

VISUALIZING CAMERA FEEDS ON A MAP

BACKGROUND

5

[0001] Traffic camera feeds deliver video images to users over the Internet. Such feeds are becoming both abundant and familiar to Internet users.

10

[0002] In one implementation, a user is presented with an array of available camera feeds for a geographic region of interest. Each feed may be represented as a thumbnail image so that multiple feeds can be seen at the same time. For easier viewing, the user can select and enlarge a thumbnail image.

15

[0003] In another implementation, a relatively conventional map (e.g., a street map) is displayed, in which the streets and highways are represented as lines. The map also includes icons that show the locations of traffic cameras. The user can click on one of the icons to select a traffic camera and display its feed.

20

[0004] The formats described above are useful to a limited extent, but they also have shortcomings that prevent their full potential from being realized. For example, it may be difficult for a viewer to understand which direction a camera is facing or which street the camera is recording. Compounding this problem, the orientation of the camera may change dynamically over the course of a day. Also, feeds from different cameras may be captured at different times, making it difficult for a viewer to assimilate the information from each of the feeds. These problems become more acute as the number of traffic cameras increases. For example, in some cities, streets are monitored using thousands of cameras.

25
30

SUMMARY

[0005] Feeds from cameras are better visualized by superimposing images based on the feeds onto a map based on a two- or three-dimensional virtual map.

[0006] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the
5 claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The accompanying drawings, which are incorporated in and form a part of this specification, illustrate embodiments and, together with the description,
10 serve to explain the principles of the embodiments:

[0008] Figure 1 is a block diagram of one embodiment of a video visualization system.

15 [0009] Figure 2 illustrates a video feed overlaid onto a virtual globe map.

[0010] Figure 3 illustrates a billboard overlaid onto a virtual globe map.

[0011] Figures 4A, 4B and 4C illustrate the extraction of certain objects
20 from a video feed.

[0012] Figure 5 is a flowchart of one embodiment of a method for visualizing a video feed.

25 [0013] Figure 6 is a flowchart of one embodiment of a method for calibrating a video camera.

DETAILED DESCRIPTION

[0014] Some portions of the detailed descriptions which follow are presented in terms of procedures, logic blocks, processing and other symbolic representations of operations on data bits within a computer memory. These descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. In the present application, a procedure, logic block, process, or the like, is conceived to be a self-consistent sequence of steps or instructions leading to a desired result. The steps are those requiring physical manipulations of physical quantities. Usually, although not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated in a computer system.

[0015] It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the following discussions, it is appreciated that throughout the present application, discussions utilizing the terms such as “accessing,” “superimposing,” “rendering,” “aligning,” “projecting,” “correlating,” “overlaying,” “simulating,” “calibrating,” “displaying,” “receiving” or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system’s registers and memories into other data similarly represented as physical quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0016] Embodiments described herein may be discussed in the general context of computer-executable instructions residing on some form of computer-usable medium, such as program modules, executed by one or more computers or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc., that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or distributed as desired in various embodiments.

[0017] By way of example, and not limitation, computer-usable media may comprise computer storage media and communication media. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, random access memory (RAM), read only memory (ROM), electrically erasable programmable ROM (EEPROM), flash memory or other memory technology, compact disk ROM (CD-ROM), digital versatile disks (DVDs) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium that can be used to store the desired information.

[0018] Communication media can embody computer-readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency (RF), infrared and other wireless media. Combinations of any of the above should also be included within the scope of computer-readable media.

[0019] According to embodiments described herein, one or more camera feeds, such as traffic camera feeds, can be visualized in a single map view. Camera feeds can include still or video images. In a sense, a video consists of a succession of still images. Therefore, although the discussion below focuses primarily on video feeds, that discussion can be readily extended to include cases that involve one or more still images.

[0020] The map can be a two-dimensional (2D) or three-dimensional (3D) map referred to herein as a "virtual globe" or "virtual world map" or, most generally, as a "virtual map." A virtual globe map may include man-made features, such as roads and buildings, as well as natural features. An example of

a virtual globe map is Virtual Earth 3D™ by Microsoft®. In general, a virtual globe map is a 3D software model or representation of the Earth. That is, the 3D model possesses 3D data about the natural terrain and any man-made features. A virtual globe map may incorporate a digital elevation map (DEM). Virtual maps
5 can also be produced for smaller scale structures such as rooms and building interiors.

[0021] Although display screens are two-dimensional surfaces, a map rendered in 3D allows a viewer to change viewing angle and position with limited
10 distortion. The discussion below focuses primarily on 3D maps. A 2D rendering is simply a special instance of a 3D rendering, and so the discussion can be readily extended to 2D maps.

