



US 20170293847A1

(19) **United States**

(12) **Patent Application Publication**  
**Robertson et al.**

(10) **Pub. No.: US 2017/0293847 A1**

(43) **Pub. Date: Oct. 12, 2017**

(54) **CRIME RISK FORECASTING**

**Publication Classification**

(71) Applicant: **Palantir Technologies, Inc.**, Palo Alto, CA (US)

(51) **Int. Cl.**

**G06N 5/04** (2006.01)

**G06Q 50/26** (2006.01)

(72) Inventors: **Duncan Robertson**, Sutton Scotney (GB); **Alexander Sparrow**, Bath (GB); **Mike Lewin**, London (GB); **Meline Von Brentano**, London (GB); **Matthew Elkherj**, Palo Alto, CA (US); **Rafael Cosman**, La Jolla, CA (US)

(52) **U.S. Cl.**

CPC ..... **G06N 5/04** (2013.01); **G06Q 50/265** (2013.01); **G06F 3/0481** (2013.01)

(57) **ABSTRACT**

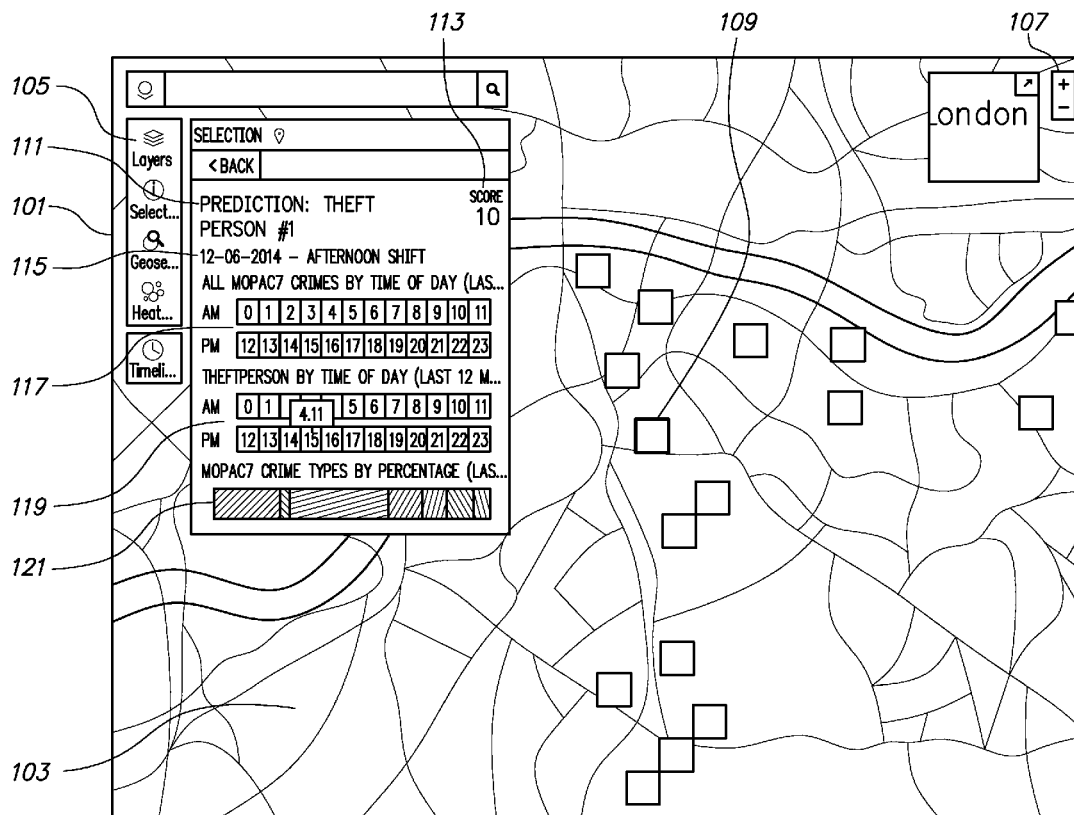
(21) Appl. No.: **15/628,832**

(22) Filed: **Jun. 21, 2017**

**Related U.S. Application Data**

(63) Continuation of application No. 14/843,734, filed on Sep. 2, 2015, which is a continuation of application No. 14/319,161, filed on Jun. 30, 2014, now Pat. No. 9,129,219.

A computer-based crime risk forecasting system and corresponding method are provided for generating crime risk forecasts and conveying the forecasts to a user. With the conveyed forecasts, the user can more effectively gauge both the level of increased crime threat and its potential duration. The user can then leverage the information conveyed by the forecasts to take a more proactive approach to law enforcement in the affected areas during the period of increased crime threat.



