US 20090231436A1

(54) **METHOD AND APPARATUS FOR TRACKING WITH IDENTIFICATION**

(76) Inventors: **Anthony E. Faltesek**, Roseville, MN (US); **Vassilios Morellas**, Plymouth, MN (US)

Correspondence Address:
KRIS FREDRICK
Honeywell Inc., Law Department
MN65-3500, 3660 Technology Drive
Minneapolis, MN 55418 (US)

(57) **ABSTRACT**

Combining and fusing the tracking of people and objects with image processing and the identification of the people and objects being tracked. Also, conditions of a person, object, area or facility can be detected, evaluated and monitored.

FIGURE 1

301

303

302

FIGURE 2c

301

303

302

FIGURE 2b

301

303

302

FIGURE 2a

SENSOR

ACTIVE RF BADGE READER ~203

ACTIVE IR BADGE READER ~204

PASSIVE RF BADGE READER ~205

PASSIVE IR BADGE READER ~206

~105

BAR CODE READER ~207

GPS BADGE READER ~208

CLOTHING MARK/COLOR READER ~209

BIOMETRIC READER ~210

SPECIALIZED INFRARED CAMERA ~211

OTHER SENSORS ~215

IDENTIFIER

~212

FIGURE 3

FIGURE 4

• Plant Sector Being Viewed:
  – Sector 1A
• Personnel on Screen:
  – John
    -Moving West
    - W.O. Open
  – Mary
    - Stationary
    - No W.O. Open
• Camera Being Viewed:
  – Camera A
  – Visible Spectrum
  – No Alarms

John

Mary

Talk

Talk

FIGURE 5

- **Asset Being Viewed:**
  - John
  - Moving- West
  - Location- Sector 1A
  - Has been-
    - Sector 1B
    - Sector 2C
  - Vector started at 7G

- **Work Order Open:**
  - W.O. #3409
  - Repack Valve on Cracking Tower #3
  - Special Instructions
  - Do Not impede flow into cracking tower #3 while repacking

John

Talk

FIGURE 6

- **Plant Sector Being Viewed:**
  - Sector 1A
- **Personnel:**
  - Two in Sector, See Main Screen
- **Cameras:**
  - Camera A, See Main Screen
- **Abnormal Situations:**
  - No abnormal situations reported

John   Mary

Talk   Talk

FIGURE 7

- <u>ALERT!!!:</u> ICE FORMATION DETECTED - CHECK SECTOR 2A.

- <u>ALERT!!!:</u> THE FLARE HAS CHANGED SHAPE!

- THE FLARE HAS CHANGED COLOR!

FIGURE 9

FIGURE 9

Hypothesis	$z_1(1)$	$z_2(1)$	$z_1(2)$	Track No.
H_1	0	0	-	0
H_2	1	0	-	1
H_3	0	2	-	2
H_4	1	2	-	1,2
H_5	1	0	3(1)	3
H_6	1	2	3(1)	2,3
H_7	0	2	4(2)	4
H_8	1	2	4(2)	1,4
H_9	0	0	5	5
H_{10}	1	0	5	1,5
H_{11}	0	2	5	2,5
H_{12}	1	2	5	1,2,5

FIGURE 10

FIGURE 11

FIGURE 12

Interpolation Pattern

FIGURE 13

Interpolation Scheme

250

249

251

(i_w, j_w)

(i_w+m, j_w+n)

(i_w-k, j_w+l)

