(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 May 2010 (27.05.2010)

PCT

(10) International Publication Number
WO 2010/059188 A2

(51) International Patent Classification:
G06F 17/30 (2006.01) *G06K 9/46* (2006.01)

(21) International Application Number:
PCT/US2009/006064

(22) International Filing Date:
12 November 2009 (12.11.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/273,600 19 November 2008 (19.11.2008) US

(71) Applicant *(for all designated States except US)*: **EASTMAN KODAK COMPANY** [US/US]; 343 State Street, Rochester, New York 14650-2201 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DAS, Madirakshi** [IN/US]; 343 State Street, Rochester, New York 14650 (US). **LOUI, Alexander C.** [US/US]; 343 State Street, Rochester, New York 14650 (US). **WOOD, Mark D.** [US/US]; 343 State Street, Rochester, New York 14650 (US).

(74) Common Representative: **EASTMAN KODAK COMPANY**; 343 State Street, Rochester, New York 14650-2201 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: METHOD FOR EVENT-BASED SEMANTIC CLASSIFICATION

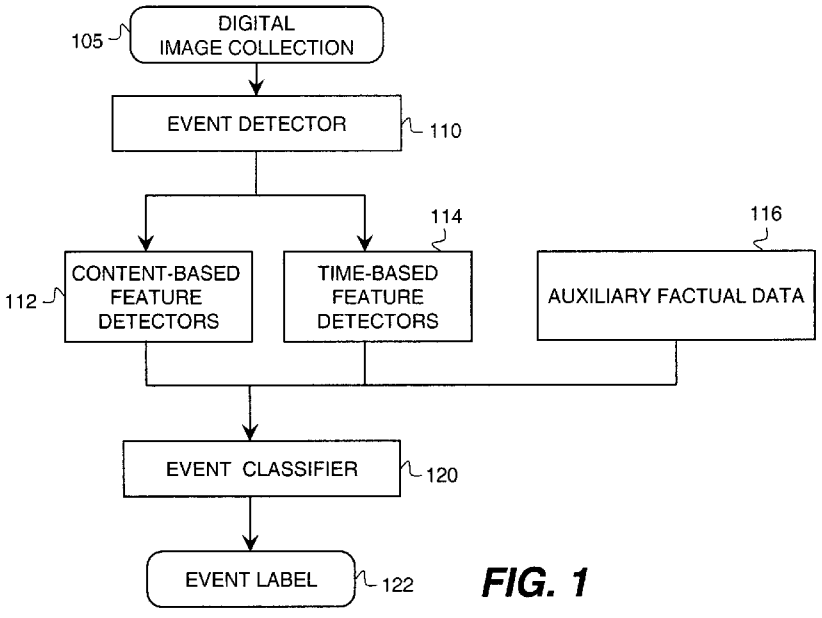

FIG. 1

(57) **Abstract**: A method of automatically classifying images in a consumer digital image collection, includes generating an event representation of the image collection; computing global time-based features for each event within the hierarchical event representation; computing content-based features for each image in an event within the hierarchical event representation; combining content-based features for each image in an event to generate event-level content-based features; and using time-based features and content-based features for each event to classify an event into one of a pre-determined set of semantic categories.

WO 2010/059188 A2

METHOD FOR EVENT-BASED SEMANTIC CLASSIFICATION

FIELD OF THE INVENTION

The invention relates generally to the field of digital image processing, and in particular to a method for classifying digital images into semantic categories.

BACKGROUND OF THE INVENTION

The proliferation of digital cameras and scanners has lead to an explosion of digital images, creating large personal image databases. The organization and retrieval of images and videos is already a problem for the typical consumer. Currently, the length of time spanned by a typical consumer's digital image collection is only a few years. The organization and retrieval problem will continue to grow as the length of time spanned by the average digital image and video collection increases, and automated tools for efficient image indexing and retrieval will be required.

Many methods of image classification based on low-level features such as color and texture have been proposed for use in content-based image retrieval. A survey of low-level content-based techniques ("Content-based Image Retrieval at the End of the Early Years," A.W.M. Smeulders et al, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), Dec 2000) provides a comprehensive listing of relevant methods that can be used for content-based image retrieval. The low-level features commonly described include color, local shape characteristics derived from directional color derivatives and scale space representations, image texture, image transform coefficients such as the cosine transform used in JPEG-coding and properties derived from image segmentation such as shape, contour and geometric invariants. Though these features can be efficiently computed and matched reliably, they usually have poor correlation with semantic image content.

There have also been attempts to compute semantic-level features from images. In WO 01/37131 A2, visual properties of salient image regions are used to classify images. In addition to numerical measurements of visual

properties, neural networks are used to classify some of the regions using semantic terms such as "sky" and "skin." The region-based characteristics of the images in the collection are indexed to make it easy to find other images matching the characteristics of a given query image. U.S. Patent No. 6,240,424 B1, discloses a method for classifying and querying images using primary objects in the image as a clustering center. Images matching a given unclassified image are found by formulating an appropriate query based on the primary objects in the given image. U.S. Patent Application Publication No. 2003/0195883 A1 computes an image's category from a pre-defined set of possible categories, such as "cityscapes."

These semantic-level features are also not the way users recall and search for images in their collection. Users' recollection of photographs is often based on the event that was captured. For example, photographs may be identified as "Grand Canyon vacation," "Mom's birthday party," "Joe's baseball league" and so on. There are mechanisms available in current software to manually enter such tags or captions to identify photographs. However, a need exists to automate this labor-intensive process, so that a user is able to search by common types of events without having to tag the images first. Further, the user can combine event type with other semantic features such as people present in the image, location or activity to narrow the search to relevant images.

SUMMARY OF THE INVENTION

It is an object of the present invention to classify images or videos in a digital image collection into one of several event categories. This object is achieved by a method of automatically classifying images in a consumer digital image collection, comprising:

(a) generating an event representation of the image collection;

(b) computing global time-based features for each event within the hierarchical event representation;

(c) computing content-based features for each image in an event within the hierarchical event representation;

(d) combining content-based features for each image in an

event to generate event-level content-based features; and

(e) using time-based features and content-based features for each event to classify an event into one of a pre-determined set of semantic categories.

The organization and retrieval of images and videos is a problem for the typical consumer. Automated tools are needed that can understand the content of the images and provide the ability to search the collection using semantic concepts such as events, people and places. This invention provides automatic classification of images in a collection into semantic event categories. This will permit the consumer to search for and browse images in the collection depicting specific events. The images and videos are automatically labeled with event category labels that can enable the automated generation of event-specific creative media outputs. This invention has the advantage of permitting users to search for images or videos in the collection that are part of specific event categories. Further, the invention has the advantage that images are automatically labeled with event category labels that can enable the automated generation of event-specific creative media outputs.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a flowchart of the method of the present invention;

FIG. 2 is a flowchart of the method for creating a hierarchical event representation from an image collection;

FIG. 3 illustrates the Bayesian net for event classification using time-based and content-based features;

FIG. 4 illustrates the Bayesian net for event classification with a broader set of features;

FIG. 5 illustrates the rule-based modification of Bayesian event classifier output;

FIG. 6 is a flowchart for training the Bayesian net for event classification; and

FIG. 7 illustrates an exploded view of the event classifier 120 of FIG. 1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention can be implemented in computer systems as will be well known to those skilled in the art. The invention has been described in detail with particular reference to certain preferred embodiments thereof, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention. Videos in a collection are treated as groups of keyframe images and included in the term "images" in the rest of the description.