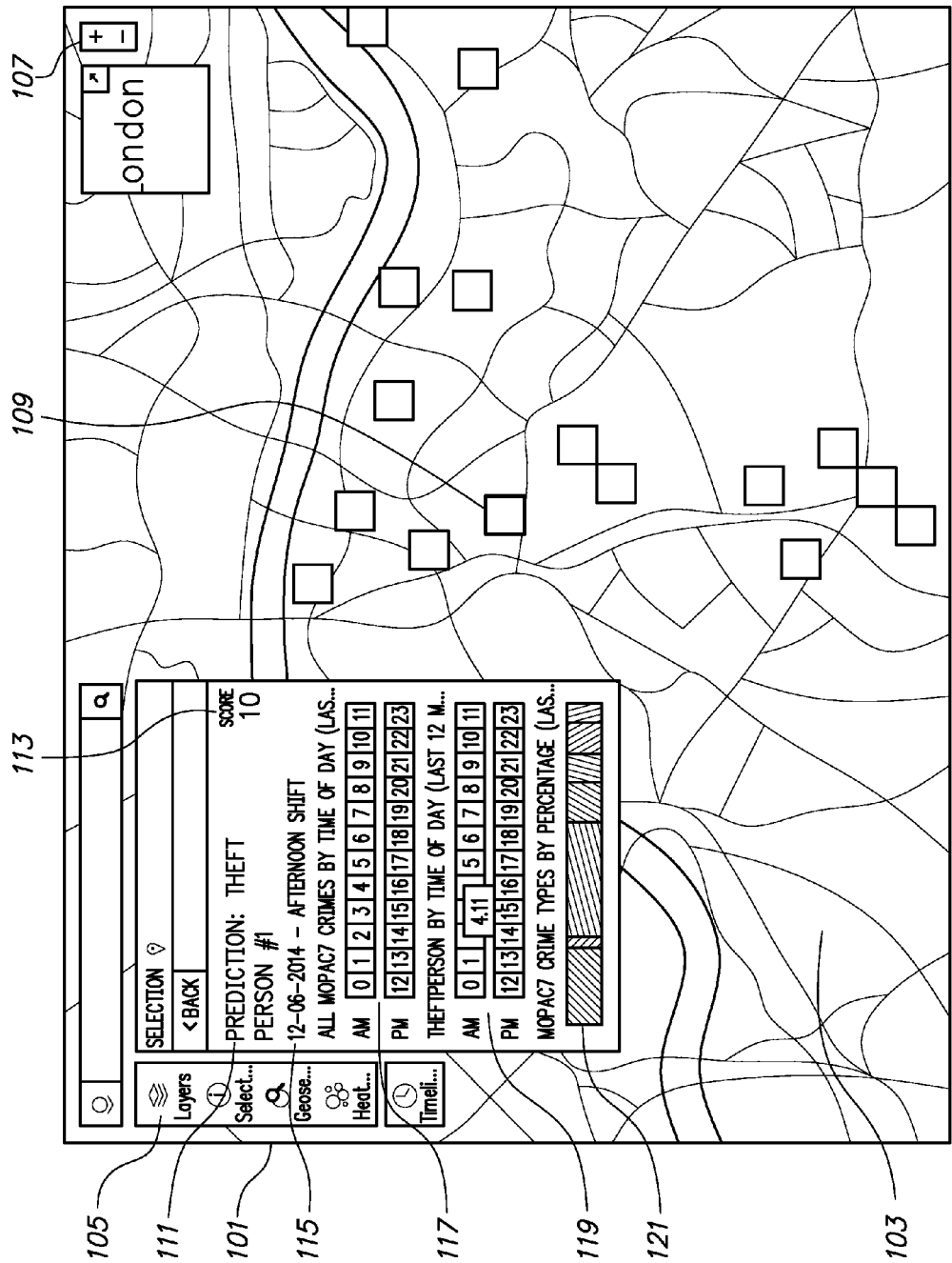


FIG. 1A

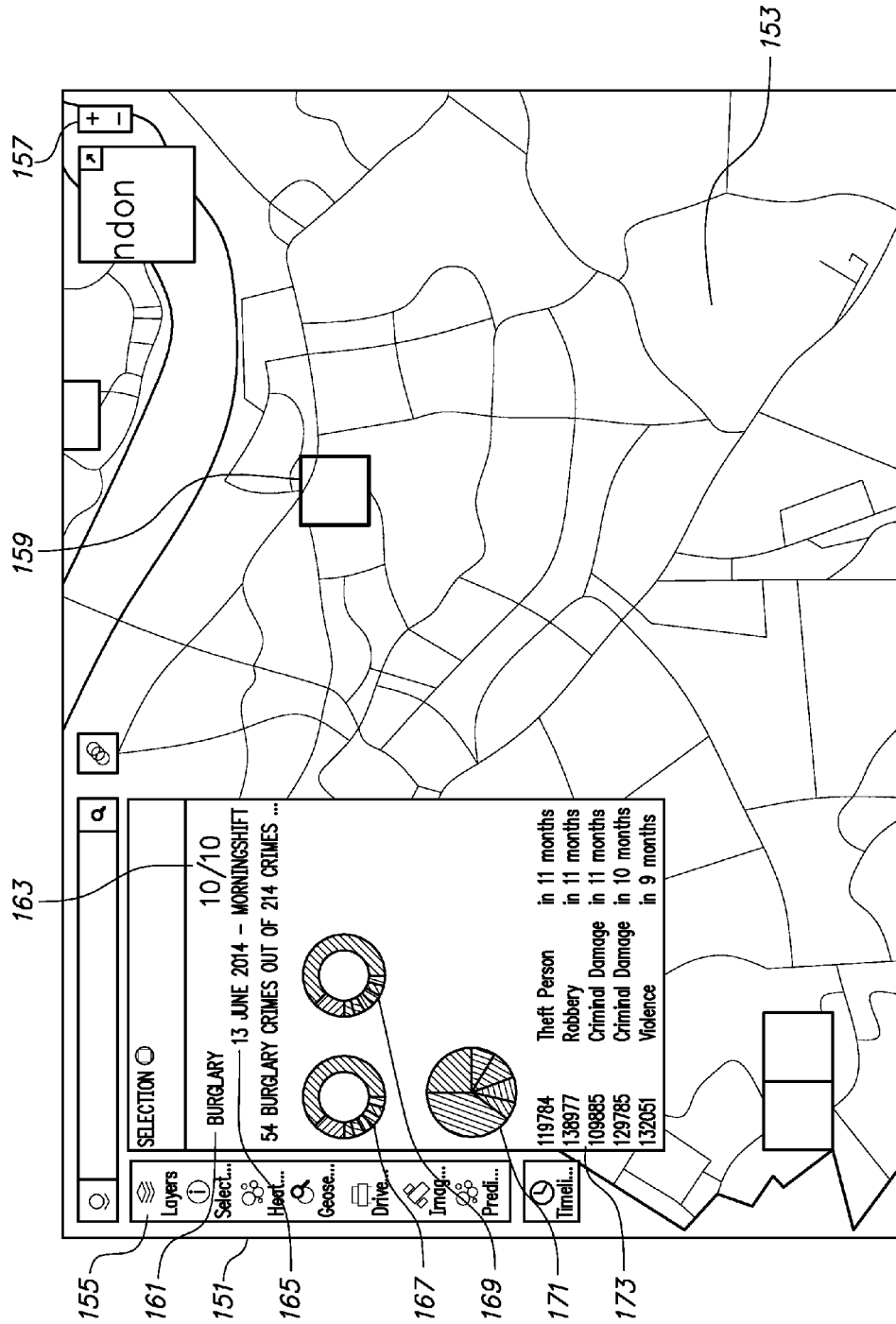
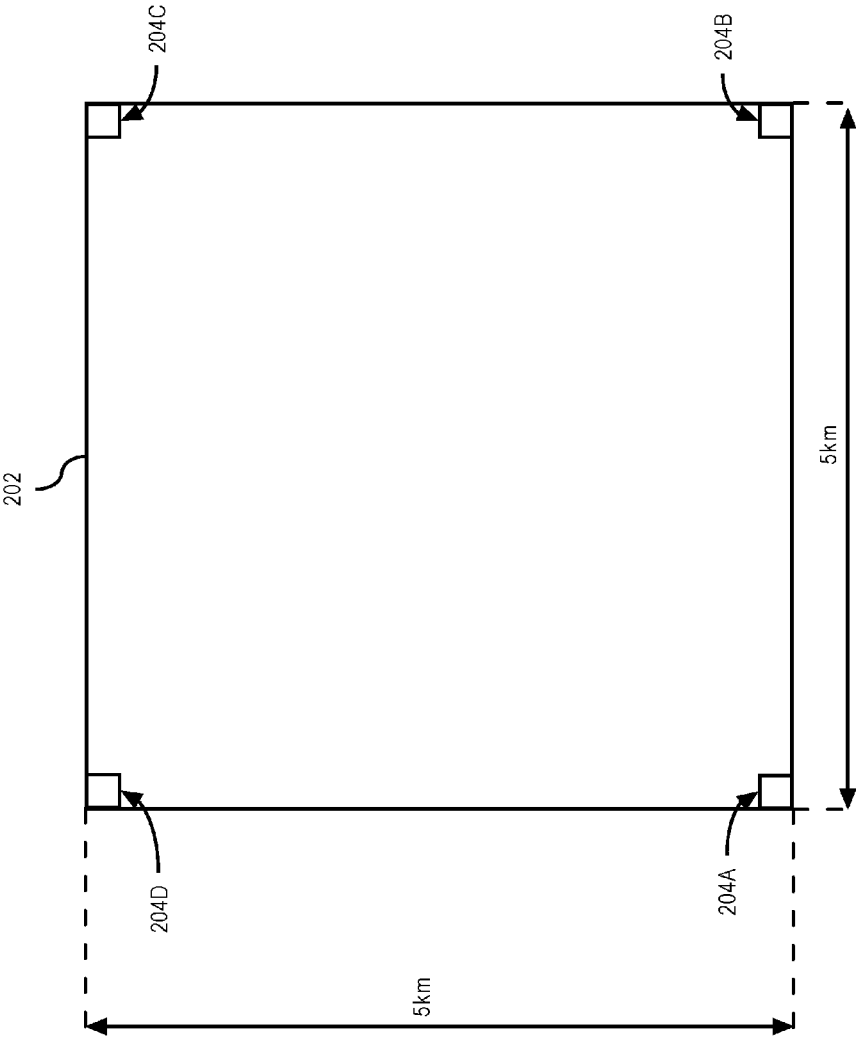


