database.

8. The system of claim 1, wherein the structured data object comprises location information
of the client device.

15 9. The system of claim 8, wherein the hash is based, at least in part, on the location
information.

20 10. The system of claim 1, wherein the structured data object comprises orientation
information of the client device.

11. The system of claim 10, wherein the hash is based, at least in part, on the orientation
information.

25 12. The system of claim 10, wherein the orientation information is based at least in part on
accelerometer data of the client device, compass data of the client device, or gyroscope
information of the client device.

30 13. The system of claim 1, wherein the structured data object comprises network connectivity
information of the client device.

14. The system of claim 13, wherein the hash is based, at least in part, on the network
connectivity information.

15. The system of claim 1, wherein the network connectivity information corresponds to a Wi-Fi connection, a cellular connection, or a Bluetooth connection.
16. The system of claim 1, wherein the action comprises an instruction to cause the client device to load a URL.
17. The system of claim 1, wherein the action comprises an instruction to cause the client device to render a virtual reality scene on a display of the client device.
18. The system of claim 1, wherein the action comprises an instruction to cause the client device to render an augmented reality scene on a display of the client device.
19. A system for identifying enrichable content and causing to be executed at least one action associated with the enrichable content, the system comprising:
- an application server comprising:
- a memory allocation storing executable instructions; and
 - a processor allocation operably coupled with the memory allocation and configured to load from the memory allocation the executable instructions thereby instantiating an instance of a backend application configured to communicably couple to a frontend application instance instantiated over a client device in network communication with the application server, the backend application instance configured to:
 - receive a structured data object comprising a static image;
 - provide the static image as input to a high-speed object classifier;
 - receive as output from the high-speed object classifier, a set of objects and corresponding bounding boxes identified within the static image;
 - generate an object graph based on the set of corresponding bounding boxes;
 - generate a hash corresponding to object graph;
 - determine whether the hash is equivalent to a hash previously stored in a database;
 - in response to determining that the hash is equivalent to a previously-stored hash, retrieve the previously-stored hash from the database;
 - provide the retrieved hash as an input query to an action database;
 - receive from the action database in response to the input query, an action associated with the retrieved hash;

provide the retrieved hash to an action distributor to cause the action to be executed by at least one of:

- a third-party system;
- the backend application; or
- 5 the client device.

20. The system of claim 19, wherein the static image is an image of a scene captured by the client device.

10 21. The system of claim 20, wherein the scene comprises an active television displaying a broadcast.

22. The system of claim 20, wherein the structured data object comprises at least one of:
location information of the client device; or
15 orientation information of the client device.

23. A method of operating a server application to ingest media content and associate the media content with one or more actions to be performed by at least one of a client device, a third-party service, or a first-party service, the method comprising:
20 receiving a structured data object comprising a static image of a scene and an attribute identifying an action to associate to the enrichable content;
provide the static image as input to a high-accuracy object classifier;
receive as output from the high-accuracy object classifier, a set of objects and corresponding bounding boxes identified within the static image;
25 generate an object graph based on the set of corresponding bounding boxes;
generate a hash corresponding to object graph; and
storing the hash and the action in a database.

30 24. The method of claim 23, wherein the static image comprises a QR code.

25. The method of claim 24, comprising extracting data encoded by of the QR coded.

26. The method of claim 25, comprising associating the action with the extracted data of the QR code.

27. The method of claim 23, wherein the structured data comprises information obtained from an NFC tag, RFID, or Bluetooth tag disposed within the scene.

5 28. The method of claim 29, comprising associating the action with the information.