FIGURE 14

METHOD AND APPARATUS FOR TRACKING WITH IDENTIFICATION

[0001] This application claims the benefit of U.S. Provisional Application No. 60/284,863, entitled "Method and Apparatus for Tracking People", filed Apr. 19, 2001, wherein such document is incorporated herein by reference.
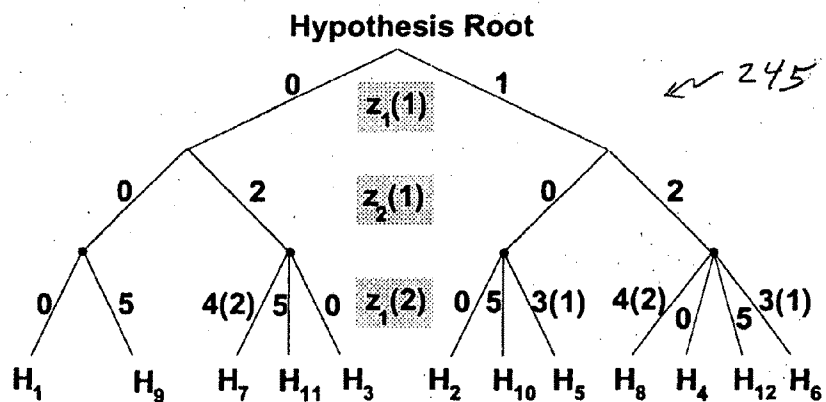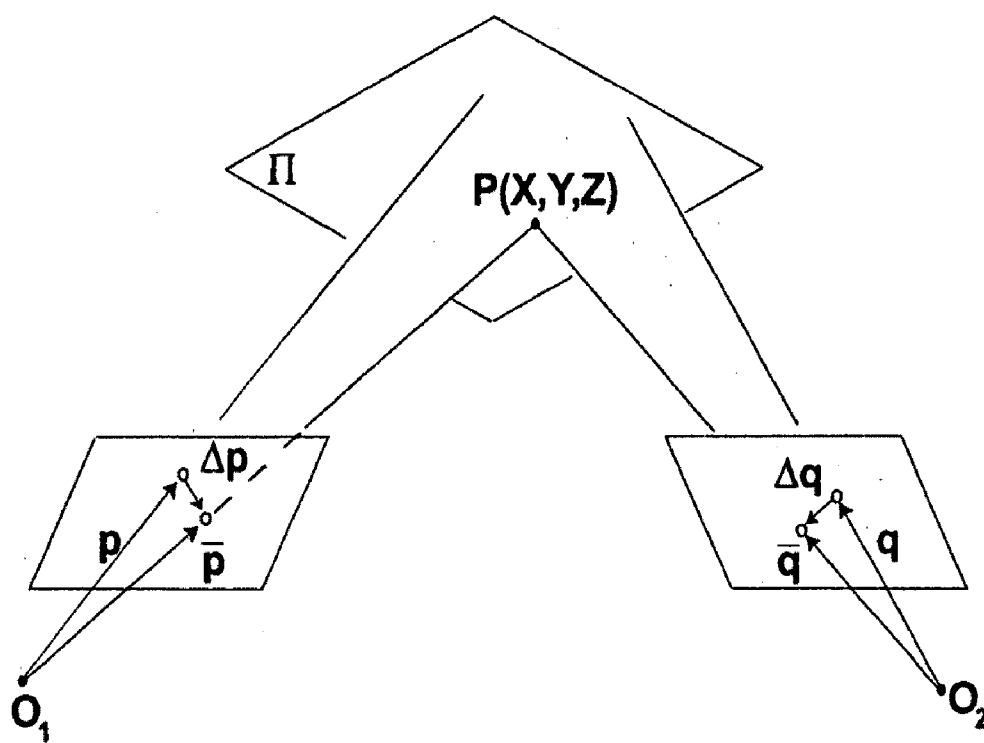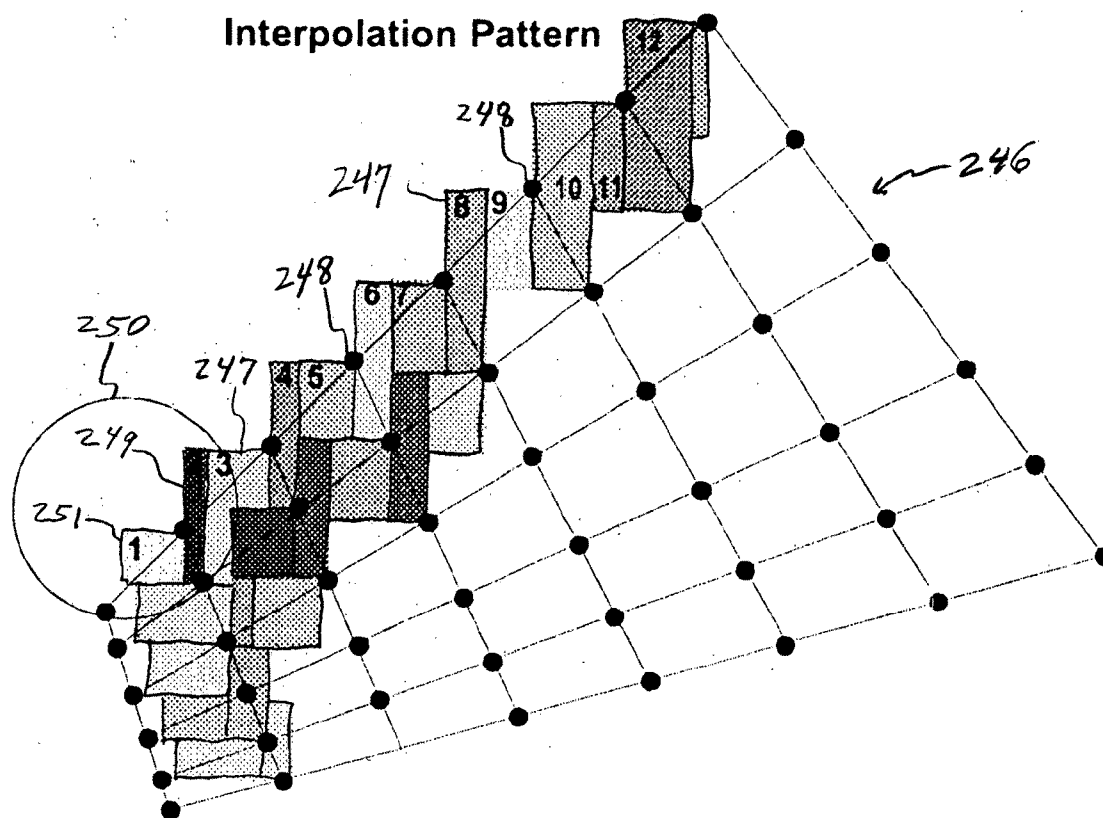
BACKGROUND

[0002] The invention relates to tracking objects and people and particularly tracking with identification.

[0003] Other applications relating to similar technology include U.S. patent application Ser. No. 10/034,696, filed Dec. 27, 2001, and entitled, "Surveillance System and Methods Regarding Same", which is incorporated herein by reference; U.S. patent application Ser. No. 10/034,780, filed Dec. 27, 2001 and entitled "Method for Monitoring a Moving Object and System Regarding Same", which is incorporated herein by reference; and U.S. patent application Ser. No. 10/034,761, filed Dec. 27, 2001 and entitled "Moving Object Assessment System and Method", which is incorporated herein by reference.

SUMMARY

[0004] The invention involves the tracking of people or objects with image processing and the identification of the people or objects being tracked. Also, conditions of an area or a facility can be detected and tracked.