Referring to FIG. 1, a user's digital image collection 105 is grouped into an event representation by the event detector 110. FIG. 2 shows the steps in creating a hierarchical event representation that has three-levels: super-events, events and sub-events. Referring to FIG. 2, the digital image collection 105 is first grouped into temporal events and sub-events using event and sub-event detector 225. Commonly assigned U.S. Patent No. 6,606,411 and U.S. Patent No. 6,351,556 disclose algorithms for clustering image content by temporal events and sub-events, the disclosures of which are incorporated herein. According to U.S. Patent No. 6,606,411 events have consistent color distributions, and therefore, these pictures are likely to have been taken with the same backdrop. For each sub-event, a single color and texture representation is computed for all background areas taken together. The above two patents teach how to cluster images and videos in a digital image collection into temporal events and sub-events. The terms "event" and "sub-event" are used in an objective sense to indicate the products of a computer mediated procedure that attempts to match a user's subjective perceptions of specific occurrences (corresponding to events) and divisions of those occurrences (corresponding to sub-events). Briefly summarized, a collection of images is classified into one or more events determining one or more largest time differences of the collection of images based on time and/or date clustering of the images and separating the plurality of images into the events based on having one or more boundaries between events where one or more boundaries correspond to the one or more largest time differences. For

each event, sub-events can be determined (if any) by comparing the color histogram information of successive images as described in U.S. Patent No. 6,351,556. This is accomplished by dividing an image into a number of blocks and then computing the color histogram for each of the blocks. A block-based histogram correlation procedure is used as described in U.S. Patent No. 6,351,556 to detect sub-event boundaries. Another method of automatically organizing images into events is disclosed in commonly assigned U.S. Patent No. 6,915,011, , which is herein incorporated by reference. Briefly summarized, according to one aspect of the above invention, an event clustering method uses foreground and background segmentation for clustering images from a group into similar events. Initially, each image is divided into a plurality of blocks, thereby providing block-based images. Utilizing a block-by-block comparison, each block-based image is segmented into a plurality of regions comprising at least a foreground and a background. One or more luminosity, color, position or size features are extracted from the regions and the extracted features are used to estimate and compare the similarity of the regions including the foreground and background in successive images in the group. Then, a measure of the total similarity between successive images is computed, thereby providing image distance between successive images, and event clusters are delimited from the image distances.

The events detected continue to be chronologically ordered in a timeline from earliest to latest. Using the method described above, it is not possible to detect single events that span a long period of time (days) and encompass a variety of activities and settings (for example, a long vacation covering multiple destinations) or events that occur in distinct parts separated by some hours from each other (for example, a sporting event with many matches or a wedding). Gaps in photo-taking corresponding to the overnight period also cause breaks in event continuity. Further processing is needed to detect these super-events, defined as a grouping of multiple contiguous events that may span multiple days.

Referring to FIG. 2, inter-event durations are computed for events

detected by the inter-event duration calculator 230. The inter-event duration is the time duration between the last image of one event and the first image of the next event on a continuous timeline. The events are then treated as single points on a time axis, separated by the inter-event durations. A density-based clustering method is applied to these points (ref. Data Mining Concepts and Techniques by Han and Kamber, Elsevier, 2006, pg. 418-420) by the density-based classifier 235. This algorithm grows regions with sufficiently high point density into clusters. This method enables points separated by relatively small duration gaps (for example, less than 18 hours) to be clustered together to form super-events. These typically represent longer duration or multi-part events in the user's collection. Smaller and isolated events appear as noise in this clustering method, and are left un-clustered. The final three-level hierarchical event representation is generated by the event hierarchy generator 240 and includes super-events, events and sub-events. After this point, the term "event" refers to the top-level of the hierarchical event representation – which can be a super-event or an event.

Referring to FIG. 1, algorithms for detecting content-based features are applied to each image in the collection using the content-based feature detectors 112. Content-based features refer to features that are computed from image pixels and that are intended to be an indicator of image content. This includes low-level features such as color and texture, and higher level features such as detection of faces or other objects. Commonly assigned U.S. Patent No.6504951 describes a method for detecting sky in an image. Classifying a scene as indoor or outdoor using camera metadata and image content is described in commonly assigned U.S. Patent Application Publication No. 2005/0105776. Scene classification (by way of example only, includes urban, field or nature scene) has been described in "Review of the state of the art in semantic scene classification," Computer Science TR 799, University of Rochester, 2002. Detection of a number of semantic classes (such as water, grass, snow, mountain, or buildings by way of example only) using supervised learning has been described in "Supervised learning of semantic classes for image annotation and

retrieval," by Carneiro et al, IEEE Transactions on PAMI, March 2007. Using the temporal context of events during the computation of the image-level feature detection process is described in commonly assigned U.S. Patent Application Publication No. 2005/0105775. A method for describing images that can contain multiple valid classes is described in "Learning multi-label scene classification" by Boutell et al in Pattern Recognition 37(2004). In addition, methods for detecting human faces are well known in the art of digital image processing, and can be used for obtaining a count of the number of faces in the image, approximate sizes of faces and locations of faces. For example, a face detection method for finding human faces in images is described in the following article: Jones, M.J.; Viola, P., "Fast Multi-view Face Detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2003. A method for assigning a face to an age category is described in U.S. Patent No. 5,781,650. The method uses ratios of facial measurements that are known to change with age due to bone structure development.

Table 1 has a list of example features that are collected from available algorithms. These features are found to be strongly correlated with the semantic event category. The first column of the table shows the name of the feature. The second column indicates the levels the feature can have. For example, in the simplest case, a feature can have only two levels – present and not present. The feature detection output can also be quantized into multiple levels to indicate either extent or degree of confidence of the feature. For some features, an experimentally determined threshold is used to test presence e.g. water is present if more than 25% of the image pixels are detected to be water (as shown in Table 1). This ensures that the feature is a significant part of the image. These features are detected at the image level. Image-level feature detection results are combined to obtain event-level content-based features. If a sufficient percentage of images in an event are tagged with the (non-zero) feature level, the event is tagged with that feature level. This threshold percentage is indicated in the third column of

Table 1. For example, 15% of the images in an event would need to have a "Present" level for the feature "Grass" for the event to have the level "Present" for the feature "Grass."