FIG. 1B



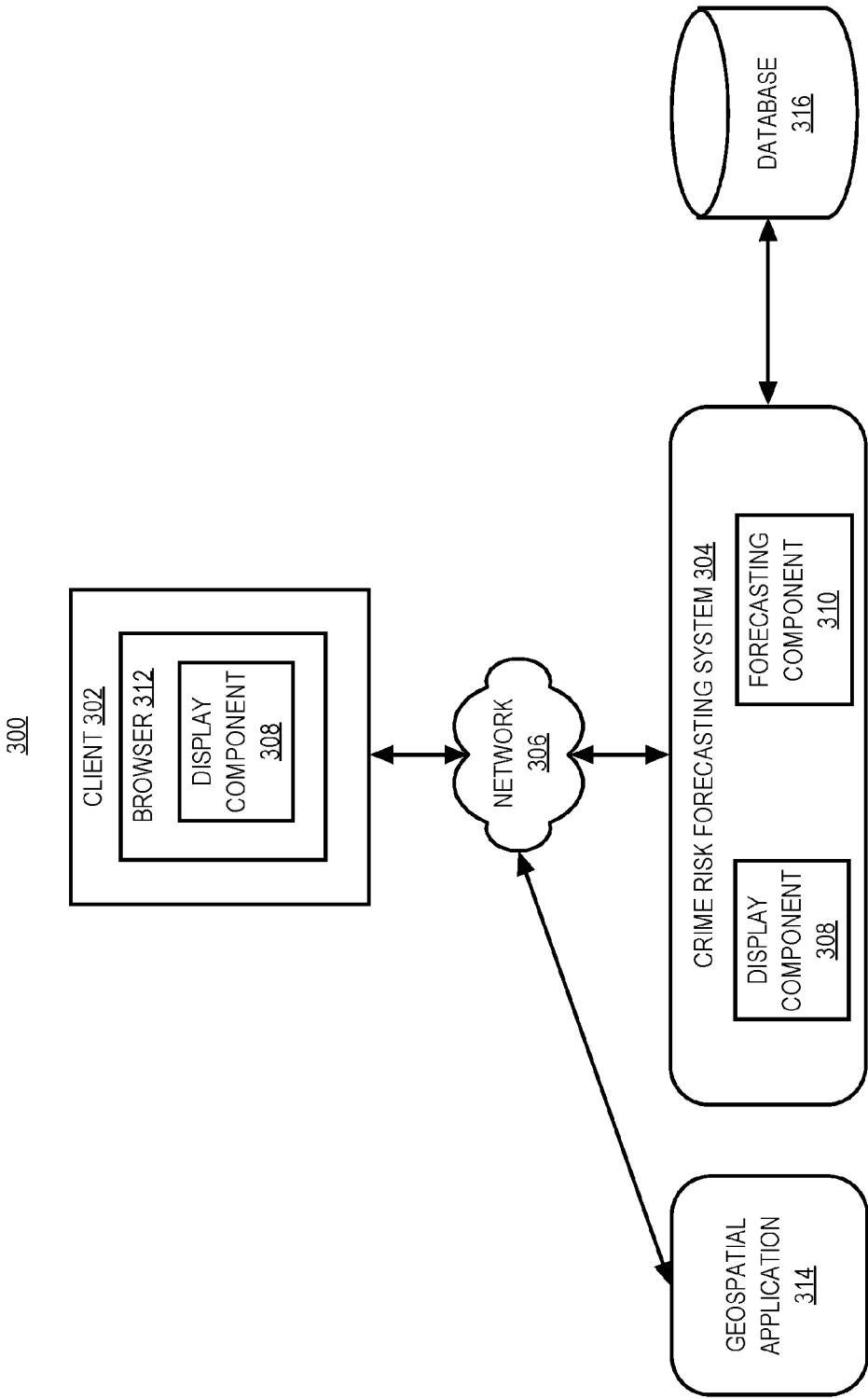
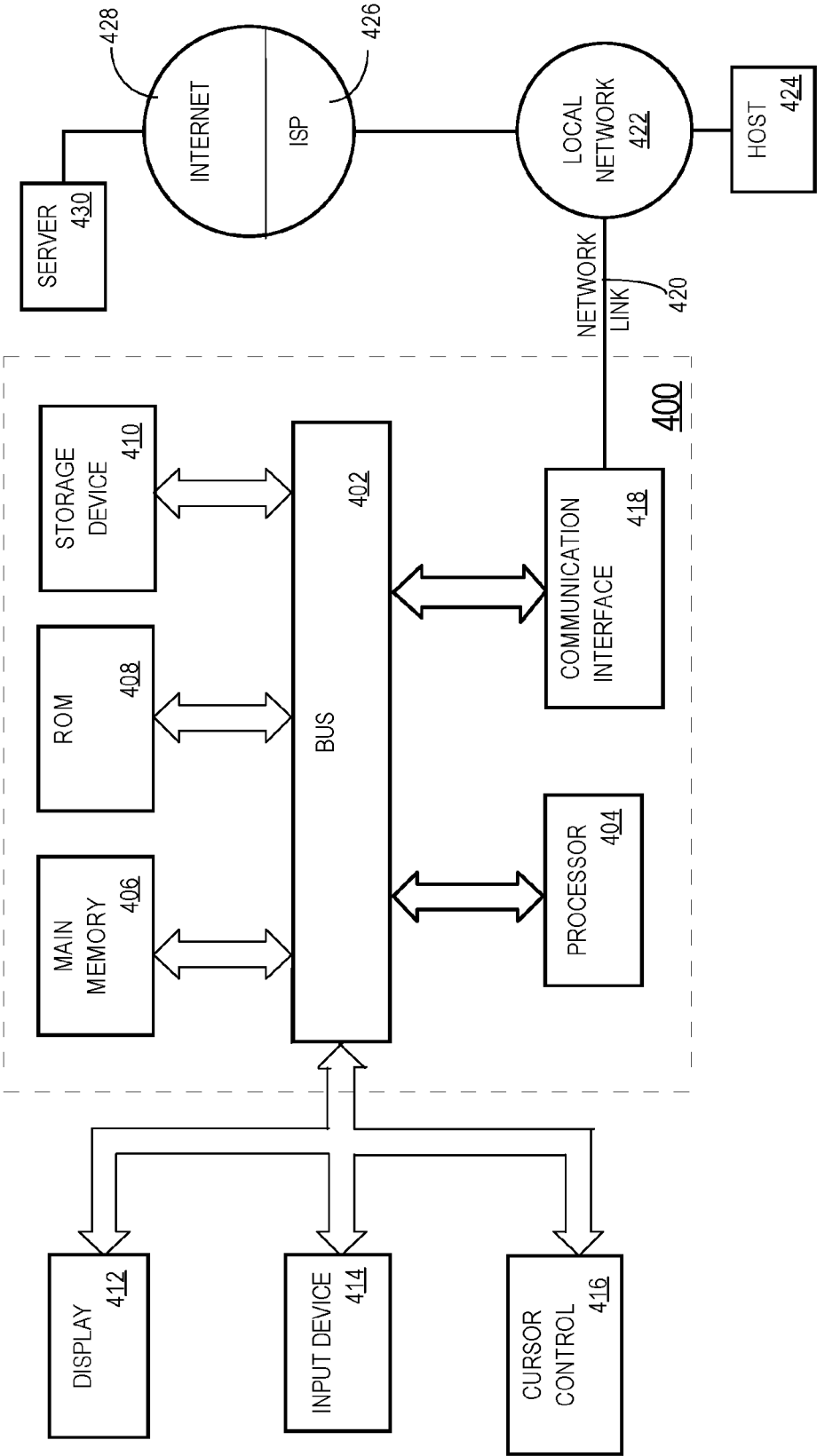


FIG. 3

FIG. 4



## CRIME RISK FORECASTING

### CROSS-REFERENCE TO APPLICATIONS

[0001] This application is a Continuation of U.S. application Ser. No. 14/843,734, filed Sep. 2, 2015, which is a Continuation of U.S. application Ser. No. 14/319,161, filed Jun. 30, 2014, now U.S. Pat. No. 9,129,219, the entire contents of each of which is hereby incorporated by reference for all purposes as if fully set forth herein. The applicant(s) hereby rescind any disclaimer of claim scope in the parent application(s) or the prosecution history thereof and advise the USPTO that the claims in this application may be broader than any claim in the parent application(s).

[0002] This application is also related to U.S. patent application Ser. No. 13/917,571, filed Jun. 13, 2013, entitled "Interactive Geospatial Map", which is incorporated by reference herein in its entirety.

### TECHNICAL FIELD

[0003] The disclosed embodiments relate generally to computing devices. More specifically, the disclosed embodiments relate to computing devices and computer-implemented methods for generating crime risk forecasts and conveying the forecasts to a user.

### BACKGROUND

[0004] Interactive geospatial maps, such as those produced by Internet-based mapping systems or other computer-based geospatial information systems (GIS), are available from several providers. These interactive maps typically comprise satellite imagery or graphical basemaps that provide an aerial or bird's-eye perspective of a curved geographic surface, such as the surface of the Earth, after being projected using a map projection (e.g., a Mercator map projection). Some of the interactive basemaps may include one or more situational data layers displayed as overlays on the basemaps that visually convey various situational features such as roads, traffic, buildings, parks, restaurants, banks, schools, and other situational features.

### SUMMARY

[0005] The claims section appended hereto provides a useful summary of some embodiments of the present invention.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0006] In the drawings:

[0007] FIG. 1A is a screen shot of a computer graphical user interface showing an example generated crime risk forecast overlay to an interactive geospatial basemap per an embodiment of the present invention.

[0008] FIG. 1B is another screen shot of a computer graphical user interface showing another example generated crime risk forecast overlay to an interactive geospatial basemap per an embodiment of the present invention.

[0009] FIG. 2 is a schematic of a 5 kilometer by 5-kilometer geographic surface region divided into 400 square grid areas, each 250 by 250 meters in length.

[0010] FIG. 3 is a block diagram of an example web-based geographic crime risk forecasting computer system in which an embodiment of the present invention is implemented.

[0011] FIG. 4 is a block diagram that illustrates a computing device upon which embodiments of the present invention may be implemented.

### DETAILED DESCRIPTION

[0012] In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form to avoid unnecessarily obscuring the present invention.

### OVERVIEW

[0013] Police departments and other law enforcement agencies would like to know when and where crimes are most likely to occur in the future to most efficiently and effectively allocate crime prevention resources. A computer-based crime risk forecasting system and corresponding method are provided for generating crime risk forecasts and conveying the forecasts to a user. The user may be a police officer or another law enforcement official, for example. With the conveyed forecasts, the user can more effectively gauge both the level of increased crime threat and its potential duration. The user can then leverage the information conveyed by the forecasts to take a more proactive approach to law enforcement in the affected areas during the period of increased crime threat.