BRIEF DESCRIPTION OF THE DRAWING

[0005] FIG. 1 is a diagram of a people and object tracker system;

[0006] FIGS. 2a, 2b and 2c reveal infrared detection of ice on an object;

[0007] FIG. 3 shows various sensors and their connection to the identifier;

[0008] FIG. 4 illustrates a user interface component with a monitor, keyboard, mouse and electronics module;

[0009] FIG. 5 shows an illustrative example display of tracking and identification in a work facility;

[0010] FIG. 6 is similar to FIG. 5 but having a focus on one of the workers with more detailed information;

[0011] FIG. 7 shows worker locations relative to a process diagram of a work facility;

[0012] FIG. 8 reveals a correlation between a plan view and process diagram of the work facility, with alarm status information;

[0013] FIG. 9 is a diagram of the architecture of a multiple hypotheses tracking algorithm;

[0014] FIG. 10 is a hypothesis matrix;

[0015] FIG. 11 a hypothesis tree diagram;

[0016] FIG. 12 shows the fields of view of two cameras having a common area;

[0017] FIG. 13 is a mapping result of one grid of an image onto a grid of a warped image; and

[0018] FIG. 14 is a close up view of a portion of the mapping revealed in FIG. 13.

DESCRIPTION

[0019] The present invention combines technologies and methods into a single integrated approach to track people or objects. It is particularly useful for tracking employees in relatively large settings, for examples, refineries and factories. The primary technology for tracking is cameras, imaging devices or visual sensors and image processing. The partnering technology involves identifying mechanisms such as specialty readers and/or sensors used for positive identification of people and objects. Positive identification of people would be made at choke points of paths of movement of the people, or at appropriate check points. Fusion of the primary and partnering technologies results in a powerful technology for tracking objects or people, such as workers in a plant or another area. This fusion or combination of the mentioned technologies is one aspect of an application of the invention.

[0020] Further application of the invention can be obtained by imbedding fused camera and positive technology into a mapping system. This system provides for easy access or visualization of information, as well as transformation of the information into a context of spatial coordinates or other forms. One instance is imbedding the fused camera and positive identification information into a global information system (GIS) mapping system to show the location of a tracked person in relation to certain equipment in a database of a factory or refinery, or in relation to a process location on a process diagram of the factory or refinery. Because a process diagram may not correspond directly to geographic or physical space in the field, the geographic or physical location of a tracked person may have to be translated or mapped to the corresponding space in the process diagram. This feature can be added to the base technology, which is the combined/fused image processing and specific identification technologies, or any variation of the base technology along with other technologies or features as desired.

[0021] In tracking an employee, additional information can be developed about the condition of an employee or the plant that the employee is at. Such information may improve the capability of the present system to track employees effectively and to most efficiently direct employee efforts in the plant for a known set of conditions in the plant. For example, specialized detection capabilities in the viewing camera and the video processing system may detect whether the employee has stopped moving, thus implying possible incapacitation of, or injury to the employee. This information would allow the system to conclude that there may be a problem with the employee and thus alert an operator to investigate the situation of the employee. Special video detection capabilities can be utilized to monitor the condition of the plant or installation in a manner that provides for a more thorough diagnosis of the plant's condition.

[0022] FIG. 1 is a diagram of an embodiment of a people or object tracker 100. The tracking of, for example, people is used here for illustrating the invention. An area or facility may be observed with multiple cameras or imaging devices 101 situated throughout, for instance, the facility. The images from the cameras or imaging devices 101 are fed electronically in digital or analog format into an image processing component 102 that includes tracking capability of people or objects detected by the cameras 101. The path or track of an object or person being tracked is noted. The image processing of component 102 is to maintain continuity of the tracking even if no cameras 101 have field-of-view overlaps with one another. Tracking may also include observation of immobile objects. One example would be the observation of a valve or pipeline by a specially designed IR camera 215 that could provide an image indicating frost or ice on the valve or pipeline thereby indicating a possible problem.

[0023] The particular properties of water have been established in the upper near-infrared spectrum (i.e., 0.4 µm to 2.4 µm). Water and objects bearing water have very low reflectivity and thus essentially are black bodies in this spectrum. This black body characteristic is noticeable in an image when the water is above a virtual mass. A light cover of rainwater will not reveal this characteristic. However, a cover of ice (i.e., concentrated water) or a human body (i.e., thick water-based mass) will have such black body characteristic. FIG. 2a shows a metal cylinder 301 having an iced lower portion 302 which appears as a black body. FIG. 2b shows the cylinder midway in de-icing. Black area 302 has started shrinking. The lower portion 302 of cylinder 301 is completely de-iced and appears the same as upper part 303 in FIG. 2c. A pseudo-coloring algorithm can be used to present in different colors the different states of ice coverage. Using the different colors, an operator is given an idea how thick the ice is and he or she is dynamically updated as to the progress of the de-icing operation. Also, the algorithm will determine if the level of illumination is sufficient for such ice detection.

[0024] For positively identifying the objects or people being tracked, one or more of a group 105 of sensors may be utilized. This group of sensors includes active RF badge readers 203, active IR badge readers 204, passive RF badge readers 205, passive RF badge readers 206, long or short range bar code readers 207, GPS badge readers 208, identifying clothing marker readers 209 such as colors, shapes and numbers on the clothing, biometric readers 210 (e.g., fingerprint or retina), specialized IR cameras 211 and other sensors 215. The sensors' 105 connection to identifier 212 is shown in FIG. 3. These are just examples of identification sensors which may be used. Such sensors would be placed in area on a grid-like layout at choke points, gates, security check points, entry points, narrow lanes of travel, and so on. Outputs of these sensors 105 go to an input of a positive identifying component 212.

[0025] The output from positive identifying component 212, which is a synthesis of the inputs from the various sensors or sensor indicating a positive identification of the object or person being tracked, goes to fusion component 213. Other inputs to component 213 come from vision processor component 102 and a database component 214. Database component 214 is a source of information for fusion component 213. Fusion component may send out signals to component 213 requesting certain information about certain objects or persons being tracked. For instance, work order systems may be a source of valuable information useful in the tracking of a worker in a factory or refinery. Database component 214 is a source of many types of information in numerous databases.