Referring to FIG. 1, global time-based features are also detected for each of the events in the collection using time-based feature detectors 114. Time-based features refer to features computed from the capture date/time information obtained from the EXIF header of images or from similar metadata from other file formats. Global time-based features refer to time-based features that are computed at the event level from the capture date/time information of images included within the event. Table 2 shows some examples of time-based features. These include the duration of the event, duration between adjacent events (inter-event duration), number of images, time of day (e.g. morning, afternoon, night) and time of year. The duration of the event is computed as the difference between the capture time (as embedded in the EXIF header of the image by the camera at capture) of the first and last image of the event; the inter-event duration is the difference between the capture time of the last image of the event and the first image of the next event. The time of day is determined by the image at the center of the event timeline. The year is divided into the main seasons and is determined based on the month of capture of the event. The time-based features are detected at the event level. The third column of Table 2 shows the threshold used to determine the level for each of the features.

Semantic event categories are pre-determined based on studies of typical consumer picture-taking occasions and cover 80-90% of all images captured. Table 3 shows four top-level semantic event categories and their sub-categories. The features computed by 112 and 114 of FIG. 1 are provided as inputs to an event classifier 120 that produces an event label(s) 122 as the output. A number of classifiers have been described in the literature that can be trained to produce a finite number of output classifications from a set of input features. These include support vector machines, neural networks and Bayesian networks (ref. Pattern Classification by Duda, Hart and Stork, 2nd ed., John Wiley and sons,

2000). Referring to FIG. 7, the event classifier 120 of FIG. 1 includes two stages. The preferred embodiment uses a Bayesian Belief Network (BBN) (ref. Probabilistic Reasoning in Intelligent Systems by Judea Pearl, Morgan Kaufmann, 1988) as a probabilistic classifier 710. Other types of classifiers, including support vector machines and neural networks, can also be used. BBN is a form of probabilistic graphical modeling that enables modeling and reasoning about uncertainty. It is a directed acyclic graph which includes nodes that represent variables and arcs that represent conditional probabilities capturing probabilistic dependencies between variables. During the training phase, the conditional probabilities corresponding to each arc of the Bayesian belief network and the *a priori* probabilities of the output labels are learned. After training, the output node then produces the probability of each of the output labels given the input observations. A key feature of BBN is that it is capable of handling incomplete information gracefully.

The Bayesian Belief Network is set up as shown in FIG. 3. The input nodes include the time-based features detected by the time-based feature detectors 114 including but not limited to those listed in Table 2, such as duration 310; and the content-based features detected by the content-based feature detectors 112, including but not limited to those listed in Table 1, such as people present 315 and % Outdoors 320. An output node 305 computes the event labels based on the time-based and content-based image features.

The accuracy of the classifier may be further improved by consideration of auxiliary factual data (information) directly or indirectly associated with the captured event, illustrated as box 116 in FIG. 1. Examples of such factual information might include, but is not limited to, knowledge of familial relationships between people, personal calendars, geographic information, and data considered common sense knowledge. Such factual information also includes the non-content based capture metadata typically recorded in common image file formats such as EXIF, including the subject

distance and the camera scene mode. Each of these types of information may be used to refine or validate the classification provided by the event classifier, and are included to create an extended feature set.

Such factual information can be applied by the system as part of the BBN. FIG. 4 illustrates how these types of data can factor into the classification. For example, training data can be collected that indicates that a greater subject distance tends to be associated with sporting events and vacations than family moments and social gatherings, helping to further improve the accuracy of the BBN. In some cases camera metadata can be combined with auxiliary knowledgebases to produce other input values for the BBN. Digital capture devices are increasingly capturing GPS coordinates at the time of capture; geolocation data can also be manually added to an image. Given the latitude and longitude of where an event occurred, such coordinates can be mapped to a geographic feature class, using an auxiliary database such as the U.S. Geological Survey's Board of Geographic Names Geographic Names Information System (GNIS). Feature classes include bodies of water as well as facilities such as schools. It is expected that a sporting event is less likely to take place at a cemetery than a school; a social gathering is more likely to take place at a church than an airport. Referring to FIG. 4, node 405 computes event labels based on the extended feature set, including but not limited to a subject distance 410 and the GNIS Feature Class 415, along with the output of node 305 of FIG. 3 that factors in the effect of time-based and content-based features described earlier. The event is then labeled with the event class that has the highest probability.

The BBN is trained on labeled input/output data to calculate the conditional probabilities at each arc of the network. The *a priori* probabilities (or priors) are also learned from the labeled training data. Referring to FIG. 6, a digital image training collection 602 spanning multiple years is collected from a number of consumers. Events and super-events are detected in the collections using the hierarchical event detector 110 from FIG. 1. The events are then labeled by a human manual event labeler 607 into one of the pre-determined set of event

categories listed in Table 3 (or "Other" if none are appropriate). The features described earlier are computed for each event using correlated feature detectors 610, which includes the content-based feature detectors 112 and the time-based feature detectors 114 of FIG. 1. This forms the training set for the Bayesian belief network. The BBN is trained for the top-level event category (in one embodiment, vacation, sports, family moments and social gathering) using an event classifier training component 620. The rule-based system described later is used to determine the sub-category.

Table 1: List of Content-based event-level features

Name	Levels	%images in event
People present	None/1-2/3+	25
Children under 4yrs	None/1-2/3+	25
Indoor/outdoor	Indoor/outdoor	25
Urban	NotUrban/Urban	20
Beach	NotBeach/Beach	5
Sky	<10% area/>10%	10
Grass	NotPresent/Present	15
Water	<25% area />25%	10
Field	NotPresent/Present	25
NatureScene	NotPresent/Present	12

Table 2: List of Time-based event-level features

Name	Levels	Threshold
Duration	Short/long	1.5 hours
Inter-event duration	Short/long	2 days
Time of day	Morning/Later	12pm
Time of year	Summer/Winter/Other	Jun-Aug/Dec-Feb/Rest
Number of Images	Few/Many	15

Table 3: List of top level event categories and their sub-categories

Vacation	Sports	Family Moments	Social Gathering
Daytrip	Field	Children	Wedding
Weekend	Water	Parents	Birthday
Getaway	Indoor	Siblings	Holiday
Holiday	Other	Grandparents	Party

Since conditional probabilities for some of the features in the extended feature set can be difficult or infeasible to obtain through training data, the system can also optionally apply heuristics or rules to validate or refine the output of the BBN of FIG. 4. Referring to FIG. 7, the output of the BBN-based classifier 710 can be passed to a rule-based classifier 520, where the details of the rule-based classifier are shown in FIG. 5. The rules can be based upon domain-specific knowledge, informal common sense knowledge, or formalized common sense knowledgebases such as Cyc or Open Mind Common Sense. For example, the classification "Family Moments" generated by the BBN can be validated by a rule that requires the event to portray at least one or more people that are known to be family members of the system user; if this condition does not hold, then the event is reclassified as "Social Gathering." Application of such a rule requires first use of a people recognition algorithm (U.S. Patent Application Publication

No. 2008/0112621; U.S. Patent Application Publication No. 2007/0177805; and U.S. Patent Application Publication No. 2007/0098303) to determine the identity of the individuals portrayed in the assets making up the event, and then consultation of the auxiliary factual database 116 of FIG. 1 to determine the relationship of each individual to the user of the system. Such relationships can be explicitly stored in the database for every combination of user and individual; alternatively, the system can store only canonical relationships, and infer specific relationships as requested.