### Crime Risk Forecasting System

[0014] Per an embodiment of the present invention, crime risk forecasts are generated and conveyed by a computer-based crime risk forecasting system. The crime risk forecasting system has at least two components: a forecasting component and a display component. The forecasting component and the display component may each be implemented in software, in hardware, or in a combination of software and hardware per the requirements of the implementation at hand.

[0015] The forecasting component can use several different algorithms for generating the crime risk forecasts. Exemplary crime risk forecasting algorithms are described in greater detail below.

[0016] The display component provides functionality for displaying generated forecasts to a user of the system. The display component incorporates a geospatial application capable of generating interactive geospatial basemaps such as, for example, those disclosed in related U.S. patent application Ser. No. 13/917,571, filed Jun. 13, 2013, entitled "Interactive Geospatial Map", which is incorporated by reference herein in its entirety.

[0017] The display component displays the generated crime risk forecast as an overlay to an interactive geospatial basemap generated by the geospatial application. By incorporating interactive geospatial basemaps generated by the geospatial application, the display component can provide environmental context to the conveyed forecast. The environmental context may indicate nearby buildings, roads, banks, automated teller machines, public transportation hubs, subway entrances and exits, and other environmental and situational information which can suggest why a geographic area is forecasted as a crime risk.

### Crime Risk Forecasts

**[0018]** The crime risk forecast is generated for a “target geographic area”, or just “target area” for short. For example, the target area may correspond to 250 by 250 square meters of a geographic area. For example, the target area may correspond to a city block, or a portion thereof. Other sized target areas are possible and the target area need not correspond to only a 250 by 250 square meter geographical area. More generally, the size of a target area may vary from implementation to implementation per the requirements of the particular implementation at hand.

**[0019]** In addition to being generated for the target area, the crime risk forecast may also be generated for a time window. As used herein, the term “time window” refers to a continuous period of time. For example, the time window for which the forecast is generated may correspond to an 8-hour law enforcement patrol shift on a particular day. For example, the day may be divided into three patrol shifts: a morning shift from 6 A.M. to 2 P.M., a late shift from 2 P.M. to 10 P.M., and a night shift from 10 P.M. to 6 A.M. A time window may correspond to one of the three patrol shifts on a given day. Other length time windows are possible and a time window need not be only 8-hours in length or span only a single day. More generally, the length of the time window may vary from implementation to implementation per the requirements of the implementation at hand.

**[0020]** In addition to being generated for the target area and the time window, the crime risk forecast may also be generated for a predefined crime type in a set of predefined crime types. For example, the set of predefined crime types may include one or more general personal and property crimes such as assault, battery, kidnapping, homicide, offenses of a sexual nature, larceny (theft), robbery (theft by force), vehicle theft, burglary, arson, and so forth. In some instances, the set of predefined crime types includes a crime type that is a combination of two or more crime types. For example, the set of predefined crime types may include a robbery and theft crime type instead of having separate robbery and theft crime types. More generally, the set of predefined crime types for which the forecast can be generated may vary from implementation to implementation per the requirements of the implementation at hand.

**[0021]** Multiple crime risk forecasts may be generated for the same target area in which each forecast corresponds to a predefined crime type and a time window. For example, a crime risk forecast may be generated for burglary during the late patrol shift on a particular day in a target area, another forecast generated for burglary during the early patrol shift on the particular date in the target area, yet another forecast generated for vehicle theft during the late patrol shift on the particular date in the target area, and so forth.

**[0022]** A generated crime risk forecast has a visible manifestation presented to the user. For example, the visible manifestation can be presented to the user in a web browser or another computer graphical user interface. The visible manifestation of the crime risk forecast includes a visually highlighted area corresponding to the target area on the interactive geospatial basemap generated by the geospatial application. The visually highlighted area indicates the geographic target area the crime risk forecast applies to.

**[0023]** The visual highlighting of the target area can take a variety of different forms including a bounding box or other geometric outline that indicates the geographic borders of the target area relative to the interactive geospatial

basemap. The interior of the outlined geographic shape may also be colored with a semi-transparent fill that allows for visual perception of the underlying basemap, at least some degree corresponding to the level of transparency of the fill color. The fill color may indicate, for example, that the visual highlighting corresponds to a crime risk forecast, as opposed to a previous crime incident.

**[0024]** In addition to indicating the target area to which the crime risk forecast applies, the visible manifestation may convey a crime risk rating for the crime type and the time window the forecast was generated for. The crime risk rating conveys to the user the crime risk level in the target area for the crime type during the time window. For example, the risk rating may be a quantitative value such as a number on a scale of 1 to 10, with 10 being the highest risk of the crime type occurring in the target area within the time window and 1 being the lowest risk of the crime type occurring in the target area and during the time window. Other risk rating scales are possible and a numerical risk rating scale of 1 to 10 is not required. For example, the crime risk rating scale could be “A”, “B”, “C”, “D”, “E” and “F” with “A” indicating the lowest (or highest) crime risk and “F” indicating the highest (or lowest) crime risk. Qualitative risk ratings such as “low risk”, “moderate risk”, and “high risk” are also possible.

**[0025]** In addition to visually highlighting the target area on the interactive geospatial basemap and providing the risk rating, the visible manifestation of the crime risk forecast may convey the algorithm used to generate the forecast. In some instances, a different algorithm is used to generate forecasts for different time windows. For example, forecasts generated for the early patrol shift, the late patrol shift, and the night patrol shift may be generated by different algorithms. The visible manifestation of the crime risk forecasts may indicate which algorithm was used to generate the risk ratings for each of the time windows. By doing so, the user can acquire a better sense of why a risk rating for a time window (e.g., the early shift) may be similar or different from a risk rating for another time window (e.g., the night shift).

**[0026]** FIG. 1A is a screen shot of a computer graphical user interface **101** showing an example generated crime risk forecast overlay to an interactive geospatial basemap **103** per an embodiment of the present invention. The interactive geospatial basemap **103** is generated by a geospatial application and the overlay is generated by the display component of the crime risk forecasting system based on a crime risk forecast generated by the forecasting component of the crime risk forecasting system. The interactive geospatial basemap **103** includes graphical user interface controls **105** and **107** for adjusting the display of the basemap **103**.

**[0027]** More specifically, controls **105** allow the user to selectively add and remove geospatial situational layers to and from the basemap **105**. The layers can include, but are not limited to, one or more vector layers that convey geographical regions, roads, bridges, buildings/structures, terrain, transportation hubs, utilities, infrastructure, street lights, hotels/motels, railroads, hospitals, other types of buildings or structures, regions, transportation objects, and/or other types of entities and events. The vector layers may overlay one or more base layers to form basemap **103**. The base layers may include, for example, overhead (e.g., aerial or satellite) imagery, topographic, blank projected (e.g., Mercator projected), basemap, and blank unprojected. Con-



trols **107** allow the user to adjust the zoom level of the basemap **103** either by increasing the zoom level to a smaller geographic area but in greater detail or by decreasing the zoom level to a larger geographic area but in less detail.

[0028] Basemap **103** includes a map of a city. More specifically, basemap **103** represents an area of London. Basemap **103** is overlaid with several squares each representing the target area of a corresponding crime risk forecast generated by the forecasting component. One of the squares **109** is currently selected by the user and the details of the corresponding crime risk forecast are shown in a panel overlay to the basemap **103** generated by the display component. More specifically, the overlay panel shows that the crime risk forecast corresponding to square **109** is for the crime type **111** Theft Person and for the time window **115** of the afternoon shift on Jun. 12, 2014. The panel overlay also includes a risk rating **113** of 10.

[0029] The overlay panel generated by the display component also includes other contextual information to aid the user in better understanding the selected crime risk forecast. More specifically, time of day graphical user interface element **117** color codes each hour of the day by the number of the historical crimes that occurred in the hour irrespective of crime type. Relatively darker coloring indicates that relatively more of the historical crime incidents occurred during that hour of the day. Relatively lighter coloring indicates that relatively fewer of the historical crime incidents occurred during that hour of the day. For example, the relatively darker coloring to the 9 a.m. hour suggests that higher crime activity may be associated with persons arriving at work in downtown London. Time of day graphical user interface element **119** is like element **117** except just for historical crime incidents of the crime type Theft Person. The relatively darker coloring to the 9 a.m. hour suggests that a relatively higher number of incidents of Theft Person crime type may also be associated with persons arriving at work in downtown London. Color bar **121** breaks down all historical crime incidents in the last twelve months based on which the corresponding crime risk forecast is generated by crime type where each different color corresponds to a different crime type and the percentage of the color bar **121** occupied by a color represents the percentage of all historical crime incidents in the last twelve months that are of the corresponding crime type.