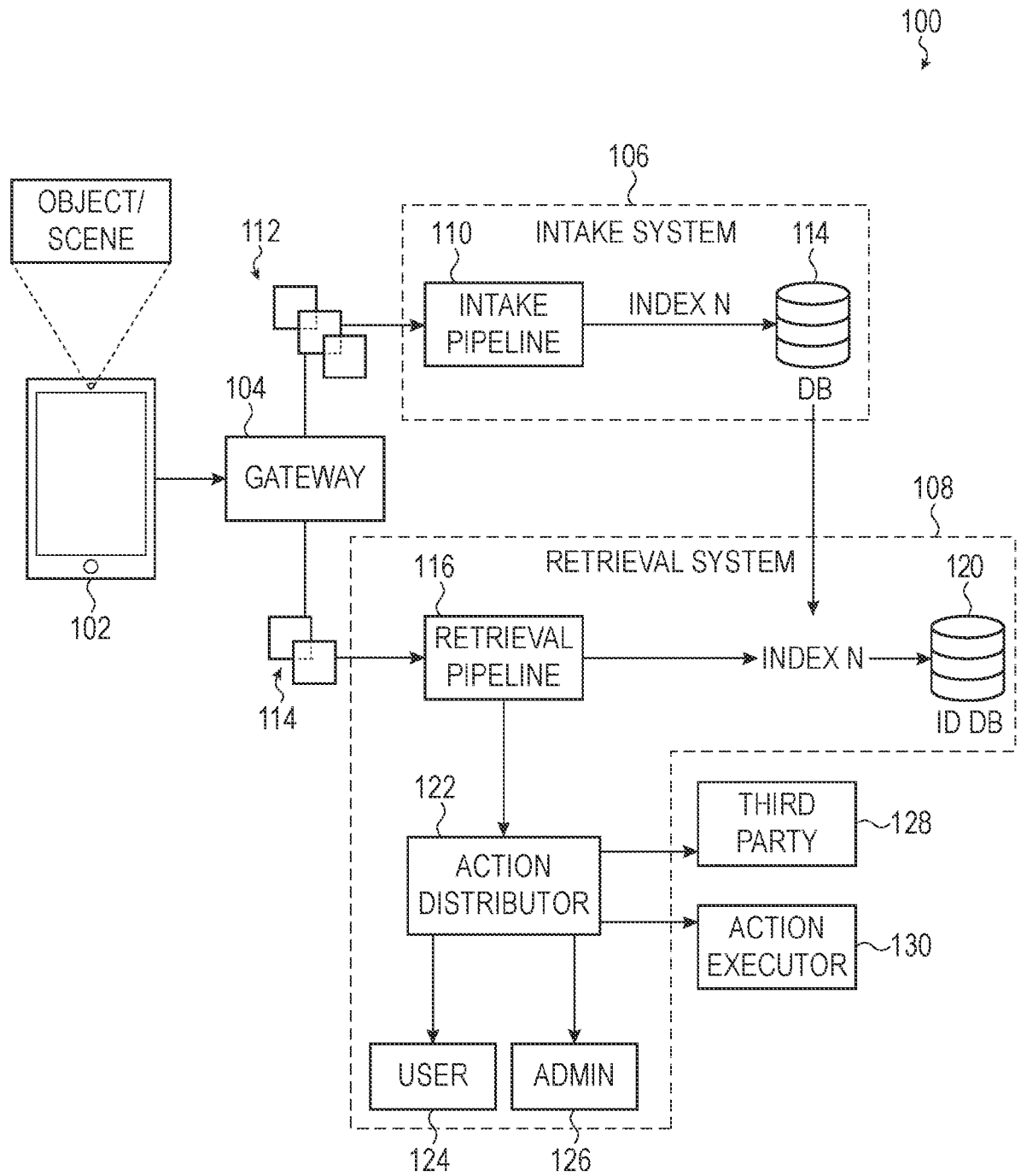


FIG. 1

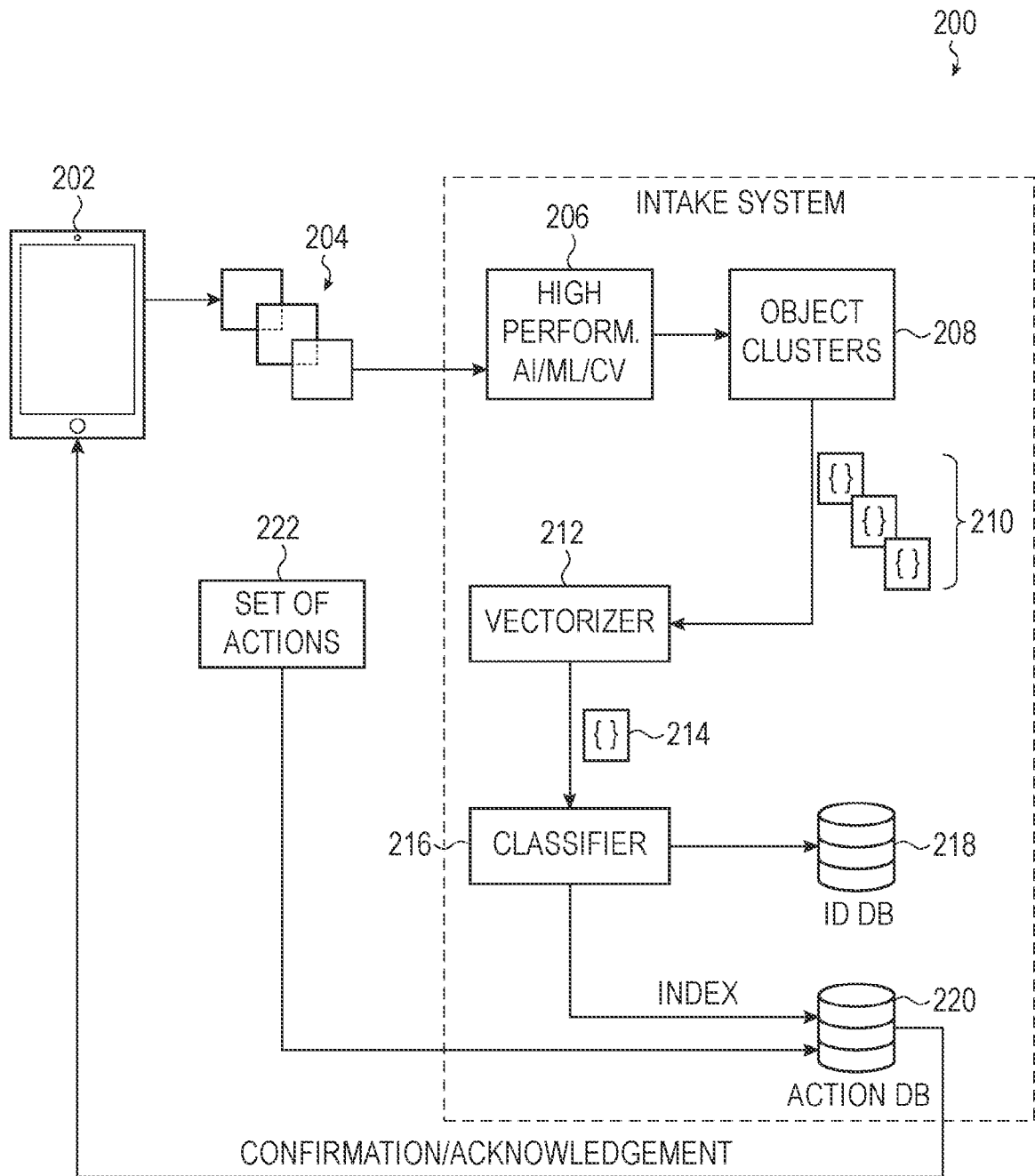


FIG. 2

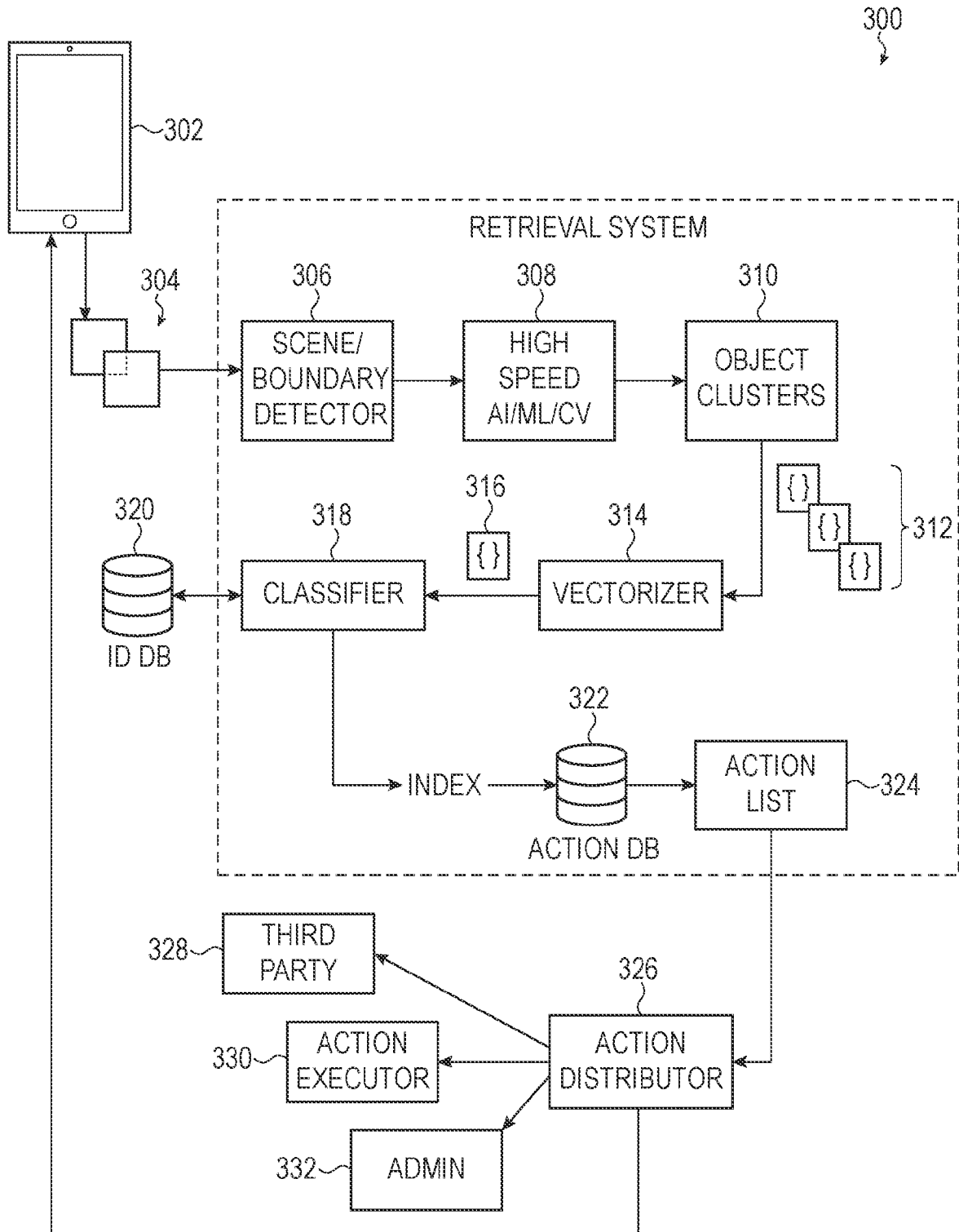


FIG. 3

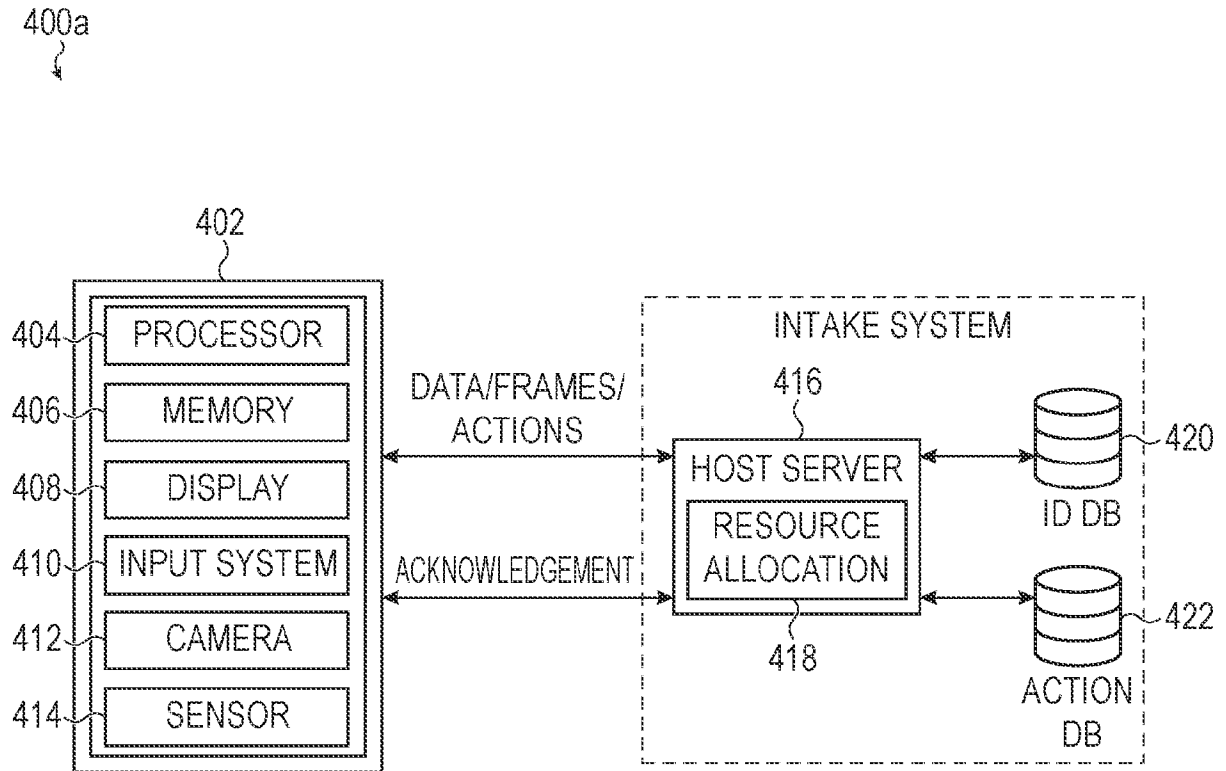


FIG. 4A

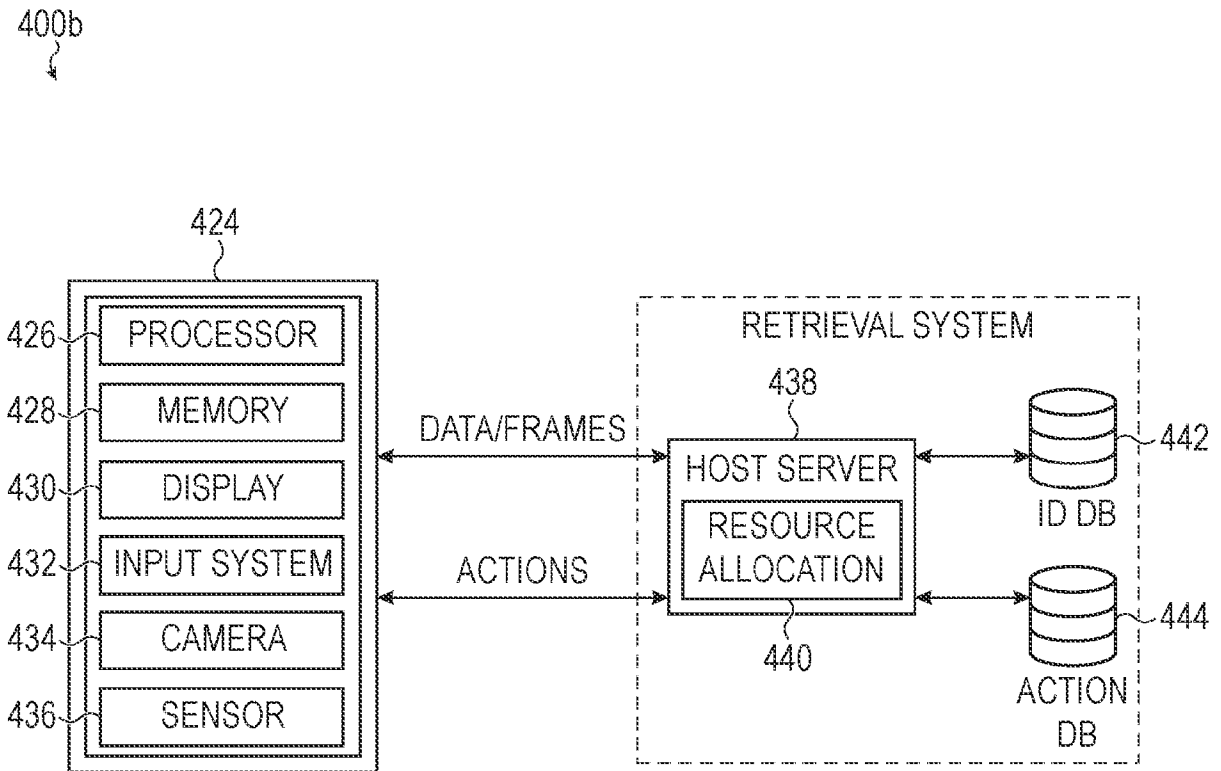


FIG. 4B

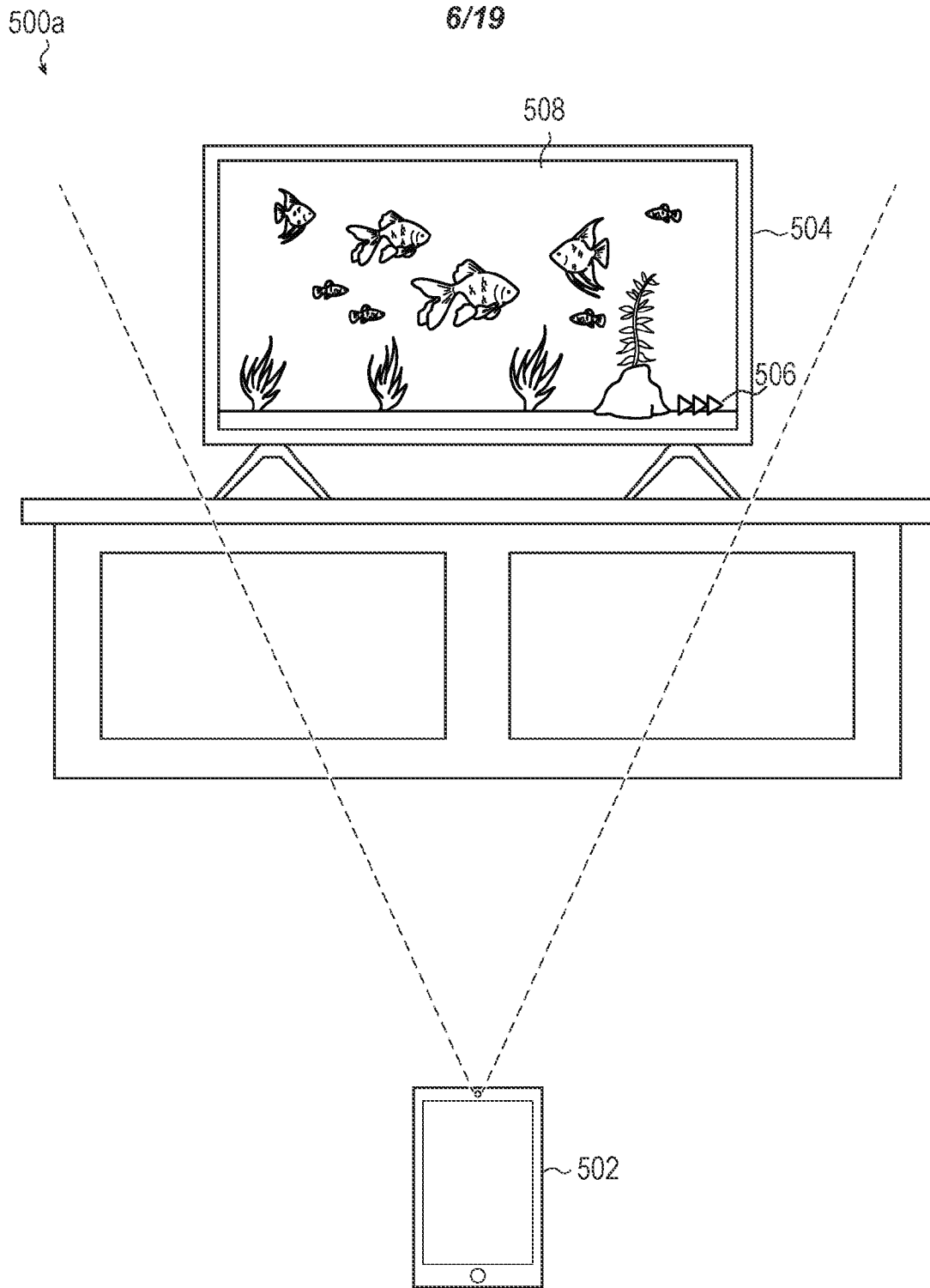


FIG. 5A

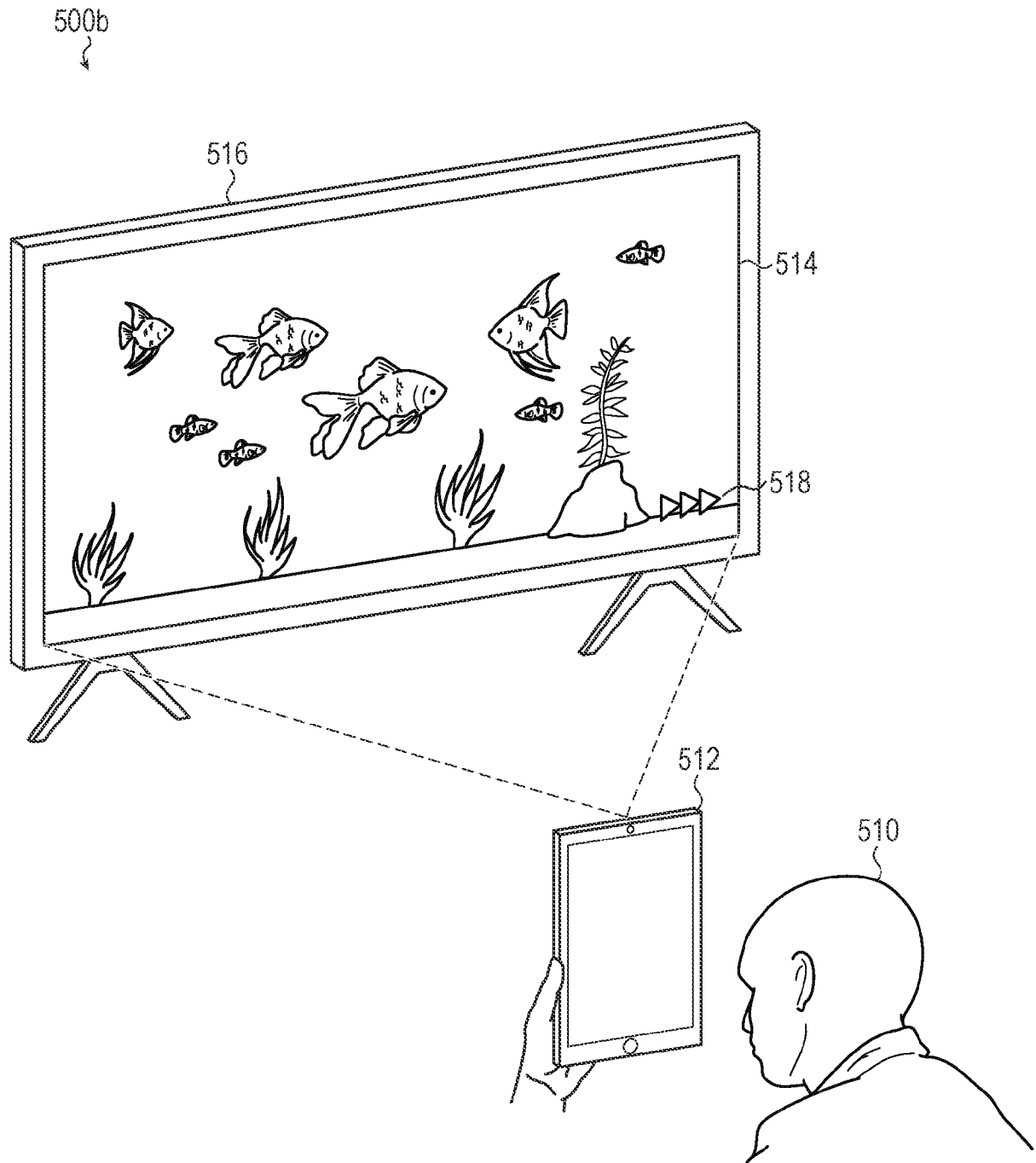


FIG. 5B

8/19

500c

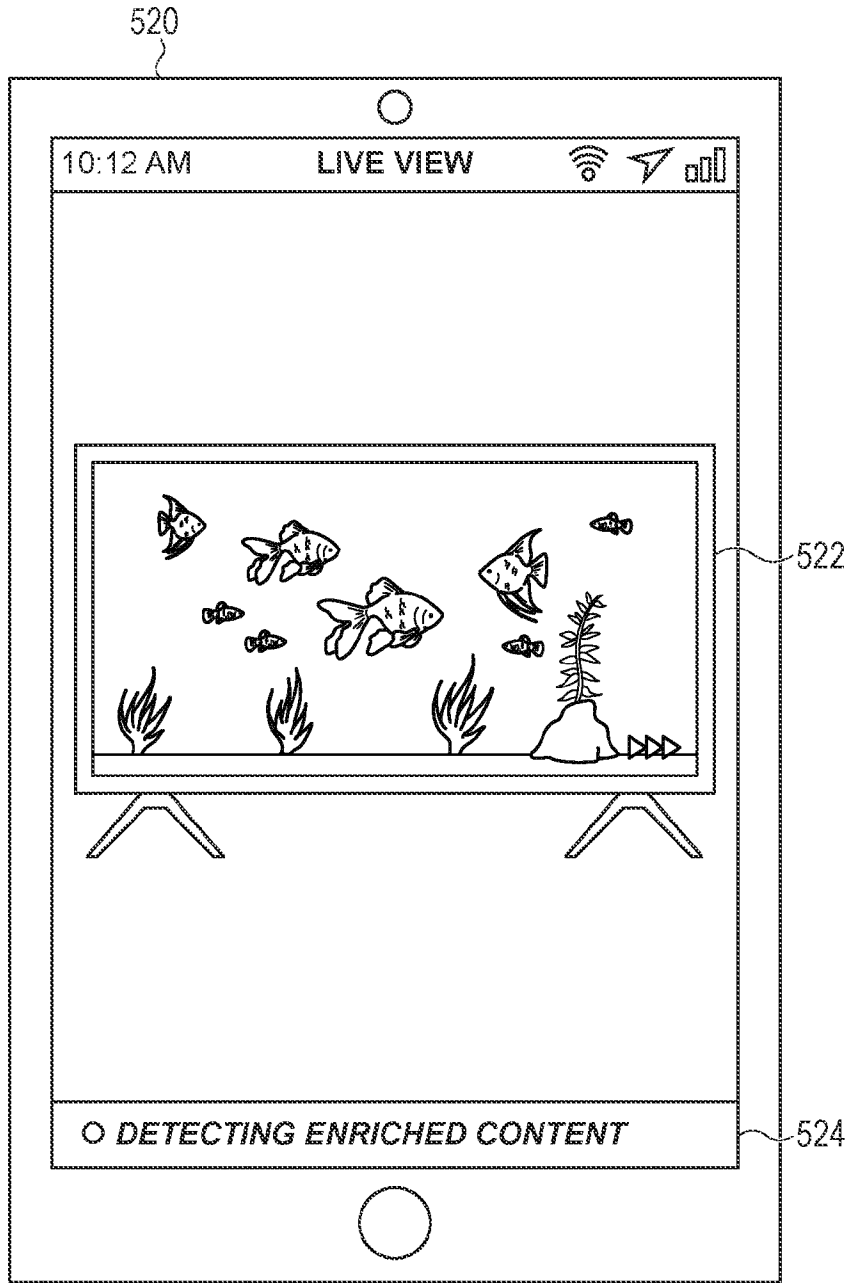