Rules are also used to further refine event categories to subcategories. For example, the BBN may not be able to significantly distinguish between the various subcategories of party for a particular event, but rules leveraging auxiliary data can make the distinction. Suppose the BBN has determined a top-level category of "Social Gathering." A rule-based approach can apply the sub-category "Social Gathering—birthday" by using the results of the people recognition algorithm to determine the people portrayed in the event and applying the common sense rule that states that a social gathering is a birthday party if a person portrayed in the event has a birthday at or near the time of the event; information about individual birthdays is part of the data stored in an auxiliary factual database 116 of FIG. 1. While this classification could also potentially be generated by the BBN of FIG. 4 given the appropriate priors, computing such priors requires sufficient training data; the use of a common sense rule eliminates the need for such data.

The BBN of FIG. 4 is used to compute a set of candidate event classifications along with their associated probabilities. Referring to FIG. 7, the second stage of the event classifier 120 of FIG. 1 is the rule-based classifier 520 of FIG. 5. FIG. 5 illustrates the steps applied in using rules to validate or refine event categories generated by the BBN in the preferred embodiment. Validation rules 510 and refinement rules 515 are run by a rule-based classifier 520. Validation rules 510 determine if, given the additional feature data from feature extractors 530 as well as the data from the auxiliary factual database 116, the event labels

405 are valid; the highest probability classification passing the validity test is output as the event class. Refinement rules 515 refine the event labels 405 using additional features from feature extractors 530, as well as the data from the auxiliary factual database 116 from FIG. 1, to determine a refined semantic event sub-category. The feature extractors 530 include but are not limited to content-based feature detectors 112 and time-based feature detectors 114 from FIG. 1. In one embodiment, the rule-based system in FIG. 5 computes the sub-categories using the following rules, after the top-level category is determined using the BBN of FIG. 4:

Vacation: "Daytrip" if event duration is within the same calendar day, "Weekend" if event falls during the weekend days, "Getaway" if the vacation is long (greater than 4 days), "Holiday" if it falls during a holiday and the location is determined to be outside the user's home area.

Sports: "Field" if results from detector for field and "grass" are positive, "Water" if result from detector for water is positive, "Indoor" if result of indoor/outdoor detector is indoor, and "Other" if none of the above detectors show positive result.

Family Moments: The category is based on the relationship of the user to the people recognized in the images.

Social Gathering: "Wedding" and "Birthday" are based on information from the personal calendar of a user, "Holiday" is based on the calendar and holidays observed in the user's geographic location, and "Party" is based on the presence of many people recognized by the people recognition system as being commonly occurring in the user's collection (i.e. not strangers).

Referring to FIG. 7, some embodiments can eliminate the probabilistic classifier 710. In this case, the rules executed by the rule-based classifier 520 do not reference the event label(s) based on the extended feature set 405 of FIG. 4. Likewise, some embodiments cannot incorporate the rule-based classifier 520, producing event labels based on the probabilistic classifier alone.

In addition to permitting the user to search and browse by event type, the semantic category of event can be used to author creative output for the

user. This semantic information together with other contextual information (including, but not limited to image metadata, date/time, GPS, location, or any combination thereof) can be input into an automatic albuming system for generating themed and personalized album based on the type of event detected. For example, a vacation in Europe will suggest the use of a relevant background design and theme that reflects the cultural and regional characteristic of the location where the event took place. A party event will evoke the use of a fun and whimsical theme and mood for the album. Other custom output products can be created using the event information including collages, cups, T-shirts, and multimedia DVD and CD that includes audio and music. In addition, the semantic event information enables more efficient searching and browsing of the user collection. For instance, the user can easily search by text when the events have been automatically annotated by the system. Yet another application will be for targeted advertising based on the event detected. For example, when an outdoor sporting event is detected, relevant advertising of sporting goods can be targeted to the user.

The invention has been described in detail with particular reference to certain preferred embodiments thereof, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention.

PARTS LIST

105	Digital image collection
110	Hierarchical event detector
112	Content-based feature detectors
114	Time-based feature detectors
116	Auxiliary factual data
120	Event classifier
122	Event label
225	Event and sub-event detector
230	Inter-event duration calculator
235	Density-based classifier
240	Event hierarchy generator
305	Event labels based on image features node
310	Duration node
315	People present node
320	% Outdoors node
405	Event labels based on extended feature set node
410	Subject distance node
415	GNIS feature class node
510	Validation rules
515	Refinement rules
520	Rule-based classifier
530	Feature extractors
602	Digital image training collection
607	Manual event labeler
610	Correlated feature detectors
620	Event classifier training
710	Probabilistic classifier

CLAIMS:

1. A method of automatically classifying images in a consumer digital image collection, comprising:

(a) generating an event representation of the image collection;

(b) computing global time-based features for each event within the hierarchical event representation;

(c) computing content-based features for each image in an event within the hierarchical event representation;

(d) combining content-based features for each image in an event to generate event-level content-based features; and

(e) using time-based features and content-based features for each event to classify an event into one of a pre-determined set of semantic categories.

2. The method of claim 1, wherein the pre-determined semantic categories include vacation, sports, family moments or social gathering.

3. The method of claim 1, wherein the global time-based features include duration of event, inter-event duration, time of day, time of year or number of images in the event.

4. The method of claim 1, wherein the content-based image-level features include people present, ages of people present, indoor or outdoor, scene type, or materials present.

5. The method of claim 1, wherein the event representation is hierarchical and includes at least two levels.

6. The method of claim 1, wherein the classification in step (e) is achieved using a probabilistic classifier, a rule-based classifier or a combination thereof.

7. The method of claim 1, wherein the method of classification in step (e) includes referencing a database of auxiliary factual

information and a set of validation rules.

8. The method of claim 1, wherein the method of classification in step (e) includes referencing a database of auxiliary factual information and a set of refinement rules.

9. The method of claim 1, wherein the method of classification in step (e) includes using a rule-based system for validating and refining the output of the classifier.

10. A method of grouping time-based events detected in a collection of consumer digital images into super-events, comprising:

(a) generating time-based event boundaries;

(b) computing inter-event durations; and

(c) grouping events into clusters that are termed super-events based on the inter-event durations.