[0022] The discussion below uses traffic camera feeds as a prime example
15 – as will be described, traffic camera feeds can be integrated into a virtual map so that the movement of cars and the like are visualized against the backdrop of the virtual map. However, embodiments described herein are not so limited and may be used in applications other than monitoring traffic. For example, the discussion below can be readily extended to visualizing the movement of pedestrians along
20 pavements or within buildings, or the movement of objects and people on the production floor of a factory.

[0023] Visualizing a camera feed on a virtual map entails aligning the feed with the features of the map. For example, to visualize a feed from a traffic
25 camera on a map, the traffic is aligned with the road layout of the map. Visualizing a camera feed on a virtual map provides context about where the camera is pointing and what road the camera is covering. When multiple feeds are simultaneously visualized, the feeds can also be aligned in time. Presenting the feeds in a single view allows a user to more quickly comprehend the coverage
30 provided by each camera feed. A user can also comprehend the relationships between camera feeds, allowing the user to infer causes and effects and to interpolate information for areas not covered by neighboring cameras. For example, if one camera shows a traffic accident and a neighboring camera shows

a traffic backup, then the user can infer that the backup is due to the accident, and that the stretch of roadway between cameras is also backed up.

[0024] Figure 1 is a block diagram of one embodiment of a video
5 visualization system 100 that achieves the above. Aspects of the system 100 can be implemented as computer-executable components stored in a computer-readable medium.

[0025] In the example of Figure 1, the system 100 includes a video capture
10 component 110, a video imaging component 120, a virtual map modeling component 130, and a video visualization component 140. While these components are represented as separate elements, they may not be implemented as such. In general, the system 100 provides functionality about to be described, and that functionality can be accomplished using a single functional element or
15 multiple functional elements, on a single device or computer system or distributed over multiple devices/computer systems.

[0026] The video capture component 110 receives and processes a feed
from one or more video cameras (not shown). For example, the video capture
20 component 110 may digitize and/or encode (compress) video images captured by the video cameras. A video capture component 110 may be implemented on each of the video cameras.

[0027] In the example of Figure 1, the video imaging component 120
25 receives and processes (e.g., decodes) the feed from the video capture component 110. The video imaging component 120 can perform other types of processing; refer to the discussion below, such as the discussion of Figures 4A-4C. for examples of some of the other types of processing that may be performed.

[0028] Continuing with reference to Figure 1, the virtual map modeling
30 component 130 generates a 2D or 3D virtual map. The video visualization component 140 combines the output of the video imaging component 120 and the virtual map modeling component 130 and renders the result. Generally speaking, the video visualization component 140 integrates the feeds from one or more

cameras with a 2D or 3D map that is based on the virtual map and renders the resultant integrated feed(s)/map. In other words, the video visualization component 140 cuts-and-pastes frames of video into a map derived from a 3D virtual map.

5

[0029] To properly visualize each feed on a map, each camera is aligned with the map. A camera can be aligned to the map by aligning features in its video feed with corresponding features in the map. For example, a traffic camera feed generally includes vehicles moving on a roadway; to align the traffic camera and the map, the roadway in the video is mapped to and aligned with the corresponding roadway in the map. Consequently, the vehicles contained in the video are also properly placed and oriented within the map.

[0030] Alignment of a camera and map can be manual, automatic, or a combination of both. Fully manual alignment involves specifying corresponding points in the video and the map. The more points specified, the better the alignment. The correspondence enables computation of a warping function that maps the frames of a video onto the correct portion of the map. Fully automatic alignment can be implemented by matching features in the video to features in the map and then aligning those features.

[0031] A combination of manual and automatic alignment techniques can be used to realize the advantages of both types of techniques while avoiding some of their shortcomings. First, corresponding points can be manually specified to calibrate the internal parameters of the camera (e.g., focal length, principal point, and skew) and the initial position and orientation of the camera. Second, features in the video can be computed using low-level vision techniques and then used to automatically realign the camera as needed. Additional information is provided in conjunction with Figure 6. A combination of manual and automatic techniques avoids some of the tediousness and human error associated with a fully manually technique; instead, manual actions are limited to the initial alignment as subsequent alignments are performed automatically. A combination of techniques is also more robust than a fully automatic technique. For example, if a particular feature in a video is being relied upon for alignment, but that feature

is missing from the map, then the attempt at alignment would likely fail without human intervention.

[0032] After initial alignment of a camera, the camera may be moved either intentionally or because of effects such as the wind. If the camera moves, the correspondence between points in the video and in the map may again be manually specified to compensate. Alternatively, the camera can be automatically realigned as mentioned above.