FIG. 5C

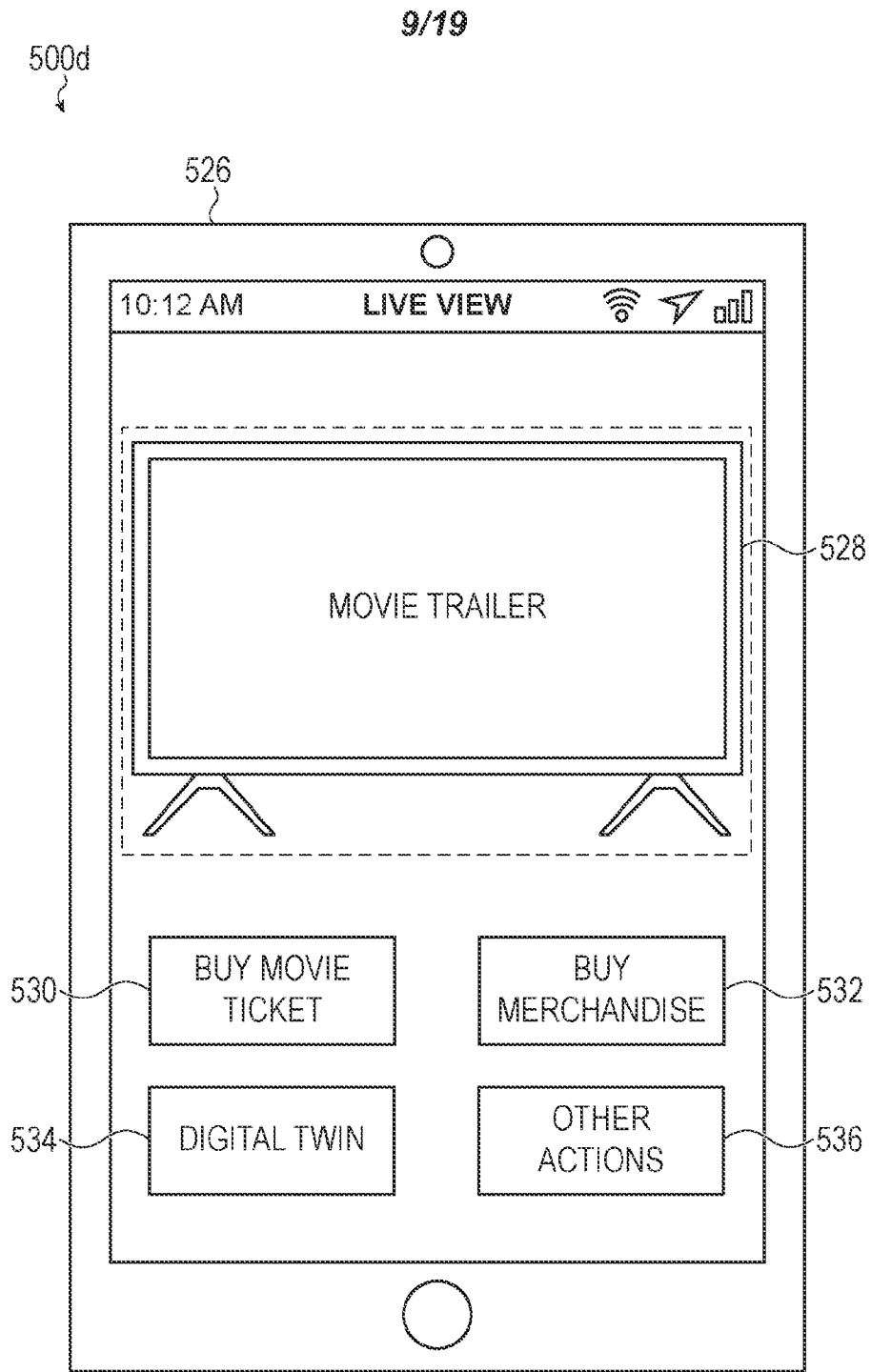


FIG. 5D

500e
↓

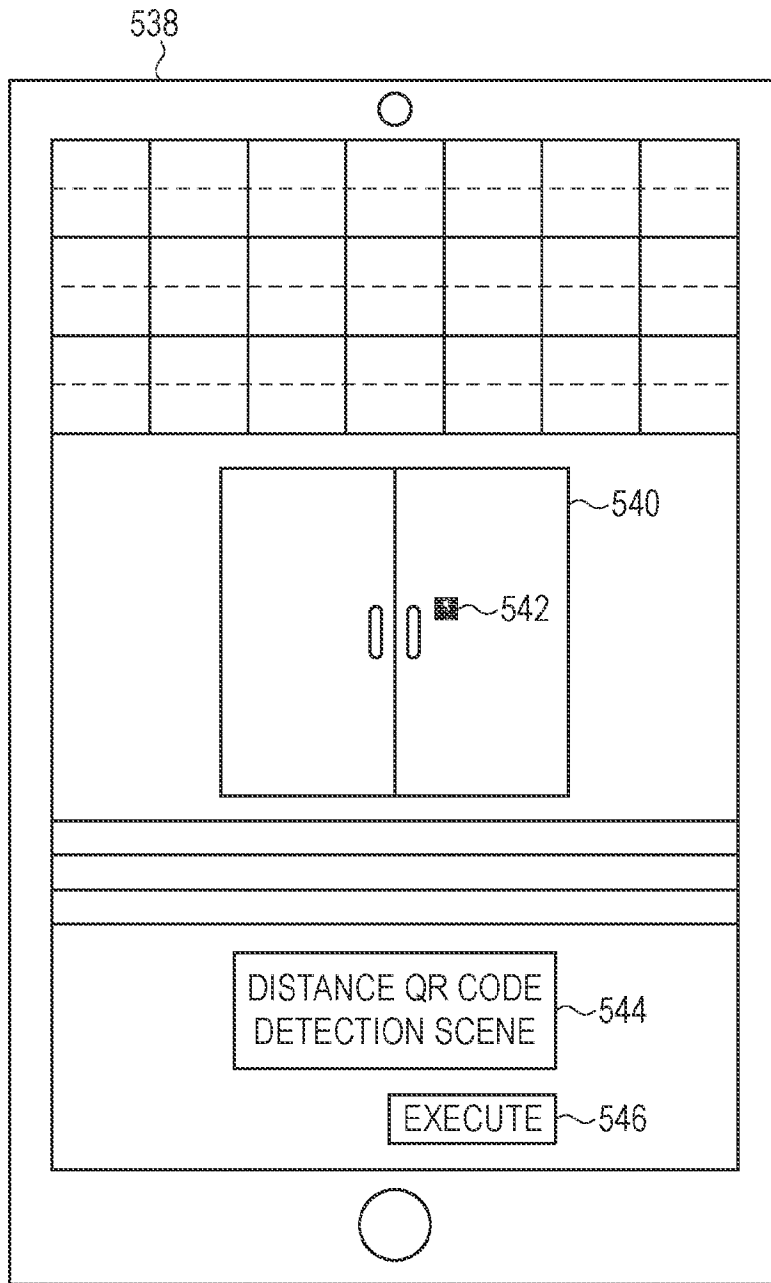


FIG. 5E

500f
↓

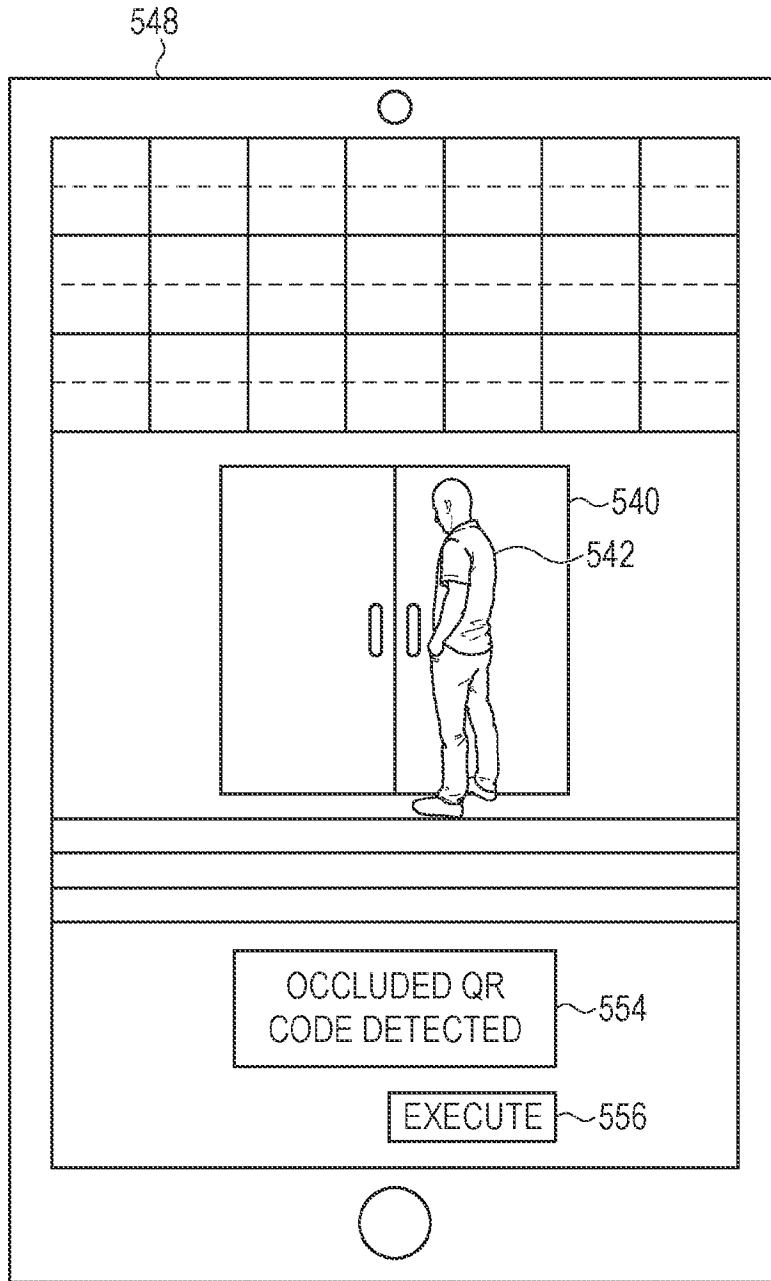


FIG. 5F

12/19

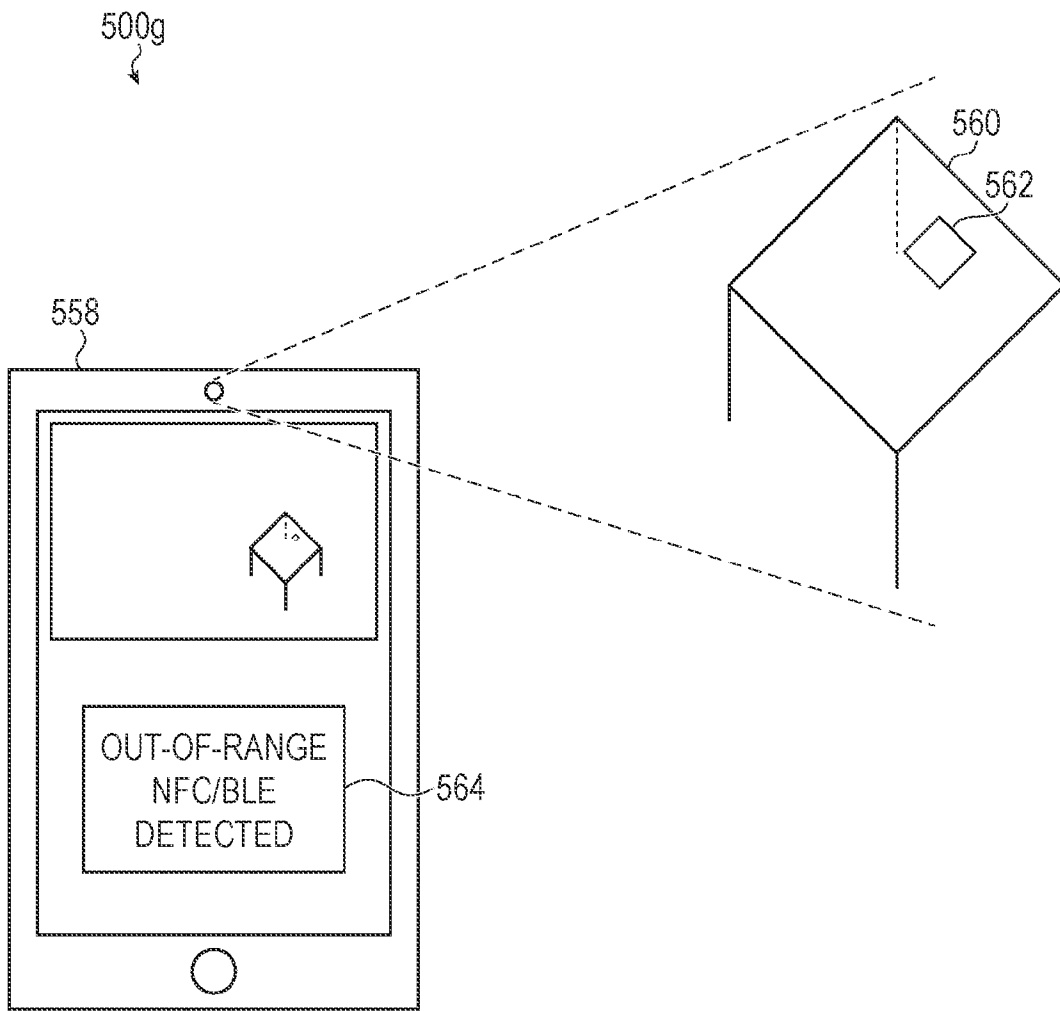


FIG. 5G

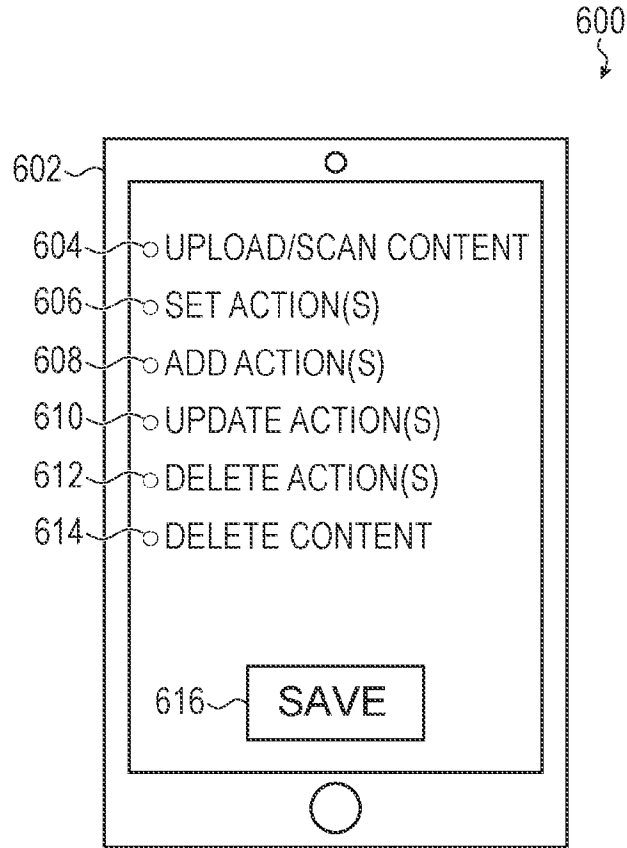


FIG. 6

700a
↓

14/19

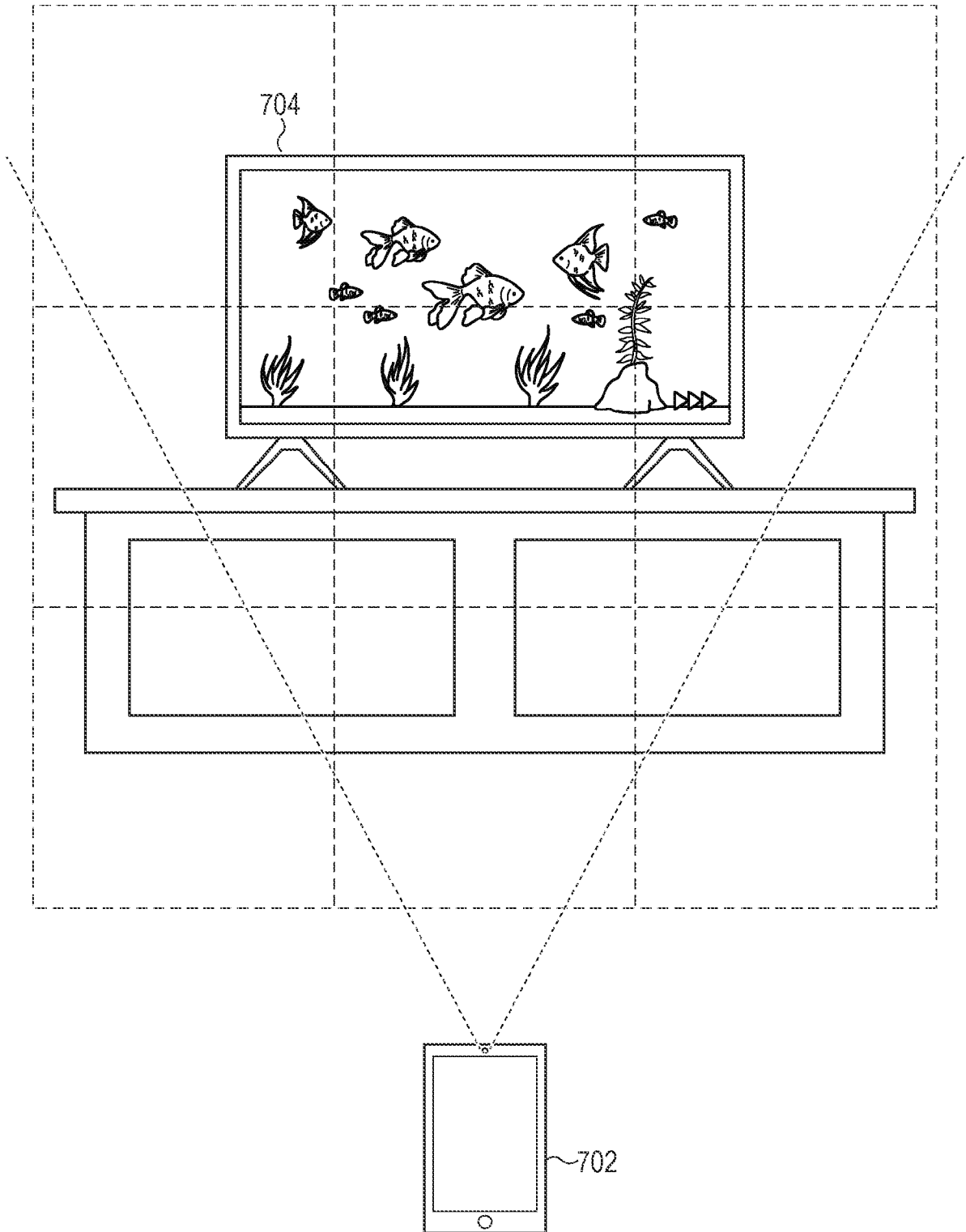


FIG. 7A

700B
↓