[0026] The output of the fusion component 213 goes to a user interface component 216. Component 216 typically would have interface electronics 234, and a screen 218 with a display 225 in FIG. 5 showing the object or person being tracked and, whenever possible, a positive identification of the tracked person or object.

[0027] An example of an application of the fused camera tracking and positive identification of system 100 is illustrated in FIG. 5. One component of user interfaces 216 is a monitor 217 having a display screen 218, a keyboard 219 and a mouse 220, as shown in FIG. 4. FIG. 5 is displayed on an operator screen 218. FIG. 5 is a display 225. A plan view 221 of an industrial installation is shown. Employees John 222 and Mary 223 are shown on the plan view. Photo 224 shows

the area where employees 222 and 223 are known to be located. On the right of display 225 of screen 218 is an information section 227 about the plant area being viewed, the employees shown, and a camera 226 of imaging device(s) 101 used. Sector 1A of the plant is being viewed by camera A (i.e., camera 226). John 222 is moving West with an open W.O. (i.e., work order). Mary 223 is stationary with no W.O. open. There are no alarms on. Camera 226 and image processing 103 are tracker 104 enable visual tracking of employees 222 and 223. The other components used are a sensor from sensor group 105 and identifier 212. These components provide positive identification of employees 222 and 223. For instance, radio frequency (RF) identification (ID) tags are worn by employees 222 and 223. These employees are positively identified by an RF tag or badge reader 203 or 205 at a selected choke point (e.g., entrance 228) in the plant or other convenient location. Besides RF ID tags, other items may be used for positive identification of an employee tracked by system 100. Such items could include IR tags, badge swiping, and fingerprint, palm, retinal or face scanning. Also included could be visual detection of badges or clothing with unique colors or shapes, bar code scanning (e.g., with a bar code located on the person's badge or clothing), or any other method or technology available for identification.

[0028] Plan view 221 of the area monitored by camera A 226, shows employees 222 and 223 to be present in the area and at their respective locations. John 222 and Mary 223 were positively identified by a reader when walking through a choke point or entrance 228 when entering the area, and were tracked to their current locations by vision processor 102. An RF reader 203 sent additional information about employees 222 and 223 to identifier 212. Identifier 212 processed and forwarded this information to vision processor 102, when employees 222 and 223 were detected by reader 203 at choke point 228. If communication is desired with John or Mary, an intercom "Talk" button 229 proximate to John or Mary's name on screen 225 may be activated with a touch of the respective button.

[0029] Additional video technologies may be used to improve tracking and identification of a person or object. One such technology is cooperative camera networks (CCN). CCN can detect change in a scene viewed by a camera. Such change is detected with the use of frame differencing and adaptive thresholding. Frames of a video are compared to detect differences between a current frame and a reference frame. The parts of the current frame that differ from the reference frame are extracted and a histogram is done of the pixels of those extracted parts. A threshold level is assigned to the histogram that provides for a division between what is actually change and what is noise. CCN can be used, for example, to evaluate the composite color of a person's clothes so as to help identify and track such person.

[0030] FIG. 6 shows additional information being integrated with the video tracking and positive identification information. Additional information inputs may include, for example, in a factory or refinery setting, information about a particular machine's performance, information about work orders that are open and held by a worker shown on display 225, information about potential or evolving abnormal situations in the industrial processes, information about a repair process that is about to be undertaken by a worker, or other types of information. An example of such information shown in information section 227 of display 225 of FIG. 6, which is on screen 218 of monitor 217. Work order information, work

detail, special instructions and other information are accessed by tracking system **100** from database(s) **214**. Additional capabilities added to display **225** increase the overall power and effectiveness of system **100**. The lower part of display **225** shows the relevant portion of plan view **227**.

[0031] The above-mentioned GIS can also improve the ability of tracking system **100**. GIS can locate and translate spatial information by implementing a fine measuring grid. The area around each intersection on the grid may be designated area of influence for that intersection. The area of influence may be correlated to the portions of a map that are not directly spatially related to the plant, factory or refinery, such as a process diagram **230** as shown in FIG. **7**. Diagram **230** is an example of mapping where GIS is used to correlate spatial information to non-spatial formats. Process diagram **230** shows John **222** and Mary **223** being located near the process or process equipment that they are standing near in the plant. Field-of-view lines of camera A **226** are bent because the camera field of view is distorted as it is translated from plan view **221** of FIG. **5** into process diagram **230** of FIG. **7**.

[0032] Special video detection features add to the diagnostic capabilities of system **100**, thereby increasing its power and effectiveness. FIG. **8** illustrates some detection features that may be particularly useful for oil refineries. For instance, ice may coat some pipes and valves in an oil refinery which can prevent valves **233** from operating properly. Improper operation of these valves **233** can lead to serious functional problems in the refinery. Ice detection capabilities can be added as a feature (i.e., ice detection based on near infrared phenomenology) as illustrated in FIGS. **2**a, **2**b and **2**c. Other features can include capabilities used for flare (flame) detection, including detecting changes in the color and/or shape of the flare. An ultra violet light detector may be used to monitor a flame. Information section **227** provides examples of alert notices of ice formation and location and of flare change which may indicate the change of quality and quantity of the flames constituting the flare.

[0033] GIS can be a resource in database component **214**. GIS may be used to assist in objectively describing the actual location of a person based on maps processed by image processing component **102**. Further information from such a system could provide the location of a person or object tracked relative to a process flow layout, such as one of a refinery.