[0030] FIG. 1B is a screen shot of a computer graphical user interface **151** showing another example generated crime risk forecast overlay to an interactive geospatial basemap **153** per an embodiment of the present invention. Like in basemap **103** of FIG. 1A, the interactive geospatial basemap **153** is generated by a geospatial application and the overlay is generated by the display component of the crime risk forecasting system based on a crime risk forecast generated by the forecasting component of the crime risk forecasting system.

[0031] Basemap **153** includes a map of a city. More specifically, basemap **153** represents an area of London. Basemap **153** is overlaid with several squares each representing the target area of a corresponding crime risk forecast generated by the forecasting component. One of the squares **159** is currently selected by the user and the details of the corresponding crime risk forecast are shown in a panel overlay to the basemap **153** generated by the display component. More specifically, the overlay panel shows that the crime risk forecast corresponding to square **159** is for the

crime type **161** Burglary and for the time window **165** of the morning shift on Jun. 13, 2014. The panel overlay also includes a risk rating **163** of 10 out of 10.

[0032] In this example, as indicated in the overlay panel, the crime risk forecast is generated based on **214** historical crime incidents, 54 of which are burglary crime types. Color wheel graphical user interface element **167** color codes each hour of the day by the number of the 214 historical crime incidents that occurred in the hour irrespective of crime type. Relatively darker coloring indicates that relatively more of the 214 historical crime incidents occurred during that hour of the day. Relatively lighter coloring indicates that relatively fewer of the 214 historical crime incidents occurred during that hour of the day. Color wheel element **169** is like element **167** except just for the 54 of the 214 historical crime incidents that are of the crime type Burglary. Pie chart **171** is like color bar **121** of FIG. 1A except that it shows percentages by pie slices instead of by coloring portions of a color bar. Listing **173** shows the last five historical crime incidents to have occurred in the geographical area represented by square **159** and when they occurred.

[0033] By generating a visual manifestation of a crime risk forecast as an overlay to an interactive geospatial basemap generated by a geospatial application, the layering features and functions of the geospatial application can be used to provide the user greater context to the forecast. More specifically, the user can add and remove situational geospatial layers to provide environment context to the forecast. The environmental context may indicate nearby buildings, roads, banks, automated teller machines, public transportation hubs, subway entrances and exits, and other environmental and situational information which can suggest why a geographic area is forecasted as a crime risk.

#### Generating a Crime Risk Forecast Based on Machine Learning

[0034] As described in greater detail below, several different algorithms may be used to generate a crime risk forecast. In some instances, a crime risk forecast is based on machine learning.

[0035] When generating a crime risk forecast based on machine learning, historical crime incident features may be considered. Such features may include, for example, frequency of crime incidents in the target area, the number of crime incidents in the target area in a past period of time (e.g., the past week), the number of crime incidents in the target area in an extended past period of time (e.g., the past few weeks), the number of crime incidents in a past period of time in neighboring areas, and so forth.

[0036] To generate a crime risk forecast based on machine learning, a machine learning algorithm such as, for example, a support vector machine, neural network, logistic regression or any other algorithm may be applied to available historical crime incident features. Example neural network and logistic regression algorithms for generating crime risk forecasts are described in greater detail below.

[0037] A crime risk forecast generated by machine learning may be based on historical crime incident data associated with the target area. The historical crime incident data may pertain to only incidents within the target area over a given period(s) of time. Alternatively, the historical crime incident data may also pertain to incidents within neighboring or surrounding areas over a given period(s) of time.

**[0038]** In addition to the historical crime incident data, machine learning can be used to generate a crime risk forecast based on “non-incident” historical information. Such non-incident historical information may include, for example, weather information for the target area over a given period(s) of time and level of law enforcement patrol presence in or near the target area over a given period(s) of time.

**[0039]** Level of law enforcement patrol presence in or near a target area can be determined from global positioning system (GPS) information obtained from radio equipment used by patrol officers over a given period(s) of time. Thus, crime risk forecasts generated according to the present invention are not limited to being generated based only on historical crime incident information but may also be based on other types of historical information such as historical weather information and historical law enforcement officer patrol activity.

**[0040]** Further, by collecting and maintaining information on historical law enforcement officer patrol activity in or near a target area and presenting such information in conjunction with a crime risk forecast for the target area, the user can see the relationship between historical law enforcement presence in or near the target area and the predicted risk of crime in the target area. For example, a first generated crime risk forecast for a first target area can indicate that the future crime risk in the first target area is relatively low and a second generated crime risk forecast for a second target area can indicate that the future crime risk in the second target area is relatively high. At the same, an indication of historical law enforcement presence in or near the first target area can indicate that the historical presence was relatively high and another indication of historical law enforcement presence in or near the second target area can indicate that the historical presence was relatively low. Based on conveying these forecasts and law enforcement presence indications to the user, the user may decide to re-allocate some of the patrols assigned to areas in or near the first target area to areas in or near the second target area.

#### Historical Crime Incident Data

**[0041]** Historical crime incident data on which generated crime risk forecasts may be based can include metadata pertaining to crime incidents. Such metadata may include the crime type, the date and time of the crime, the geographic location of the crime and the method of the crime.

**[0042]** The crime type can be one of an enumeration such as burglary, robbery, theft of vehicle, theft from vehicle, criminal damage, violence, and so forth.

**[0043]** The date and time may correspond to a law enforcement or police dispatch time, for example, or other date and time that indicates when the corresponding crime incident occurred or was reported to law enforcement.

**[0044]** In some instances, the date and time of the crime is a date/time range as opposed to a discrete point in time. In these instances, the date and time of the crime can be treated as a uniform probability distribution over the date/time range.

**[0045]** In an embodiment, a geographic surface region is divided into uniformly sized areas in a grid-like fashion and location metadata for a crime incident specifies or indicates one of the areas in a region.

**[0046]** FIG. 2 is a schematic of a 5 kilometers (km) by 5 km region **202** divided into 400 square grid cells, each 250

by 250 meters (m) in length. In this example, only four of the 400 square grid cells are illustrated and are labeled **204A**, **204B**, **204C**, and **204D**. Region **202** and cells within region **202** can be defined with respect to a geographic surface by geographic coordinates such as latitude and longitude coordinates. Region **202**, for example, may correspond to a town or a neighborhood of a city. Cells within region **202** may correspond to a portion of the town or neighborhood such as, for example, a portion of a city block.

**[0047]** Region **202** may be of square or rectangular dimensions other than 5 km by 5 km or any other shape. Similarly, cells with region **202** may be of square or rectangular dimensions other than 250 m by 250 m or any other shape, per the requirements of the implementation at hand.

**[0048]** Location metadata for a historical crime incident can specify a cell in a region where the crime incident occurred. For example, the location metadata can include a region identifier that specifies the region within which the crime incident occurred.

**[0049]** The location metadata can also include grid coordinates that specifies a cell **204** within the identified region **202** within which the crime incident occurred. For example, each cell within a region may be identified by an x, y grid coordinate, where x corresponds to the horizontal axis of the region and y corresponding to the vertical axis of the region. For example, cell **204A** in region **202** may be identified by the grid coordinate x=1, y=1; cell **204B** identified by the grid coordinates x=20, y=1; cell **204C** identified by the grid coordinates x=20, y=20; cell **204D** identified by the grid coordinates x=1, y=20; and so forth. Other grid coordinate schemes as possible and the present invention is not limited to any manner for specifying a cell within a region in location metadata.

**[0050]** In addition to or instead of explicitly specifying a region and a cell within the region within which the crime incident occurred, the location metadata can indicate the cell and region where the crime incident occurred without explicitly identifying the regions and the cell. In this case, the metadata indicating the cell and region will need to be resolved to determine the cell **204** and the region **202** the metadata indicates. For example, if the location metadata is a street address location of the crime incident, the street address can be resolved to geographic coordinates (e.g., latitude/longitude coordinates) using a geocoding application. Then, the geographic coordinates returned for the street address by the geocoding application can be resolved to a cell within a region based on geographic coordinates associated with the cell and the region. For example, the cell and region can be determined by identifying the region and the cell within the region the geographic coordinates for the street address are within or nearest to. Alternatively, the location metadata can indicate a region and a cell within the region with geographic coordinates. In this case, consulting a geocoding application to resolve a street address would not be necessary.