[0033] Figure 2 illustrates an image (e.g., video) frame 210 overlaid onto a region 220 of a virtual globe map 200. Although objects (e.g., the vehicles, trees, building and roadway) are illustrated in cartoon-like fashion in Figure 2, in actual practice those objects are more realistically rendered. In actual practice, the quality of the video determines the quality of the objects within the frame 210, while the objects outside the frame 210 are generally of photographic quality, as in Virtual Earth 3D™, for example.

[0034] Utilizing a video visualization system such as system 100 (Figure 1), the frame 210 is captured by a camera (e.g., a traffic camera) and integrated into the map 200. Frame 210 represents an image of the area that is within the camera's field of vision at a particular point in time. As shown in the example of Figure 2, the frame 210 is essentially embedded (overlaid or superimposed) into the map 200. That is, the frame 210 is pasted into the region 220 of the map 200, while the remainder of the map 200 is used to fill in the areas around the region 220. As mentioned above, the camera and the map 200 can be calibrated so that each frame is properly placed and properly oriented within the map display.

[0035] Successive (video) frames can be overlaid onto the map 200 in a similar fashion such that a viewer would see the movement of traffic, for example, within the region 220. In essence, dynamic information (the video feed) is integrated within the static image that constitutes the map 200. The viewer is thereby provided with an accurate representation of, for example, actual traffic conditions.

[0036] Furthermore, if there are multiple cameras covering the geographic area encompassed by the displayed portion of the map 200, then frames from those cameras can be similarly overlaid onto corresponding regions of the map. Each of the cameras can be calibrated with the map 200 so that their respective
5 videos are properly placed and properly oriented within the map display.

[0037] Should the position of any camera change from its initial position, the camera can be recalibrated (realigned), as mentioned above. Consequently, its feed would be visualized at a different point in the map 200, so that the video
10 captured by the repositioned camera would still be properly aligned with the other features of the map. In other words, the region 220 is associated with a particular camera; if the camera is moved, then the region 220 would also be moved to match.

[0038] Camera feeds can be visualized in ways different from that just described. Another way to visualize a camera feed is to render the video as an animated texture on a "billboard" that is superimposed onto a virtual globe map. Figure 3 illustrates a billboard 310 overlaid onto a map 300. As discussed above, objects are illustrated in cartoon-like fashion in Figure 3, but in actual practice
20 those objects are more realistically rendered.

[0039] The billboard 310 is planar and rectangular (although it appears to be non-rectangular when viewed in perspective) and aligned to the features (e.g., the roadway) of map 300 in the manner described above. Each texture in the
25 animation plays a frame from the video so that the effect is that traffic, for example, appears to move along the billboard 310. As in the example of Figure 2, alignment of the billboard 310 with the map 300 is relatively straightforward, and the viewer is provided with an accurate representation of, for example, actual traffic conditions.

30

[0040] An alternative to the above is to use a digital elevation map (DEM) that includes a road layout, to provide a 3D effect that otherwise may not be apparent with the use of a billboard. In a sense, a camera can be thought of as a slide projector that projects its "slides" (frames) onto the DEM, creating the effect

that objects in the video (e.g., traffic) are following the contours of the DEM. The resultant display would be similar to that shown by Figure 2. Another advantage to the use of a DEM is that, if the camera changes direction, the video feed can still be visualized with limited distortion.

5

[0041] The feed from a video camera includes information that may be of limited interest to a user. For example, a traffic camera may capture details of the environment surrounding a roadway when a viewer is only interested in the volume of traffic on the roadway. To address this, the background can be removed from the video feed so that, for example, only the traffic remains. One way to accomplish this is to extract only those objects in the video that are moving. To do this, the median value for each pixel in the video is determined over a period of time. Then, each per-pixel median value is subtracted from the corresponding pixel value in each frame of the video to produce an alpha matte that shows only the moving objects.

[0042] Figures 4A, 4B and 4C illustrate the extraction of certain objects from a video feed. As in the examples above, objects are illustrated in cartoon-like fashion, but in actual practice those objects can be more realistically rendered. Figure 4A illustrates a median image 400. Figure 4B shows a typical frame 410. Figure 4C shows one frame of an alpha-matted image 420. The alpha-matted image 420 can then be superimposed onto a virtual globe map or DEM in a manner similar to that described above. The resultant display would be similar to that shown by Figure 2.