[0034] Image processing component **102** consists of processing for multicamera surveillance and object or person tracking. Component **102** has a moving object segmentor, a tracker and a multi-camera fusion module. One object detection method is based on a mixture of Normal representation at the pixel level. Each normal reflects the expectation that samples of the same scene point are likely to display Gaussian noise distributions. The mixture of Normals reflects the expectation that more than one process may be observed over time. The method used here is similar in that a multi-Normal representation is used at the pixel level. But that is the extent of the similarity. The present method uses an Expectation-Maximization (EM) algorithm to initialize models in the present invention. The EM algorithm provides strong initial statistical support that facilitates fast convergence and stable performance of the segmentation operation. The Jeffreys (J) divergence measure is used as the measuring criterion between Normals of incoming pixels and existing model Normals. When a match is found, the model update is performed using a method of moments. When a match is not

found, the update is performed in a way that guarantees the inclusion of the incoming distribution in the foreground set.

[0035] The method just described permits the identifying foreground pixels in each new frame while updating the description of each pixel's mixture model. The identified and labeled foreground pixels can then be assembled into objects using a connected components algorithm. Establishing a correspondence of objects between frames (i.e., tracking) is accomplished by using a linearly predictive multiple hypotheses tracking algorithm which incorporates both position and size. The object tracking method is described below.

[0036] Although overlapping or fusion of fields of view (FOV) are not required for the image processing with tracking component **102**, fusion of FOV's is discussed here. Fusion is useful since no single camera can cover large open spaces in their entirety. FOV's of various cameras may be fused into a coherent super picture to maintain global awareness. Multiple cameras are fused (calibrated) by computing the respective nomography matrices. The computation is based on the identification of several landmark points in the common FOV between camera pairs. The landmark points are physically marked on the scene and sampled through the user interface. The achieved calibration is very accurate.

[0037] The present FOV fusion system has a warping algorithm to accurately depict transformed views. This algorithm computes a near optimal camera configuration scheme since the cameras are often far apart and have optical axes that form angles which vary quite much. Resulting homographies produce substantially skewed frames where standard warping fails but the present warping succeeds.

[0038] Object or person tracking by image processing component can substantively begin with an initialization phase. The goal of this phase is to provide statistically valid values for the pixels corresponding to the scene. These values are then used as starting points for the dynamic process of foreground and background awareness. Initialization needs to occur only once. There are no stringent real-time processing requirements for this phase. A certain number of frames N (N=70) are accumulated and then processed off-line.

[0039] Each pixel x of an image (of the scene) is considered as a mixture of five time-varying trivariate Normal distributions:

$$x \sim \sum_{i=1}^{5} \pi_i N_3 \left(\mu_i, \sum_i \right),$$

where:

$$\pi_i \geq 0, \, i = 1, K, 5$$

and

$$\sum_{i=1}^{5} \pi_i = 1$$

are the mixing proportions (weights) and N_3 (μ, Σ) denotes a trivariate Normal distribution with vector mean μ and variance-covariance matrix Σ. The distributions are trivariate to account for the three component colors (i.e., red, green and blue) of each pixel in the general case of a color camera.

[0040] Initialization of the pixel values here involves partially committing each data point to all of the existing distributions. The level of commitment is described by appropriate

4

weighting factors. This is accomplished by the EM method noted above. The EM algorithm is used to estimate the parameters of the initial distribution π_1, μ_1 and Σ_i, i=1, . . . , 5 for every pixel x in the scene. Since the EM algorithm is applied off-line over N frames, there are N data points in time available for each pixel. The data points x_j, j=1, . . . , N are triplets:

$$x_j = \begin{pmatrix} x_j^R \\ x_j^G \\ x_j^B \end{pmatrix}$$

where x_j^R, x_j^G, and x_j^B stand for the measurement received from the red, green and blue channels of the camera for the specific pixel at time j. This data $x_1, x_2, \ldots, x_N$ are assumed to be sampled from a mixture of 5 trivariate Normals:

$$x_j \approx \sum_{i=1}^{5} \pi_i N_3 \left[\begin{pmatrix} \mu_i^R \\ \mu_i^G \\ \mu_i^B \end{pmatrix}, \sigma_i^2 I \right],$$

where the variance-covariance matrix is assumed to be diagonal with x_j^R, x_j^G, and x_j^B having identical variance within each Normal component, but not across all components (i.e., $\sigma_k^2 \neq \sigma_1^2$ for k≠1 components).

[0041] Originally, the algorithm is provided with some crude estimates of the parameters of interest: $\pi_1^{(0)}$, $\mu_i^{(0)}$, and $(\sigma_i^{(0)})^2$. These estimates are obtained with a K-means method which commits each incoming data point to a particular distribution in the mixture model. Then, the following loop is applied.

[0042] For k=0, 1, . . . calculate:

$$z_{ij}^{(k)} = \frac{\pi_i^{(k)} (\sigma_i^{(k)})^{-3/2} \exp\left\{ -\frac{1}{2(\sigma_i^{(k)})^2} (x_j - \mu_i^{(k)})'(x_j - \mu_i^{(k)}) \right\}}{\sum_{t=1}^{(5)} \pi_t^{(k)} (\sigma_i^{(k)})^{-3/2} \exp\left\{ -\frac{1}{2(\sigma_i^{(k)})^2} (x_j - \mu_i^{(k)})'(x_j - \mu_i^{(k)}) \right\}},$$

$$\pi_i^{(k+1)} = \frac{1}{N} \sum_{j=1}^{N} z_{ij}^{(k)},$$

$$\mu_i^{(k+1)} = \frac{1}{N\pi_i^{(k+1)}} \sum_{j=1}^{N} z_{ij}^{(k)} x_j,$$

$$(\sigma_i^{(k+1)})^2 = \frac{1}{3N\pi_i^{(k+1)}} \sum_{j=1}^{N} z_{ij}^{(k)} (x_j - \mu_i^{(k+1)})'(x_j - \mu_i^{(k+1)}),$$

for i=1, . . . , and j=1, . . . , N. Then, set k=k+1 and repeat the loop.