**[0051]** While in some embodiments a geographic area is divided into region(s) and grid cells before crime risk forecasts are generated for the area, the area is divided in a grid-like fashion after crime risk forecasts are generated for the area in other embodiments. In other words, an algorithm for generating a crime risk forecast for a geographic area may not require the area to have been divided into grid cells prior to performing the algorithm to generate the forecast. However, a different algorithm may require an area to have

been divided into grid cells before the algorithm can generate a forecast for a grid cell or cell(s) in the area.

**[0052]** Crime incident metadata for a historical crime incident can include information about the crime incident other than just the crime type, crime date/time, and crime location. For example, if the crime incident involves entry into physical premises such as a building, home, or office, then the crime metadata for the incident may indicate the entry method used by the perpetrator. For example, if the crime incident was a burglary of a home, then the metadata for the incident may specify whether the perpetrator entered through a window, through communal doors, and so forth. Including entry method metadata information about certain crime incidents allows separate crime incidents to be linked or associated together by entry method. More generally, crime incident metadata for a crime incident can indicate a primary crime type (e.g., burglary) and one or more sub-types or characteristics of the primary crime type (e.g., entry method). This allows separate crime incidents of the same primary type to be linked or associated together by their common sub-types or characteristics.

**[0053]** Crime incident metadata for a crime incident can also include free-form text provided by a reporting law enforcement officer. Such text may be a short description of the crime incident in the law enforcement officer's own words. Keywords may be extracted from the text and used to determine a location for the crime incident where no specific location is provided. For example, the text entered by the law enforcement officer may state that an incident occurred "in Appletown somewhere along First Street between Maple and Elm Avenues". Based on this text description, an approximate street address can be determined and fed to a geocoding application to obtain approximate geographic coordinates or a line or area where the incident occurred. The returned geographic coordinates or line or area can then be resolved to a grid cell(s) in a region as described above.

#### Custody Data

**[0054]** In addition to historical crime incident data, generation of a crime risk forecast may be based on custody data. The custody data may indicate persons apprehended or placed in law enforcement custody that are suspected to have or known to have committed certain crime incidents. Such custody information may be used to mathematically de-emphasize the significance of these crime incidents when generating crime risk forecasts based on these crime incidents on the theory that such crime incidents are less likely to occur again now that the perpetrators of those crime incidents are in custody.

**[0055]** The custody data may also indicate when and where persons will be released from custody. Based on temporal and spatial custody release information, the significance of historical crime incidents in or near a target area where persons will be released from custody may be mathematically emphasized when generating a crime risk forecast for the target area based on these crime incidents on the theory that the persons released from custody have a significant probability of re-offending in or near the area they are released at. For example, when generating the crime risk forecast for the target area, historical crime incidents of the type perpetrated by the persons being released from custody in or near the target area may be mathematically emphasized

on the theory that the persons released are more likely to re-commit the same types of crimes as opposed to other types of crimes.

#### Computer-Aided Dispatch (CAD) Data

**[0056]** In addition to custody data, generation of a crime risk forecast may be based on computer-aided dispatch (CAD) data. The CAD data may be obtained from a computer-aided dispatch (CAD) system or other computer system for dispatching public emergency personnel to respond to reported incidents. CAD data obtained for an incident may be used to create or supplement the historical crime incident data for the incident. More specifically, information in the CAD data may be used to generate or populate the crime type, crime date/time, and/or the crime location metadata for the incident.

#### Prospective Hotspotting

**[0057]** Per an embodiment of the present invention, a prospective hotspotting algorithm is used to generate a crime risk forecast for a target area. To do so, historical crime incidents in or near the target area are analyzed. A set of possible historical crime incidents analyzed for generating the forecast can be circumscribed by time and space threshold parameters. More specifically, only recent crime incidents within a specified geographic distance (e.g., a radial distance) of the target area may be analyzed. Here, recent crime incidents can be established by identifying nearby crime incidents that occurred, for example, after a specified time in the past. For example, the set of crime incidents used for generating the forecast can include crime incidents that occurred in the past week.

**[0058]** Per an embodiment of the prospective hotspotting technique for generating a crime risk forecast for a target area, the crime risk for the target area is a weighted sum over all previous crime incidents within a space and time threshold. Each previous crime incident within the space and time thresholds can be weighted using a decay function that considers how far in space and in time the crime incident is from the target area. In addition to the decay function, a spatiotemporal threshold can be applied so that, for example, only those events occurring within a 1 km geographic radial distance from the center of the target area and four (4) weeks of the current date/time are included in the calculation.

**[0059]** For example, the crime risk for a grid cell at time can be computed from a sum of a set of previous crime incidents:

$$A_{j,k}(t) = \sum_{i: t-T_i < T, \Delta S < S} \frac{1}{1 + \Delta T} \frac{1}{1 + \Delta S}$$

**[0060]** Here,  $A_{j,k}(t)$  is the crime risk in grid cell (j, k) in a geographic region at time t. T and S are the time and space threshold cut-off parameters, respectively.  $\Delta T$  is the number of days between  $T_i$  and t divided by T/10 where  $T_i$  is the time of occurrence of one of n number of previous crime incidents over which the sum is computed.  $\Delta S$  is the geographic distance between a point (e.g., the centroid) of the geographic area covered by the grid cell (j, k) and a geographic location of the previous crime incident  $S_i$ .

[0061] As can be seen from the above equation, only previous crime incidents with locations and times within the time and space threshold cut-off parameters T and S, respectively, are included in the sum.

[0062] The prospective hotspotting technique can generate a more accurate crime risk forecast for a target area because it incorporates previous crime incidents for nearby areas and because it discounts previous nearby crime incidents as a function of their spatial and temporal distance from current forecast.

[0063] When generating a crime risk forecast for a particular crime type and a particular time window in addition to generating the crime risk forecast for the target area, the previous crime incidents within the time and space thresholds that are included in the sum can be limited to only those previous crime incidents that are of one or more particular crime types and that occurred within one or more particular time windows. For example, when generating a crime risk forecast for burglary during the late shift on a future date, the previous crime incidents can be limited to previous burglaries that occurred during the late shift within the space and time thresholds.

[0064] While in the above example the prospective hotspotting technique space and time thresholds are used to limit the previous crime incidents considered in the sum, a decay function can be used to limit the previous crime incidents included in the sum in other embodiments. For example, an exponential decay function can be used to limit the previous crime incidents included in the sum.

#### Histogram

[0065] Per an embodiment, a crime risk forecast generated for a target area is computed as the sum of previous crime incidents in the target area. For example, the crime risk for a grid cell at a time can be computed from a sum of a set of previous crime incidents:

$$A_{j,k}(t) = \sum_{i: T_i < t} I(J_i = j, K_i = k)$$

[0066] The above-equation effectively functions as a simple histogram estimator.  $A_{j,k}(t)$  is the crime risk in grid cell (j, k) in a geographic region at time t. The symbol I represents an indicator function that returns a positive numerical value (e.g., one) or zero for each previous crime incident in the set of previous crime incidents  $T_i < t$ , where  $T_i$  is the time of the ith previous crime incident in the set. The indicator function I returns a positive numerical value if the location of the ith previous crime incident ( $J_i, K_i$ ) was in the same grid cell (j,k) as the grid cell of the target area. Otherwise, the indicator function returns I zero. Thus, only previous crime incidents that occurred in the target area are considered in the histogram approach.

[0067] Like with the prospective hotspotting technique, the previous crime incidents included in the sum can be limited to those of a crime type and corresponding to a time window.

#### Machine Learning Features

[0068] In one embodiment, logistic regression is used to generate a crime risk forecast from machine learning features. In one embodiment, the output of the logistic regres-

sion approach for a given target grid cell is whether that grid cell will or will not experience a crime at time t. This can be modeled as a random variable  $C_{j,k,t} = \{0, 1\}$ , where j, k, and t are space-time grid coordinates. In an embodiment, time t is discretized to a date.

[0069] Machine learning features can be built as a function f of previous crime incidents and the discretized space-time coordinates of the target grid cell. For example, features  $f_j(j, k, t, J_m, K_m, L_m)$  for time  $L_m < t$  can be built.

[0070] For the regression of the random variable onto the features, dates to be predicated can be excluded from the regression.

[0071] In an embodiment, built features can be static features. For example, built features may be sparse vectors encoding the cell position (j, k). As another example, built features can be historical features that count previous crimes over a time window.

[0072] Instead of logistic regression, any other machine learning algorithm for example a support vector machine or feed forward neural network may be used to learn the effects of previous crime incidents.

[0073] The machine learning algorithm may also include different input such as historical weather information associated with the previous crime incidents. Other inputs may include computer-aided dispatch data, arrest data, and law enforcement patrol activity data.