11. The method of claim 10, wherein the grouping in step c) is achieved using density-based clustering.

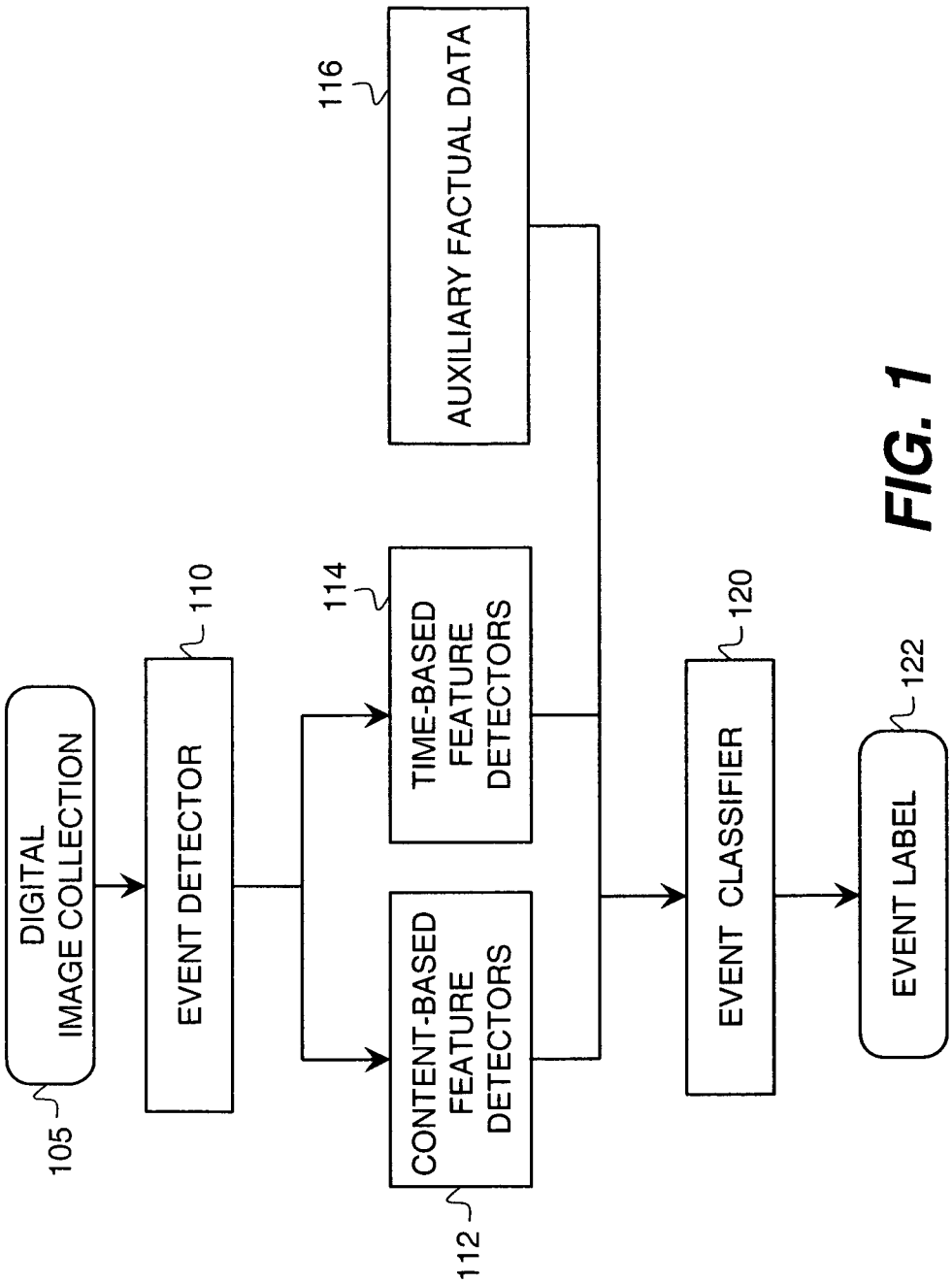
DIGITAL IMAGE COLLECTION
105
EVENT DETECTOR
110
TIME-BASED FEATURE DETECTORS
114
CONTENT-BASED FEATURE DETECTORS
112
AUXILIARY FACTUAL DATA
116
EVENT CLASSIFIER
120
EVENT LABEL
122

FIG. 1

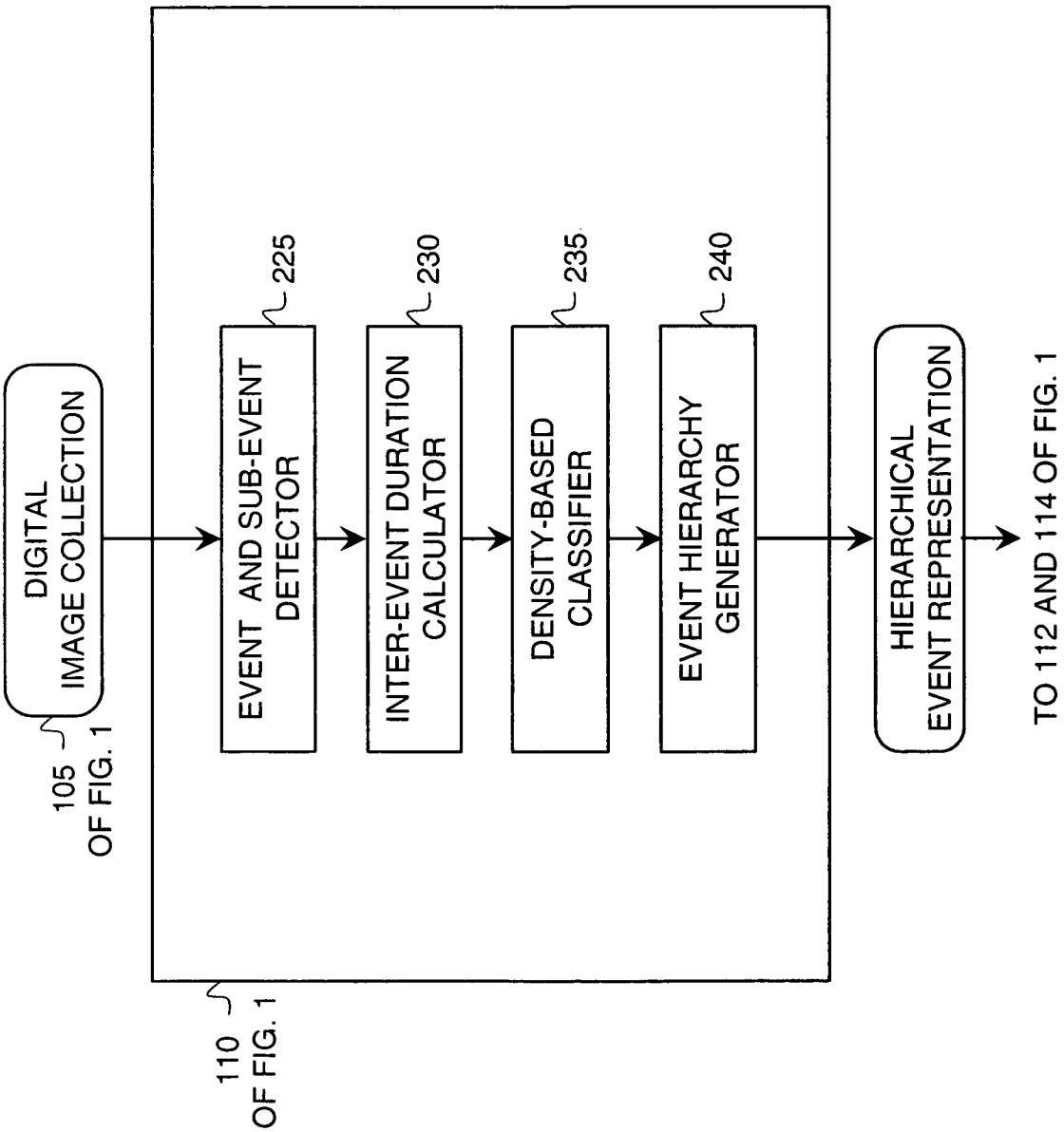
DIGITAL IMAGE COLLECTION
105 OF FIG. 1
EVENT AND SUB-EVENT DETECTOR
225
INTER-EVENT DURATION CALCULATOR
230
DENSITY-BASED CLASSIFIER
235
EVENT HIERARCHY GENERATOR
240
HIERARCHICAL EVENT REPRESENTATION
TO 112 AND 114 OF FIG. 1
110 OF FIG. 1