[0043] The condition for terminating the loop is:

$$|\pi_i^{(k+1)} - \pi_i^{(k)}| < \epsilon, \ i=1, \ldots, 5,$$

where ϵ is a 'small' positive number (10^{-2}) $z_{ij}^{(k)}$ are the posterior probabilities that x_j belongs to the i-th distribution and they form a 5×N matrix at the k-th step of the computation. The EM process is applied for every pixel in the focal plane array (FPA) of the camera. The result is a mixture model of

five Normal distributions per pixel. These Normal distributions represent five potentially different states for each pixel. Some of these states could be background states and some could be transient foreground states. The EM algorithm is computationally intensive, but since the initialization phase takes part off line, this is a non-issue.

[0044] There is a segmentation of moving objects. The initial mixture model is updated dynamically thereafter. The update mechanism is based on the incoming evidence (i.e., new camera frames). Several items could change during an update cycle:

[0045] 1. The form of some of the distributions could change (weight π_i mean μ_1, and variance σ_1^2).

[0046] 2. Some of the foreground states could revert to background and vice versa.

[0047] 3. One of the existing distributions could be dropped and replaced with a new distribution.

At every point in time, the distribution with the strongest evidence is considered to represent the pixel's most probable background state.

[0048] The update cycle for each pixel proceeds as follows:

[0049] 1. First, the existing distributions are ordered in descending order based on their weight values.

[0050] 2. Second, the algorithm selects the first B distributions that account for a predefined fraction of the evidence T:

$$B = \operatorname*{argmin}_{b} \left\{ \sum_{i=1}^{b} w_i > T \right\},$$

[0051] where w_i, i=1, . . . , b are the respective distribution weights. These B distributions are considered as background distributions while the remaining 5-B distributions are considered foreground distributions.

[0052] 3. Third, the algorithm checks if the incoming pixel value can be ascribed to any of the existing Normal distributions. The matching criterion is the Jeffreys (J) divergence measure.

[0053] 4. Fourth, the algorithm updates the mixture of distributions and their parameters. The nature of the update depends on the outcome of the matching operation. If a match is found, the update is performed using the method of moments. If a match is not found, then the weakest distribution is replaced with a new distribution. The update performed in this case guarantees the inclusion of the new distribution in the foreground set.

[0054] There is a matching operation. The Kullback-Leibler {KL) number between two distributions f and g is defined as:

$$K(f, g) = E_f \left[\log\left(\frac{f}{g}\right) \right] = \int \log\left(\frac{f(x)}{g(x)}\right) f(x) d_x$$

A formal interpretation of the use of the KL information number is of whether the likelihood ration can discriminate between f and g when f is the true distribution.

[0055] For the purpose of the algorithm, one needs to define some divergence measure between two distributions, so that if the divergence measure between the new distribution and one of the existing distributions is "too small" then these two distributions will be pooled together (i.e., the new data point

will be attached to one of the existing distributions). For a divergence measure d(f,g), it is necessary to satisfy (at least) the following three axioms:

$$d(f,f)=0 \qquad\qquad (a)$$

$$d(f,g)\geqq 0 \qquad\qquad (b)$$

$$d(f,g)=d(g,j). \qquad\qquad (c)$$

[0056] The KL information number between two distributions f and g does not satisfy (c), since:

$$K(f, g) = E_f\left[\log\left(\frac{f}{g}\right)\right] \neq E_g\left[\log\left(\frac{g}{f}\right)\right] = K(g, f)$$

i.e., the KL information number is not symmetric around its arguments and thus it can not be considered as a divergence measure.

[0057] The Jeffreys divergence measure between two distributions f and g is the following:

$$J(f, g) = \int [f(x) - g(x)]\log\left[\left(\frac{f(x)}{g(x)}\right)\right]dx.$$

This divergence measure is closely related to the KL information number, as the following Lemma indicates:

[0058] Lemma 1:

$$J(f,g)=K(f,g)+K(g,f)$$

[0059] Proof:

$$J(f, g) = \int [f(x) - g(x)]\log\left(\frac{f(x)}{g(x)}\right)dx$$
$$\int f(x)\log\left(\frac{f(x)}{g(x)}\right)dx +$$
$$\int g(x)\log\left(\frac{g(x)}{f(x)}\right)dx$$
$$= K(f, g) + K(g, f).$$

[0060] The J(f,g) is now symmetric around its arguments since:

$$J(f,g)=K(f,g)+K(g,f)=K(g,f)+K(f,g)=J(g,f)$$

and satisfies also axioms (a) and (b). Thus J(f,g) is a divergence measure between f and g.

[0061] J(f,g) is used to determine whether the new distribution matches or not to one of the existing five distributions. The five existing Normal distributions are: $f_i \sim N_3(\mu_i, \sigma_i^2 I)$, i=1, . . . , 5. The incoming distribution is $g \sim N_3(\mu_g, \sigma_g^2 I)$. We assume that:

$$\mu_g = x_t \text{ and } \sigma_g^2 = 25,$$

where x_t is the incoming data point. The five divergence measures between g and f_i, i=1, . . . , 5 will be given by the following formula:

$$J(f_i, g) = \frac{3}{2}\left(\frac{\sigma_i}{\sigma_g} - \frac{\sigma_g}{\sigma_i}\right)^2 + \frac{1}{2}\left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_g^2}\right)(\mu_g - \mu_i)'(\mu_g - \mu_i).$$

[0062] Once the five divergence measures have been calculated, the distribution $f_j (1 \leqq j \leqq 5)$ is found for which:

$$J(f_j, g) = \min_{1 \leq i \leq 5}\{J(f_i, g)\}$$

and there is a match between f_j and g if and only if $J(f_j,g) \leqq K^*$, where K^* is a prespecified cutoff value. In the case where $J(f_j,g) > K^*$ then the new distribution g cannot be matched to any of the existing distributions.