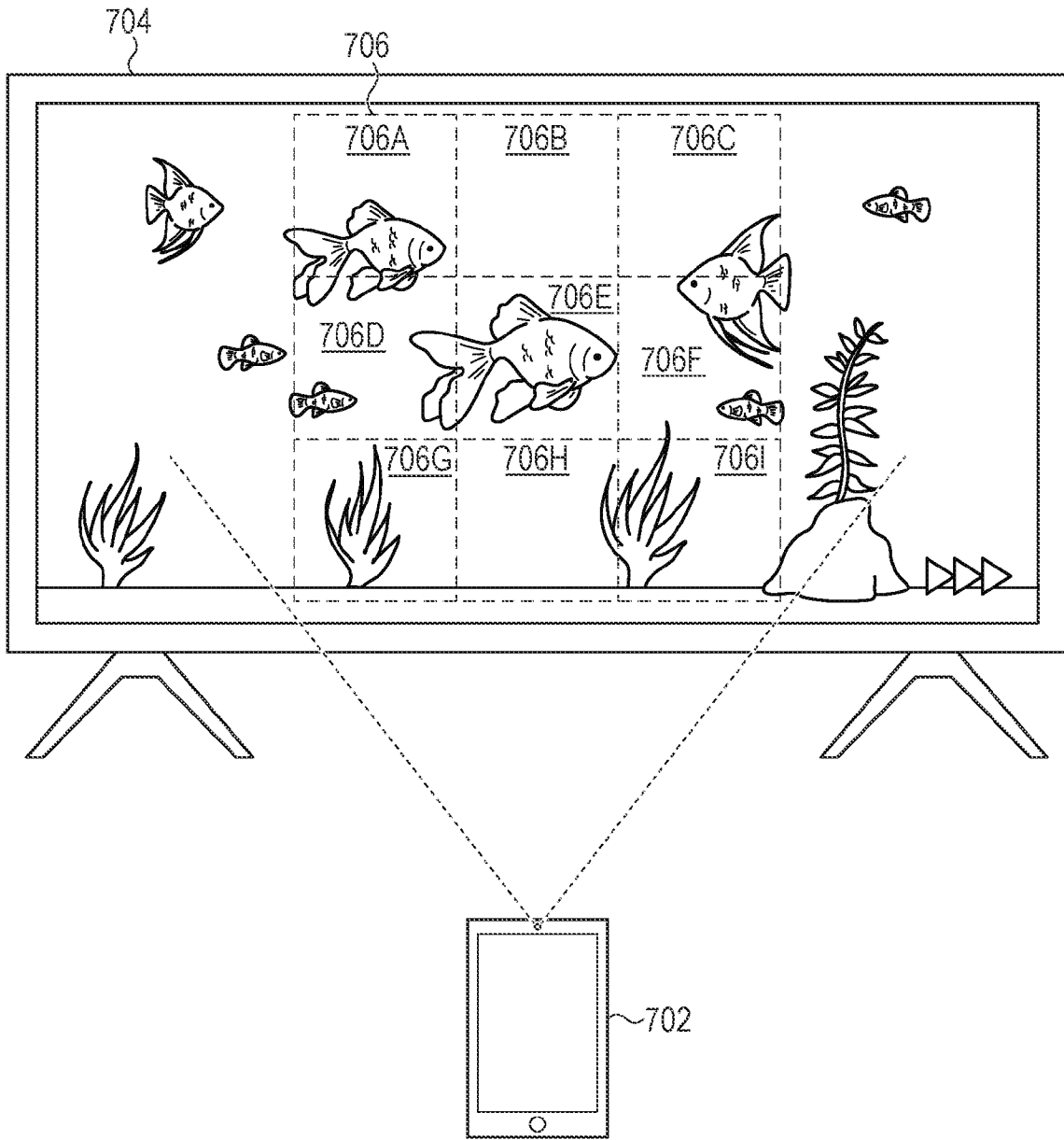


FIG. 7B

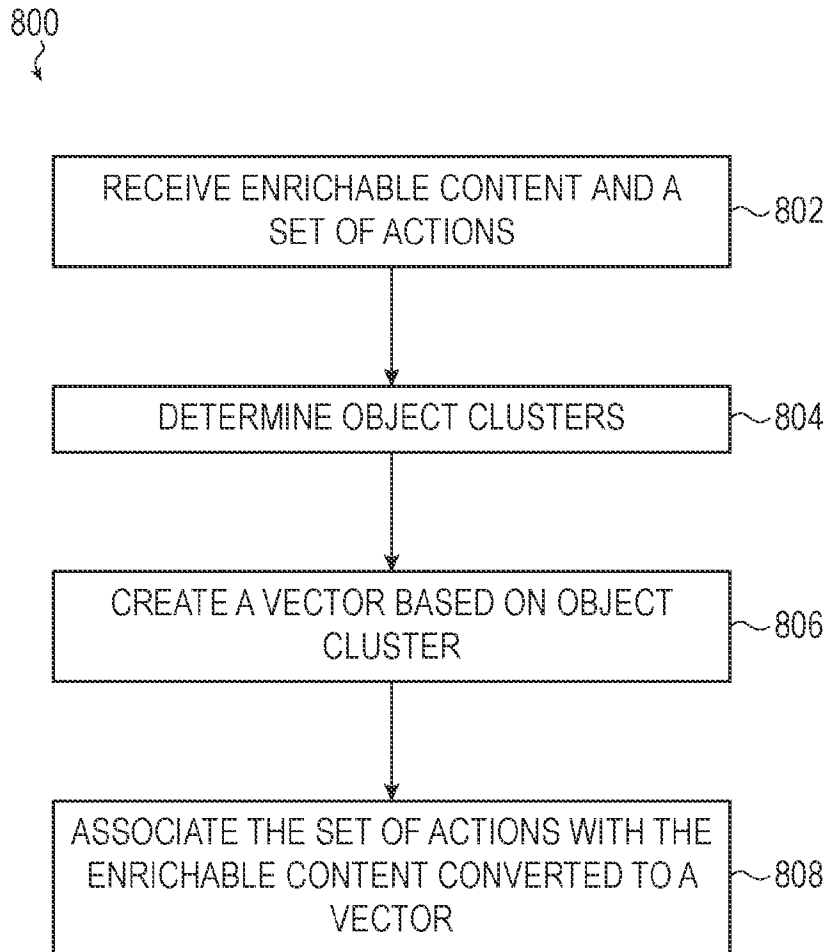


FIG. 8

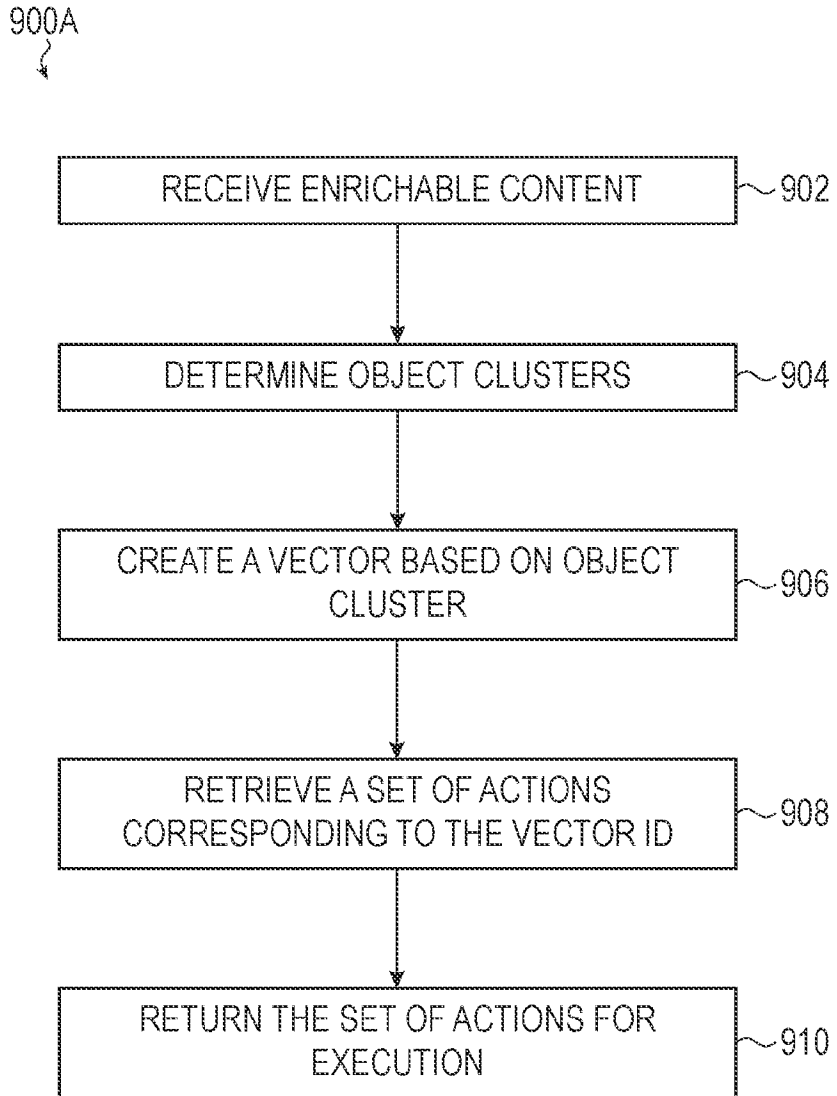


FIG. 9A

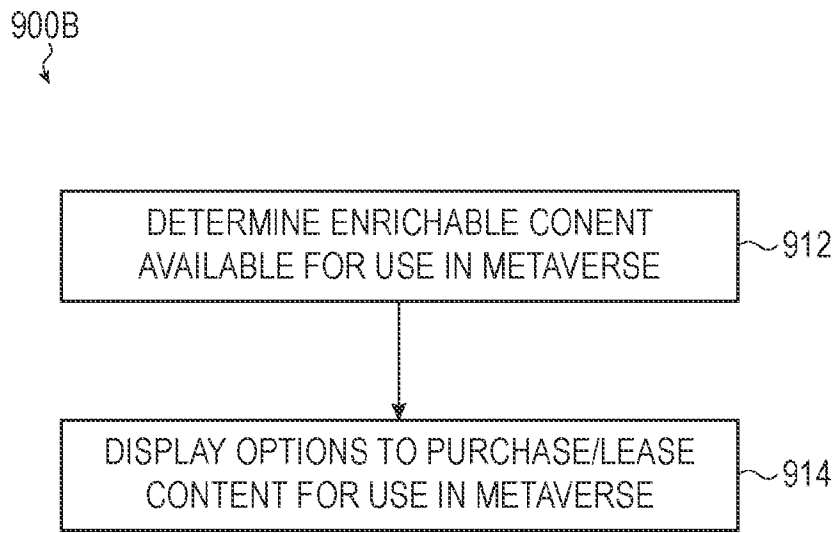
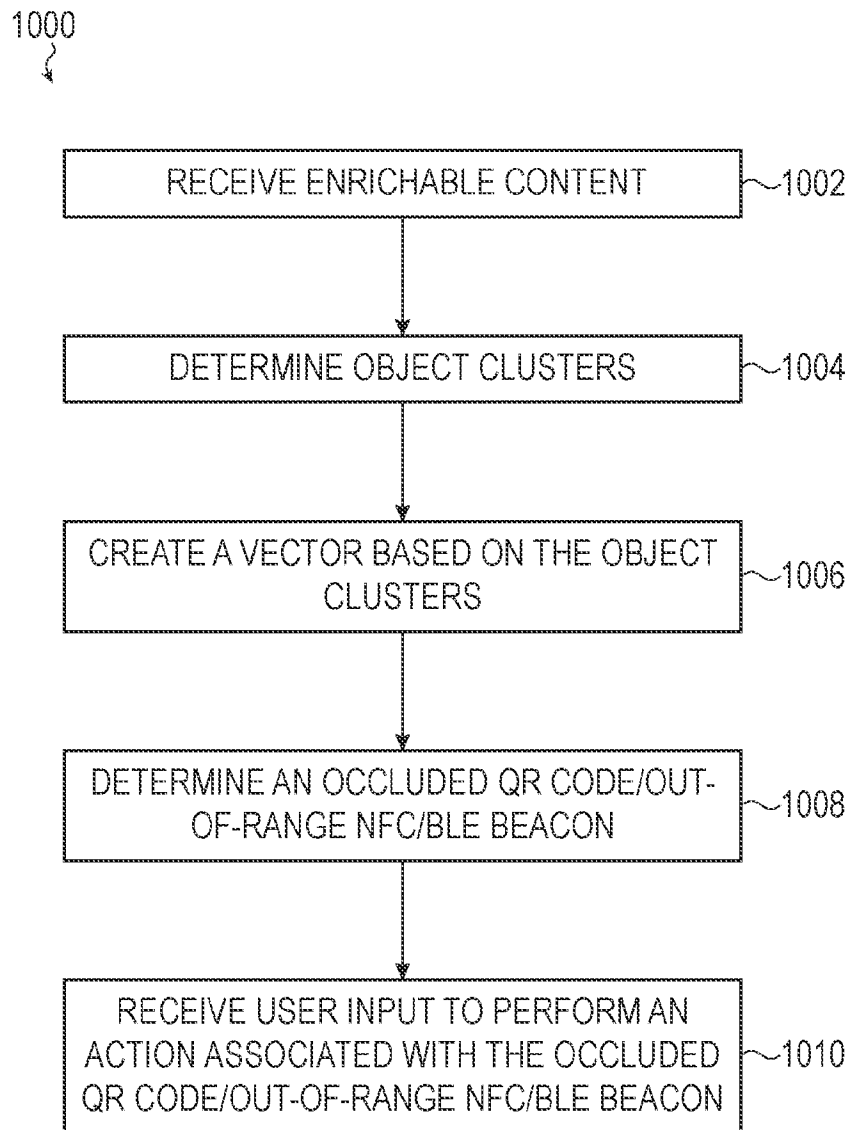


FIG. 9B

19/19

**FIG. 10**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2023/021728

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - INV. - G06F 3/0482; G01S 3/786; G06F 16/26; G06F 16/38 (2023.01)