FIG. 2

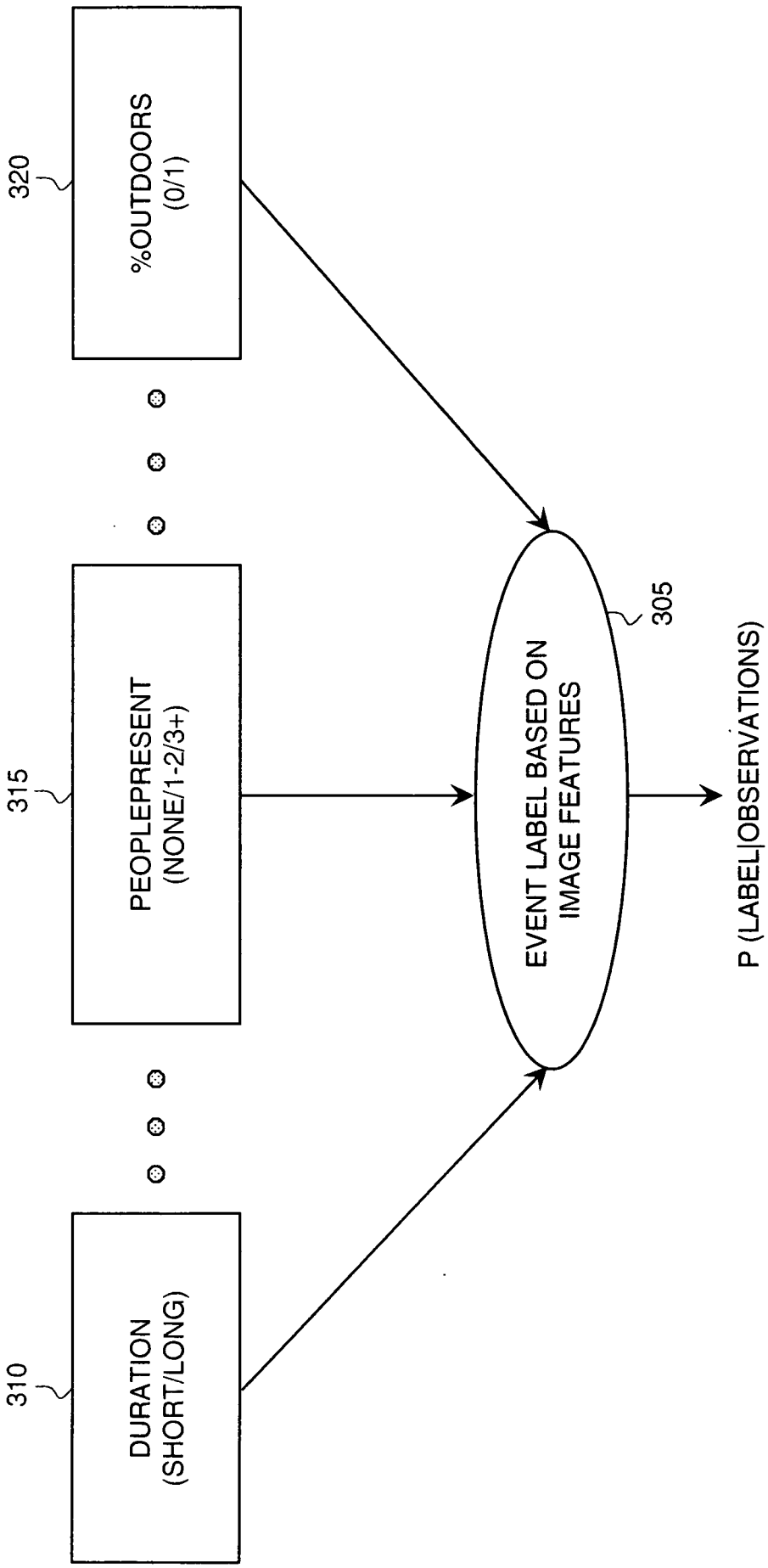
320
%OUTDOORS
(0/1)
315
PEOPLEPRESENT
(NONE/1-2/3+)
310
DURATION
(SHORT/LONG)
EVENT LABEL BASED ON
IMAGE FEATURES
305
P (LABEL|OBSERVATIONS)