[0063] There is a model update when a match is found. If the incoming distribution matches to one of the existing distributions, these two are pooled together to a new Normal distribution. This new Normal distribution is considered to represent the current state of the pixel. The state is labeled either background or foreground depending on the position of the matched distribution in the ordered list of distributions.

[0064] The parameters of the mixture are updated with the method of moments. First introduced is some learning parameter a which weighs on the weights of the existing distributions. A 100α% weight is subtracted from each of the five existing weights and it is assigned to the incoming distribution's weight. In other words, the incoming distribution has weight α since:

$$\sum_{i=1}^{5} \alpha\pi_i = \alpha\sum_{i=1}^{5} \pi_i = \alpha$$

and the five existing distributions have weights: $\pi_i(1-\alpha)$, i=1, . . . , 5.

[0065] Obviously, for α, 0<α<1 is needed. The choice of a depends mainly on the choice of K^*. The two quantities are inversely related. The smaller the value of K^*, the higher the value of α and vice versa. The values of K^* and α are also affected by how much noise there is in the monitoring area. So if, for example, an outside region was being monitored and had much noise due to environmental conditions (i.e., rain, snow, etc.), then a "high" value of K^* and thus a "small" value of a would be needed since a non-match to one of the distributions is very likely to be caused by background noise.

[0066] On the other hand, if the recording was being done indoors where the noise is almost non-existent, then a "small" value of K^* and thus a higher value of α would be preferred, because any time there is not a match to one of the existing five distributions, it is very likely to occur due to some foreground movement (since the background has almost no noise at all).

[0067] Assuming that there is a match between the new distribution g and on of the existing distributions f_j where $1 \leqq j \leqq 5$, then the weights of the mixture model are updated as follows:

$$\pi_{i,t}(1-\alpha)\pi_{i,t-1} \; i=1, . . . , 5 \text{ and } i \neq j$$

$$\pi_{j,t}=(1+\alpha)\pi_{j,t-1}+\alpha$$

[0068] The mean vectors and variances are also updated. If w_1 is $(1-\alpha)\pi_{j,t-1}$, i.e., w_1 is the weight of the j-th component (which is the winner in the match) before pooling it with the new distribution g, and if $w_2=\alpha$, i.e., the weights of the new observation then define:

$$\rho = \frac{w_2}{w_1 + w_2} = \frac{\alpha}{(1-\alpha)\pi_{j,t-1}} \cdot + \alpha$$

Using the method of moments leads to:

$$\lambda_{j,t} = (1-\eta)\mu_{j,t-1} + \rho\mu_g$$

$$\sigma_{j,t}^2 = (1-\rho)\sigma_{j,t-1}^2 + \rho\sigma_g^2 + \rho(1-\rho)(x_t - \mu_{j,t-1})'(x_t - \mu_{j,t-1}),$$

while the other 4 (unmatched) distributions keep the same mean and variance that they had at time t-1.

[0069] There is a model update when a match is not found. In the case where a match is not found (i.e., $\min_{1 \leq I \leq 5} K(f_i, g) > K^*$), then the current pixel state is committed to be foreground and the last distribution in the ordered list is replaced with a new one. The parameters of the new distribution are computed as follows:

   [0070]   1. The mean vector σ_5^2 is replaced with the incoming pixel value.

   [0071]   2. The variance σ_5^2 is replaced with the minimum variance from the list of distributions.

   [0072]   3. The weight of the new distribution is computed as follows:

$$w_{5,t+1} = \frac{1-T}{2},$$

   [0073]   where T is the background threshold index.

[0074] This formula guarantees the classification of the current pixel state as foreground. The weights of the remaining four distributions are updated according to the following formula:

$$w_{i,t+1} = w_{i,t} + \frac{w_{5,t} - (1-T)/2}{4}.$$

[0075] Multiple hypotheses are developed for predictive tracking. In the above, there was described a statistical procedure to perform on-line segmentation of foreground pixels corresponding to moving objects of interest, i.e., people and vehicles. Here, how to form trajectories traced by the various moving objects is described. The basic requirement for forming object trajectories is the calculation of blob centroids (corresponding to moving objects). Blobs are formed after a standard 8-connected component analysis algorithm is applied to the foreground pixels. The connected component algorithm filters out blobs with an area less than A=3×9=27 pixels as noise. This is the minimal pixel footprint of the smallest object of interest (e.g., a human) in the camera's FOV.

[0076] A multiple hypotheses tracking (MHT) is then employed that groups the blob centroids of foreground objects into distinct trajectories. MHT is considered to be the best approach to multi-target tracking applications. It is a recursive Bayesian probabilistic procedure the maximizes the probability of correctly associating input data with tracks. Its superiority against other tracking algorithms stems from the fact that it does not commit early to a trajectory. Early commitment usually leads to mistakes. MHT groups the input data into trajectories only after enough information has been collected and processed. In this context, it forms a number of candidate hypotheses regarding the association of input data with existing trajectories. MHT has shown to be the method of choice for applications with heavy clutter and dense traffic. In difficult multi-target tracking problems with crossed trajectories, MHT performs very well.

[0077] FIG. 9 depicts the architecture of the multiple hypotheses tracking (MHT) algorithm involving a blob centroid 235. The modules of this algorithm are prediction 236, validation 237, hypothesis generation 238 and hypothesis evaluation 239. An integral part of any tracking system is prediction module 236. Prediction provides estimates of moving objects' states and in the present system is implemented as a Kalman filter. Kalman filter predictions are made based on prior models for target dynamics and measurement noise.