25

[0043] Other approaches can be used to extract objects from a video feed. For example, a classifier can be trained to recognize cars and other types of traffic. The use of a classifier may increase computational complexity, but it also allows nonmoving objects (e.g., cars stopped in traffic) to be recognized and extracted from a video feed.

30

[0044] An extension to the approach just described is to replace an extracted object with an abstraction or simulation of that object. For example, instead of rendering actual images of vehicles from a traffic feed as described in

conjunction with Figures 4A-4C, a simulated vehicle can be displayed on the map. In essence, the vehicles in the alpha-matted display 420 are replaced with synthetic representations of vehicles. More specifically, the location and speed of each of the objects in the alpha-matted display 420 can be tracked from frame to frame to compute the position of each object versus time. Only the position of each object needs to be sent from the video imaging component 120 to the video visualization component 140 (Figure 1), reducing bandwidth requirements. The resultant display would be similar to that shown by Figure 2 or Figure 3, but now the objects within the region 220 or the billboard 310 would perhaps be cartoon-like, although more detailed simulations can be used.

[0045] In addition to reducing bandwidth requirements, the use of simulated objects has a number of other advantages. The tracking of each object's position versus time allows the position of the object to be extrapolated to areas outside the field of vision of the camera. Also, the rendering of the objects is standardized (simulated), and therefore it is independent of the camera resolution, lighting conditions, the distance of the object from the camera, and other influences. In addition, the visualization of the camera feed is more tolerant of noise resulting from misalignment of the camera.

[0046] The maps, including video feed(s), may be part of a larger display. For example, the map/feed visualizations may be accompanied by text, including symbols, numbers and other images, that provide additional information such as statistics, traffic summaries, weather conditions, links to other Web sites, and the like. Furthermore, such information may be superimposed onto the map. Also, such information can be added to the video feed before it is superimposed onto the map. In general, additional information may be presented in the display areas bordering the map or within the map itself.

[0047] Figure 5 is a flowchart 500 summarizing one embodiment of a method for visualizing a video feed. Figure 6 is a flowchart 600 of one embodiment of a method for calibrating a video camera to a virtual map. Although specific steps are disclosed in the flowcharts 500 and 600, such steps are exemplary. That is, various other steps or variations of the steps recited in the

flowcharts 500 and 600 can be performed. The steps in the flowcharts 500 and 600 may be performed in an order different than presented. Furthermore, the features of the various embodiments described by the flowcharts 500 and 600 can be used alone or in combination with each other. In one embodiment, the

5 flowcharts 500 and 600 can be implemented as computer-executable instructions stored in a computer-readable medium.

[0048] With reference first to Figure 5, in block 502, a user is presented with an array of available camera feeds. In block 504, the user selects one or

10 more of the camera feeds. In block 506, the selected feeds are accessed and received by, for example, the video visualization component 140 of Figure 1. In block 508, the video visualization component 140 also accesses a virtual map. In block 510, the selected video feeds are superimposed onto the virtual map and the results are displayed. More specifically, depending on the user's selections,

15 one or more feeds can be visualized simultaneously against a 2D or 3D background. Additional information is provided in the discussion of Figures 2, 3 and 4A-4C, above.

[0049] With reference now to Figure 6, in block 602, the video feed from the

20 camera is manually aligned with the 3D geometry of the map. Internal parameters of the camera (such as its focal length, principal point, and skew) as well as other parameters (such as the camera's orientation) are determined. For example, a user (not necessarily the end user, but perhaps a user such as a system administrator) can interactively calibrate ("geoposition") an image (e.g., a frame)

25 from the feed to a known 3D geometry associated with the map. In essence, the actual (real world) location of the camera is mapped to a location in the 3D map model.

[0050] In block 604, calibration and low-level vision techniques are used to

30 automatically update the camera's parameters, which may change due to, for example, wind effects or scheduled rotation of the camera. For a traffic camera feed, the low-level vision techniques are used to determine the general direction of traffic flow (a "traffic vector") shown in the feed. Associated with the 3D map model is metadata that identifies, for example, the direction of the road (a "road

vector”) being observed by the camera. The camera feed can be initially aligned with the roadway in the map by aligning the traffic vector and the road vector. As the camera moves, the incremental transform of the camera position that will align the road and traffic vectors can be computed, so that the traffic flow in the feed is
5 aligned with the roadway in the map.