FIG. 3

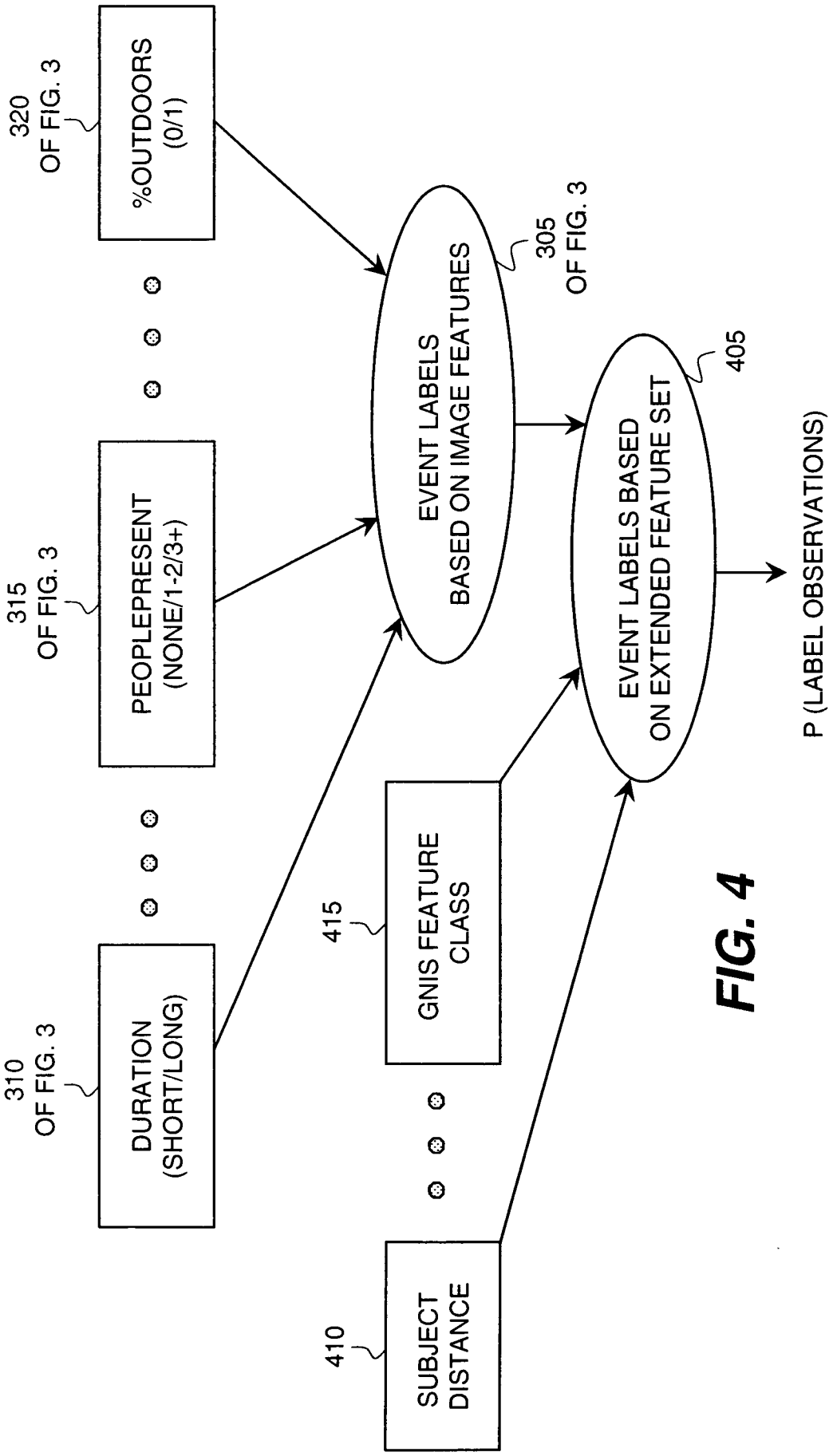
320
OF FIG. 3
%OUTDOORS
(0/1)
315
OF FIG. 3
PEOPLEPRESENT
(NONE/1-2/3+)
310
OF FIG. 3
DURATION
(SHORT/LONG)
415
GNIS FEATURE
CLASS
410
SUBJECT
DISTANCE
EVENT LABELS
BASED ON IMAGE FEATURES
305
OF FIG. 3
EVENT LABELS BASED
ON EXTENDED FEATURE SET
405
P (LABEL OBSERVATIONS)

FIG. 4

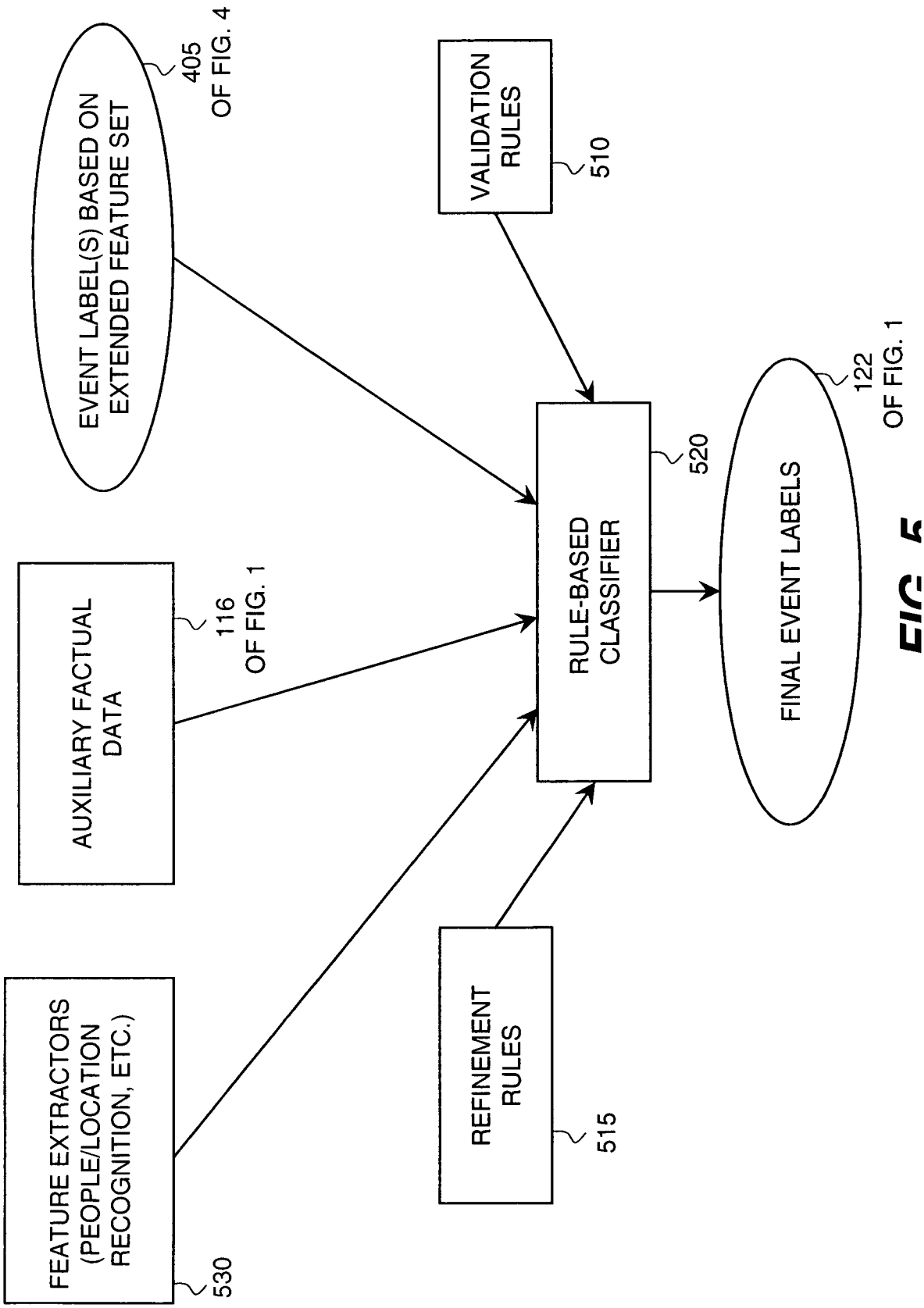
EVENT LABEL(S) BASED ON EXTENDED FEATURE SET
405 OF FIG. 4
VALIDATION RULES
510
AUXILIARY FACTUAL DATA
116 OF FIG. 1
RULE-BASED CLASSIFIER
520
FINAL EVENT LABELS
122 OF FIG. 1
FEATURE EXTRACTORS (PEOPLE/LOCATION RECOGNITION, ETC.)
530
REFINEMENT RULES
515