ADD. - G06K 17/00; G06T 7/162; G06V 10/50 (2023.01)

CPC - INV. - G06F 3/0482; G01S 3/7865; G06F 16/26; G06F 16/381 (2023.05)

ADD. - G06K 17/0016; G06T 7/162; G06V 10/50 (2023.05)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic database consulted during the international search (name of database and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2016/0117061 A1 (MIWORLD TECHNOLOGIES INC.) 28 April 2016 (28.04.2016) entire document	1-28
A	US 2017/0286901 A1 (BOSSA NOVA ROBOTICS IP, INC.) 05 October 2017 (05.10.2017) entire document	1-28
A	LAI et al. "Instance-aware hashing for multi-label image retrieval." IEEE Transactions on Image Processing 25.6 (2016): 2469-2479. Retrieved on 29 June 2023 (29.06.2023) from < https://ieeexplore.ieee.org/abstract/document/7438833 > entire document	1-28
A	KUMAR et al. "Object detection in real time based on improved single shot multi-box detector algorithm." EURASIP Journal on Wireless Communications and Networking 2020.1 (2020): 1-18. Retrieved on 29 June 2023 (29.06.2023) from < https://jwcn-urasipjournals.springeropen.com/articles/10.1186/s13638-020-01826-x > entire document	1-28
A	US 9,002,831 B1 (O'MALLEY et al.) 07 April 2015 (07.04.2015) entire document	1-28
A	US 2014/0247278 A1 (SAMARA et al.) 04 September 2014 (04.09.2014) entire document	1-28

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 June 2023

Date of mailing of the international search report

AUG 10 2023

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Taina Matos

Telephone No. PCT Helpdesk: 571-272-4300