[0051] For example, the camera may rotate 10 degrees every 10 minutes to produce a panorama of a roadway for which there is 3D road vector data available. The camera can be calibrated with respect to one frame at a given
10 time. As the camera rotates, the one unknown is its rotation angle. By computing the optical flow (traffic vector) in an image, the direction of the road in the image can also be determined (the direction of the road in the image corresponds to the traffic vector). When the 3D road vector data is projected onto the image, the misalignment of the predicted road vector and the optical flow – due to the
15 aforementioned rotation – is detected. The amount of camera rotation that best, or at least satisfactorily, aligns the predicted road vector and the measured optical flow can then be determined. When a camera has a known position but unknown orientation, this technique of finding the amount of camera rotation that aligns the predicted road vector and the measured optical flow can be used to automatically
20 find the camera’s orientation.

[0052] Camera calibration is facilitated if, for example, there are more than two roadways in the image, particularly if the roadways form a junction. Also, landmarks other than roadways can be utilized to perform the calibration. In
25 general, a known 3D geometry is corresponded to known 2D features (e.g., the optical flow of traffic in an image) to automatically update (realign) a camera’s position.

[0053] In summary, feeds from cameras are better visualized by
30 superimposing images based on the feeds onto a 2D or 3D map that, in turn, is based on a 3D virtual map. Consequently, the feeds retain their geographic context, making it easier for a user to understand which directions the cameras are pointing and which objects (e.g., streets) they are recording. Moreover, a user can more quickly understand the relation of one camera feed to another.

[0054] In the foregoing specification, embodiments have been described with reference to numerous specific details that may vary from implementation to implementation. Thus, the sole and exclusive indicator of what is the invention, and is intended by the applicant to be the invention, is the set of claims that issue from this application, in the specific form in which such claims issue, including any subsequent correction. Hence, no limitation, element, property, feature, advantage, or attribute that is not expressly recited in a claim should limit the scope of such claim in any way. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.

CLAIMS

What is claimed is:

1. A computer-readable medium (100) having computer-executable
5 components comprising:
 a virtual map model (130);
 a video image generator (120) coupled to said virtual map model and
 operable for generating first images based on a first video feed received from a
 first camera; and
10 a visualization model (140) coupled to said virtual map model and said
 video images generator, wherein said virtualization model is operable for
 overlaying said first images onto a map generated from said virtual map model.
2. The computer-readable medium of Claim 1 wherein said first video feed
15 comprises images of vehicles traveling on a roadway and wherein said map
 comprises an image of at least said roadway, wherein said visualization model is
 operable for said images of vehicles with said image of said roadway.
3. The computer-readable medium of Claim 1 wherein said video images
20 generator is further operable for generating second images based on a second
 video feed received from a second camera, wherein said visualization model is
 further operable for overlaying said second images onto said map, wherein said
 first images and said second images are aligned in time.
- 25 4. The computer-readable medium of Claim 1 wherein said first images are
 rendered as an animated texture on a billboard that is superimposed onto said
 map.
5. The computer-readable medium of Claim 1 wherein said map comprises
30 a digital elevation map (DEM) and wherein said first video feed is projected onto
 said DEM.
6. The computer-readable medium of Claim 1 wherein only moving objects
 from said first video feed are overlaid onto said map.

7. The computer-readable medium of Claim 1 wherein said first images comprise simulations of objects in said first video feed, said simulations used in place of actual images of said objects.

5

8. The computer-readable medium of Claim 1 wherein said computer-executable components further comprise a calibration model operable for calibrating said first camera using metadata associated with said map and information based on said first video feed.

10

9. In a computer system having a graphical user interface including a display and a user selection device, a method (500) of visualizing a video feed, said method comprising:

(502) displaying information that identifies a plurality of video camera feeds comprising a first video camera feed from a first video camera at a first location; (504) receiving a selection identifying said first video camera feed; and (510) displaying first images based on said first video camera feed, wherein said first images are superimposed on a virtual map that encompasses said first location.

20

10. The method of Claim 9 wherein said first video camera feed comprises images of vehicles traveling on a roadway and wherein said virtual map comprises an image of said roadway, wherein said method further comprises aligning said images of vehicles with said image of said roadway.

25

11. The method of Claim 9 further comprising:
receiving a selection identifying a second video camera feed from a second video camera at a second location; and
displaying second images based on said second video camera feed,
wherein said second images are superimposed on said virtual map, wherein said first images and said second images are aligned with respect to time.

30

12. The method of Claim 9 further comprising displaying text-based information concurrent with said displaying of said first images.

13. The method of Claim 9 wherein said map display comprises a digital elevation map (DEM) and wherein said method further comprises projecting said first feed onto said DEM.

5

14. The method of Claim 9 further comprising rendering only moving objects from said first feed onto said map display.

15. The method of Claim 9 wherein said first image comprises simulations
10 of objects in said first feed, said simulations used in place of actual pictures of said objects taken by said first camera.