FIG. 5

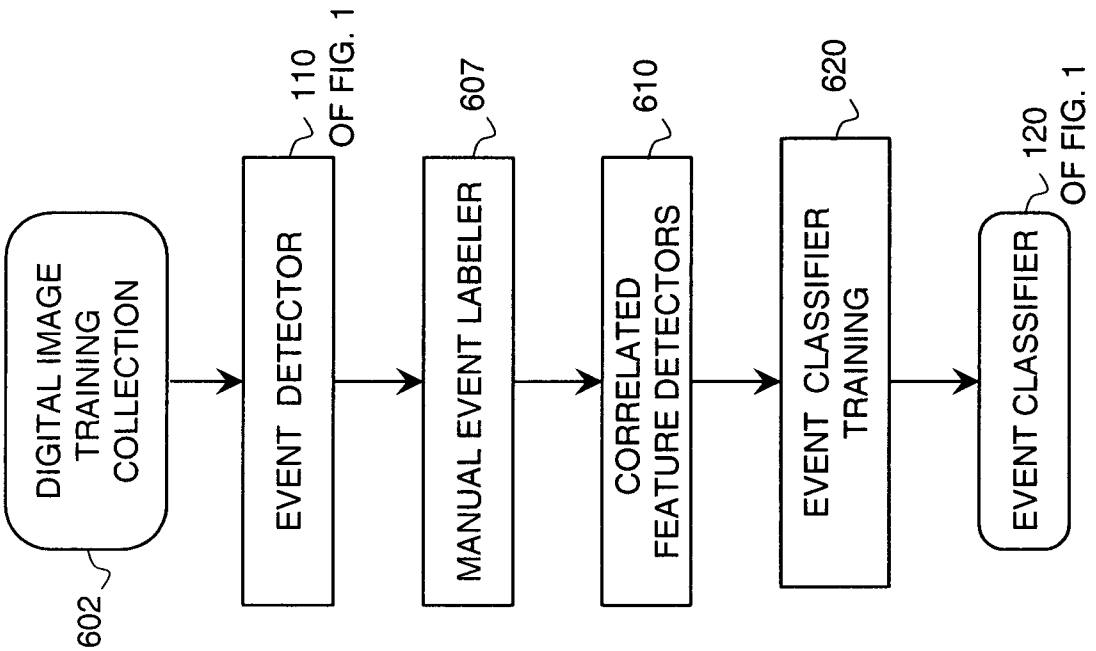
DIGITAL IMAGE TRAINING COLLECTION
602
EVENT DETECTOR
110 OF FIG. 1
MANUAL EVENT LABELER
607
CORRELATED FEATURE DETECTORS
610
EVENT CLASSIFIER TRAINING
620
EVENT CLASSIFIER
120 OF FIG. 1

FIG. 6

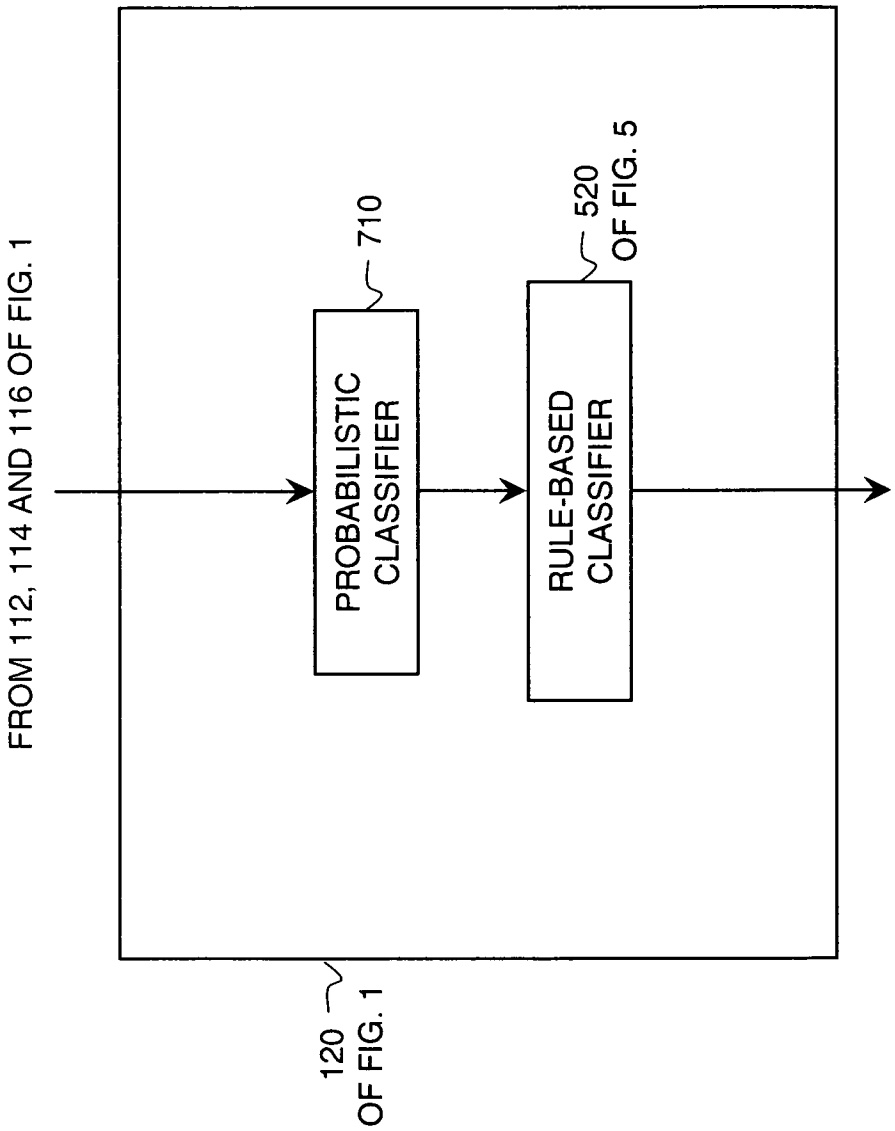
FROM 112, 114 AND 116 OF FIG. 1
PROBABILISTIC CLASSIFIER
710
RULE-BASED CLASSIFIER
520 OF FIG. 5
TO 122 OF FIG. 1
120 OF FIG. 1

FIG. 7