[0074] To improve the performance of machine learning algorithms discussed herein, various features are extracted from the underlying data and used as inputs to the machine learning algorithms.

[0075] An example of a feature extraction method is a non-linear transformation in time and space, for example “distance to the nearest police station” or “square of the time elapsed since pub closing time”.

[0076] Another example of a feature extraction method is known as “spatiotemporal cuboids”. Previous crime incidents are projected onto a space-time “cuboid”. Machine learning features are then built from the cuboid for each cell on each date in a date set. An example of one of the machine learning features that could be built is: how many crimes (or crimes of a crime type and during a corresponding time window) occurred in the past 7 days within a 3 by 3 grid-cell neighborhood of the grid cell corresponding to the target area.

[0077] Another example of a feature extraction method is seasonal decomposition. Crime events are normalized per seasonal patterns on various time scales (day of the week, month of the year etc.) to isolate the periodic component and reveal the underlying trend in the data.

#### Ensemble Method

[0078] In some embodiments, a forecast is generated from a weighted combination of forecasts produced by multiple algorithms. For example, a forecast for a target area for a crime type for a time window can be generated from a weighted combination of a) a first forecast for the target area for the crime type for the time window produced by the prospective hotspotting approach described above and b) a second forecast for the target area for the crime type for the time window generated by the histogram technique described above. Weights for each algorithm can be determined by comparing forecasts produced by the algorithm to the actual crime activity that occurred. Algorithms that produce more accurate forecasts receive greater weighting in

the combination and algorithms that produce less accurate forecasts receive less weighting in the combination.

[0079] In one ensemble method approach, outputs of multiple criminological algorithms are treated as machine learning features. For example, the outputs of the prospective hotspotting and histogram approaches described above can be treated as machine learning features. The outputs of the multiple algorithms are then provided as input to a machine learning algorithm, potentially along with other machine learning feature inputs such as those described in the following section. The machine learning algorithm can be a feed forward neural network with back propagation or logistic regression, as just some examples.

[0080] A function of the ensemble machine learning algorithm is to decide as to which algorithm(s) to use for the prediction using the input data as a guide. The input algorithms may have varying time scales, area, parameters and thus the machine learning step is effectively choosing how far back in time to look, how wide a geographic area to consider, etc. Where a neural network machine learning algorithm is used, this decision surface can be non-linear and thus potentially account for crime patterns varying over time, space, season, weather, patrol levels, custody release levels, etc. This approach in effect fuses the advantages of machine learning and automated pattern recognition with the more intuitive comprehensibility of the criminological approaches.

[0081] A benefit of the ensemble method is that it provides a way to more accurately account for the trade-off faced with the criminological approaches between capturing areas of long-term elevated risk and finding short-term fluctuations. For example, the ensemble method can more accurately decide when to pick a criminological algorithm with a longer time window.

[0082] In one embodiment, the training of the machine learning algorithm used in the ensemble method is restricted to a subset of the training examples. For example, the training set could be restricted to geographic areas of only high inherent risk. This would then bias the ensemble method forecast to weight towards algorithms that work well in these circumstances.

#### Web-Based Geographic Crime Risk Forecasting Computer System

[0083] FIG. 3 is a block diagram of an exemplary web-based geographic crime risk forecasting computer system 300. System 300 includes a client computing device 302 operatively coupled to crime risk forecasting system 304 by data network 306. The crime risk forecasting system 304 includes a display component 308 and a forecasting component 310.

[0084] Client computing device 302 executes a web browser 312. The web browser 312 may in turn execute a portion of display component 308.

[0085] Web browser 312 may be a conventional web browser offering a well-known web browser platform such as MICROSOFT INTERNET EXPLORER, MOZILLA FIREFOX, GOOGLE CHROME, and so forth.

[0086] The portion of display component 308 executed by web browser 312 may be implemented in standard client-side web application technologies such as HyperText Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript (JS), and so forth.

[0087] The portion of display component 308 executed by web browser 312 may be delivered to client 302 for execution by web browser 312 over network 306 by the display component 308 of crime risk forecasting system 304. Such delivery may be made per a standard Internet networking protocol such as the Transmission Control Protocol/Internet Protocol (IP), the HyperText Transfer Protocol (HTTP), and so forth.

[0088] The portion of display component 308 executed by web browser 312 is configured to retrieve over network 306 interactive geospatial basemaps from geospatial application 314 and visually overlay the retrieved basemaps with generated visual manifestations of crime risk forecasts generated by forecasting component 310. The forecasts generated by forecasting component 310 may also be sent to client 302 using standard Internet networking protocol(s). The basemaps with the forecast overlays are then presented in a window of the web browser 312 to a user of client 302.

[0089] Database 316 stores generated forecast information and information from which crime risk forecasts are generated such as region and grid cell geographic information, historical crime incident data, custody data, dispatch data, and so forth.

[0090] Geospatial application 314 can be any geospatial application capable of providing interactive geospatial basemaps such as those disclosed in related U.S. patent application Ser. No. 13/917,571, filed Jun. 13, 2013, entitled "Interactive Geospatial Map", which is incorporated by reference herein in its entirety. Other possible geospatial applications that could be used as geospatial application 314 for providing interactive geospatial basemaps include Internet map servers that support Web Map Service (WMS) standards developed by the Open Geospatial Consortium such as ESRI ARCGIS SERVER, GEOSERVER, ARCGIS.COM, GOOGLE MAPS API, MICROSOFT BING MAPS, and so forth.

#### Implementing Computing Device

[0091] In some embodiments, the techniques disclosed herein are implemented on one or more computing devices. For example, FIG. 4 is a block diagram that illustrates a computing device 400 in which some embodiments of the present invention may be embodied. Computing device 400 includes a bus 402 or other communication mechanism for communicating information, and a hardware processor 404 coupled with bus 402 for processing information. Hardware processor 404 may be, for example, a general purpose microprocessor or a system on a chip (SoC).

[0092] Computing device 400 also includes a main memory 406, such as a random access memory (RAM) or other dynamic storage device, coupled to bus 402 for storing information and instructions to be executed by processor 404. Main memory 406 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 404. Such instructions, when stored in non-transitory storage media accessible to processor 404, render computing device 400 into a special-purpose machine that is customized to perform the operations specified in the instructions.

[0093] Computing device 400 further includes a read-only memory (ROM) 408 or other static storage device coupled to bus 402 for storing static information and instructions for processor 404.

[0094] A storage device 410, such as a magnetic disk, optical disk, or solid-state drive is provided and coupled to bus 402 for storing information and instructions.

[0095] Computing device 400 may be coupled via bus 402 to a display 412, such as a liquid crystal display (LCD) or other electronic visual display, for displaying information to a computer user. Display 412 may also be a touch-sensitive display for communicating touch gesture (e.g., finger or stylus) input to processor 404.

[0096] An input device 414, including alphanumeric and other keys, is coupled to bus 402 for communicating information and command selections to processor 404.

[0097] Another type of user input device is cursor control 416, such as a mouse, a trackball, or cursor direction keys for communicating direction information and command selections to processor 404 and for controlling cursor movement on display 412. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to specify positions in a plane.

[0098] Computing device 400 may implement the techniques described herein using customized hard-wired logic, one or more application-specific integrated circuits (ASICs), one or more field-programmable gate arrays (FPGAs), firmware, or program logic which, in combination with the computing device, causes or programs computing device 400 to be a special-purpose machine. Per some embodiments, the techniques herein are performed by computing device 400 in response to processor 404 executing one or more sequences of one or more instructions contained in main memory 406. Such instructions may be read into main memory 406 from another storage medium, such as storage device 410. Execution of the sequences of instructions contained in main memory 406 causes processor 404 to perform the process steps described herein. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions.

[0099] The term “storage media” as used herein refers to any non-transitory media that store data and/or instructions that cause a machine to operate in a specific fashion. Such storage media may comprise non-volatile media and/or volatile media. Non-volatile media includes, for example, optical disks, magnetic disks, or solid-state drives, such as storage device 410. Volatile media includes dynamic memory, such as main memory 406. Common forms of storage media include, for example, a floppy disk, a flexible disk, hard disk, solid-state drive, magnetic tape, or any other magnetic data storage medium, a CD-ROM, any other optical data storage medium, any physical medium with patterns of holes, a RAM, a PROM, and EPROM, a FLASH-EPROM, NVRAM, any other memory chip or cartridge.

[0100] Storage media is distinct from but may be used in conjunction with transmission media. Transmission media participates in transferring information between storage media. For example, transmission media includes coaxial cables, copper wire and fiber optics, including the wires that comprise bus 402. Transmission media can also take the form of acoustic or light waves, such as those generated during radio-wave and infra-red data communications.

[0101] Various forms of media may be involved in carrying one or more sequences of one or more instructions to processor 404 for execution. For example, the instructions may initially be carried on a magnetic disk or solid-state drive of a remote computer. The remote computer can load

the instructions into its dynamic memory and send the instructions over a telephone line using a modem. A modem local to computing device 400 can receive the data on the telephone line and use an infra-red transmitter to convert the data to an infra-red signal. An infra-red detector can receive the data carried in the infra-red signal and appropriate circuitry can place the data on bus 402. Bus 402 carries the data to main memory 406, from which processor 404 retrieves and executes the instructions. The instructions received by main memory 406 may optionally be stored on storage device 410 either before or after execution by processor 404.

[0102] Computing device 400 also includes a communication interface 418 coupled to bus 402. Communication interface 418 provides a two-way data communication coupling to a network link 420 that is connected to a local network 422. For example, communication interface 418 may be an integrated services digital network (ISDN) card, cable modem, satellite modem, or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, communication interface 418 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN. Wireless links may also be implemented. In any such implementation, communication interface 418 sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

[0103] Network link 420 typically provides data communication through one or more networks to other data devices. For example, network link 420 may provide a connection through local network 422 to a host computer 424 or to data equipment operated by an Internet Service Provider (ISP) 426. ISP 426 in turn provides data communication services through the world-wide packet data communication network now commonly referred to as the “Internet” 428. Local network 422 and Internet 428 both use electrical, electromagnetic or optical signals that carry digital data streams. The signals through the various networks and the signals on network link 420 and through communication interface 418, which carry the digital data to and from computing device 400, are example forms of transmission media.

[0104] Computing device 400 can send messages and receive data, including program code, through the network (s), network link 420 and communication interface 418. In the Internet example, a server 430 might transmit a requested code for an application program through Internet 428, ISP 426, local network 422 and communication interface 418.

[0105] The received code may be executed by processor 404 as it is received, and/or stored in storage device 410, or other non-volatile storage for later execution.

[0106] A software system is typically provided for controlling the operation of computing device 400. The software system, which is usually stored in main memory 406 and on fixed storage (e.g., hard disk) 410, includes a kernel or operating system (OS) which manages low-level aspects of computer operation, including managing execution of processes, memory allocation, file and network input and output (I/O), and device I/O. The OS can be provided by a conventional operating system such as, for example, MICROSOFT WINDOWS, SUN SOLARIS, or LINUX.

[0107] One or more application(s), such as client software or “programs” or set of processor-executable instructions, may also be provided for execution by computer 400. The

application(s) may be “loaded” into main memory 406 from storage 410 or may be downloaded from a network location (e.g., an Internet web server). A graphical user interface (GUI) is typically provided for receiving user commands and data in a graphical (e.g., “point-and-click” or “touch gesture”) fashion. These inputs, in turn, may be acted upon by the computing device in accordance with instructions from the OS and/or application(s). The graphical user interface also serves to display the results of operation from the OS and application(s).

#### Extensions and Alternatives

[0108] The foregoing description, for purpose of explanation, has been described regarding specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms described. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described to best explain the principles of the invention and its practical applications, to thereby enable other skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the use contemplated.

1. A computing system, comprising:

one or more processors;

storage media;

one or more programs stored in the storage media and configured for execution by the one or more processors, the one or more programs comprising instructions configured for:

storing crime incident data reflecting crime incidents that occur over a period of time;

based, at least in part, on the crime incident data, predicting a crime risk for a geographic area, a time window, and a crime type;

wherein the time window is in the future relative to the period of time over which the crime incidents occur; generating and displaying a graphical user interface overlay on an interactive geospatial basemap;

wherein the interactive geospatial basemap is generated by a geospatial application; and

wherein the graphical user interface overlay visually indicates at least all of:

the geographic area,

the time window,

the crime type,

the crime risk, and

a number of the crime incidents that at least occur in the geographic area over the period of time during a periodically recurring continuous period of time.

2. The system of claim 1, wherein:

the periodically recurring continuous period of time is a day of the week, the week having seven days of which the day is one;

the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time is a number of the crime incidents that at least occur in the geographic area over the period of time on the day of the week.

3. The system of claim 1, wherein:

the periodically recurring continuous period of time is an hour of the day, the day having 24 hours of which the hour is one;

the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time is a number of the crime incidents that at least occur in the geographic area over the period of time on the hour of the day.

4. The system of claim 1, wherein each of the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time are of the crime type.

5. The system of claim 1, wherein one or more of the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time are not of the crime type.

6. The system of claim 1, wherein:

the time window corresponds to a law enforcement patrol shift on a particular day; and

the graphical user interface overlay visually indicates the particular day and a time of the law enforcement patrol shift on the particular day.

7. tem of claim 1, wherein the instructions are further configured for:

generating a weighted sum of a plurality of crime incidents, of the crime incidents, within a space threshold and a time threshold; and

wherein the predicting the crime risk is based, at least in part, on the weighted sum.

8. The system of claim 1, wherein the instructions are further configured for:

generating a sum of a plurality of crime incidents, of the crime incidents, that occurred in the geographic area; and

wherein the predicting the crime risk is based, at least in part, on the sum.

9. The system of claim 1, wherein the instructions are further configured for:

obtaining global positioning system (GPS) information from radio equipment used by law enforcement officers patrolling at least the geographic area;

determining a level of law enforcement patrol presence for at least the geographic area based, at least in part, on the GPS information; and

wherein the predicting the crime risk is based, at least in part, on the level of law enforcement patrol presence for at least the geographic area.

10. The system of claim 1, wherein the instructions are further configured for:

obtaining computer-aided dispatch data reflecting crime type, crime date, and crime location of one or more of the crime incidents;

wherein the crime incident data comprises the computer-aided dispatch data; and

wherein the predicting the crime risk is based, at least in part, on the computer-aided dispatch data.

11. A method performed by a computing system comprising one or more processors and storage media, the method comprising:

storing crime incident data reflecting crime incidents that occur over a period of time;

based, at least in part, on the crime incident data, predicting a crime risk for a geographic area, a time window, and a crime type;

wherein the time window is in the future relative to the period of time over which the crime incidents occur; generating and displaying a graphical user interface overlay on an interactive geospatial basemap;

wherein the interactive geospatial basemap is generated by a geospatial application; and

wherein the graphical user interface overlay visually indicates at least all of:

the geographic area,

the time window,

the crime type,

the crime risk, and

a number of the crime incidents that at least occur in the geographic area over the period of time during a periodically recurring continuous period of time.

**12.** The method of claim **11**, wherein:

the periodically recurring continuous period of time is a day of the week, the week having seven days of which the day is one;

the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time is a number of the crime incidents that at least occur in the geographic area over the period of time on the day of the week.

**13.** The method of claim **11**, wherein:

the periodically recurring continuous period of time is an hour of the day, the day having 24 hours of which the hour is one;

the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time is a number of the crime incidents that at least occur in the geographic area over the period of time on the hour of the day.

**14.** The method of claim **11**, wherein each of the number of the crime incidents that at least occur in the geographic area over the period of time during the periodically recurring continuous period of time are of the crime type.

**15.** The method of claim **11**, wherein one or more of the number of the crime incidents that at least occur in the

geographic area over the period of time during the periodically recurring continuous period of time are not of the crime type.

**16.** The method of claim **11**, wherein:

the time window corresponds to a law enforcement patrol shift on a particular day; and

the graphical user interface overlay visually indicates the particular day and a time of the law enforcement patrol shift on the particular day.

**17.** The method of claim **11**, further comprising:

generating a weighted sum of a plurality of crime incidents, of the crime incidents, within a space threshold and a time threshold; and

wherein the predicting the crime risk is based, at least in part, on the weighted sum.

**18.** The method of claim **11**, further comprising:

generating a sum of a plurality of crime incidents, of the crime incidents, that occurred in the geographic area; and

wherein the predicting the crime risk is based, at least in part, on the sum.

**19.** The method of claim **11**, further comprising:

obtaining global positioning system (GPS) information from radio equipment used by law enforcement officers patrolling at least the geographic area;

determining a level of law enforcement patrol presence for at least the geographic area based, at least in part, on the GPS information; and

wherein the predicting the crime risk is based, at least in part, on the level of law enforcement patrol presence for at least the geographic area.

**20.** The method of claim **11**, wherein the instructions are further configured for:

obtaining computer-aided dispatch data reflecting crime type, crime date, and crime location of one or more of the crime incidents;

wherein the crime incident data comprises the computer-aided dispatch data; and

wherein the predicting the crime risk is based, at least in part, on the computer-aided dispatch data.

\* \* \* \* \*