[0078] The state vector describing the motion of a foreground object (blob) consists of the position and velocity of its centroid expressed in pixel coordinates, i.e.,

$$x_k = (x_k, \dot{x}_k, y_k, \dot{y}_k)^T.$$

[0079] The state space model is a constant velocity model given by:

$$x_{k+1} = F_k x_k + u_k,$$

with transition matrix F_k:

$$F_k = \begin{pmatrix} 1 & dt & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & dt \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

[0080] The process noise is white noise with a zero mean and covariance matrix:

$$Q_k = E[u_k u_k^T] = \begin{pmatrix} \dfrac{dt^3}{3} & \dfrac{dt^2}{2} & 0 & 0 \\ \dfrac{dt^2}{2} & dt & 0 & 0 \\ 0 & 0 & \dfrac{dt^3}{3} & \dfrac{dt^2}{2} \\ 0 & 0 & \dfrac{dt^2}{2} & dt \end{pmatrix} q,$$

where q is the process variance. The measurement model describes how measurements are made and it is defined by:

$$z_k = Hx_k + v_k$$

with

$$H = \begin{pmatrix} 1000 \\ 0010 \end{pmatrix},$$

and a constant 2×2 covariance matrix of measurement noise given by:

$$R_k = E[v_k v_k^T].$$

[0081] Based on the above assumptions, the Kalman filter provides minimum mean squared estimates $\hat{x}_{k|k}$ of the state vector according to the following equations:

$$K_k = P_{k|k-1} H^T [H P_{k|k-1} H^T + R_k]^{-1}$$

$$P_{k|k} = [I - K_k H] P_{k|k-1}$$

$$P_{k+1|k} F k P_{k|k} F_k^T + Q_k$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k [Z_k - H \hat{x}_{k|k-1}]$$

$$\hat{x}_{k+1} = F_k \hat{x}_{k|k}.$$

[0082] Validation 237 is a process which precedes the generation of hypotheses 238 regarding associations between input data (blob centroids 235) and the current set of trajectories (tracks). Its function is to exclude, early-on, associations that are unlikely to happen thus limiting the number of possible hypotheses to be generated. The vector difference between measured and predicted states v_k is a random variable characterized by the covariance matrix S_k:

$$v_k = Z_k - H \hat{x}_{k|k-1}$$

$$S_k = H P_{k|k-1} H^T + R_k.$$

[0083] For every track from the list of current tracks there exists an associated gate. A gate can be visualized as an area surrounding a track's predicted location (next move). In the present case, a gate is an elliptical shape defined by the squared Mahalanobis distance:

$$d^2 = v_k^T S_k^{-1} v_k.$$

[0084] An incoming measurement (blob centroid 235) is associated with a track only when it falls within the gate of the respective track. Mathematically this is expressed by:

$$d^2 \leq D_{threshold}.$$

The result of validating a new set of blob centroids takes the form of an ambiguity matrix. An example of an ambiguity matrix corresponding to a hypothetical situation of an existing set of two tracks (T_1 and T_2) and a current set of three measurements ($z_1(k)$, $z_2(k)$ and $z_3(k)$) is given in Equation (1).

$$\Omega = \begin{pmatrix} T_F & T_1 & T_2 & T_N \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} z_1(k) \\ z_2(k) \\ z_3(k) \end{matrix} \qquad \text{Equation (1)}$$

The columns of the ambiguity matrix denote the current set of tracks with the first and last columns being reserved for false alarms (T_F) and new tracks (T_N), respectively. The rows correspond to the particular measurements of blob centroids made on the current frame. Non zero elements of the ambiguity matrix signal that the respective measurements are contained in are in the validation region of the associated track. The assignments are further constrained in the ambiguity matrix by allowing each measurement in the current scan to be associated with only one track. Further, it is assumed that a track is paired with at most one measurement per iteration. Therefore, the number of non zero elements in any row or

column (barring the first and last columns) is limited to one. Thus, the ambiguity matrix is made a cost matrix as it is defined in linear assignment problems. This formulation makes the ambiguity matrix a representation of a new set hypotheses about blob centroid-track pairings.

[0085] Central to the implementation of the MHT algorithm is the generation 238 and representation of track hypotheses. Tracks are generated based on the assumption that a new measurement may:

[0086] 1. belong to an existing track;

[0087] 2. be the start of a new track;

[0088] 3. be a false alarm.

[0089] Assumptions are validated through the validation process 237 before they are incorporated into the hypothesis structure. The complete set of track hypotheses can be represented by a hypothesis matrix 240 as shown in FIG. 10. The hypothetical situation in Table I corresponds to a set of two scans of 2 and 1 measurements made respectively on frame k=1 and k+1=2. Some notation clarification is in order. A measurement $z_j(k)$ is the jth observation (blob centroid 235) made on frame k. In addition, a false alarm is denoted by 0 while the formation of a new track (T_{newID}) generated from an old track (T_{oldID}) is shown as $T_{newID}(T_{oldID})$

[0090] The first column 241 in this table is the Hypothesis index. In the example case, there are a total of 4 hypotheses generated during scan 1 shown in column portion 242, and 8 more are generated during scan 2 shown in column portion 243. The last column 244 lists the tracks that the particular hypothesis contains (e.g., hypothesis H_8) contains tracks no 1 and no. 4). The row cells in the hypothesis table denote the tracks to which the particular measurement $z_j(k)$ belongs (e.g., under hypothesis H_{10} the measurement $z_1(2)$ belongs to track no. 5). A hypothesis matrix is represented computationally by a tree structure 245 as it is schematically shown in FIG. 11. The branches of the tree are in essence the hypotheses about measurements-track associations.

[0091] As it is evident from the above example, hypothesis tree 245 can grow exponentially with the number of measurements. Two measures are applied to reduce the number of hypotheses. The first measure is to cluster the hypotheses into disjoint sets. In this sense, tracks which do not compete for the same measurements compose disjoint sets which in turn are associated with disjoint hypothesis trees. The second measure is to assign probabilities on every branch of hypothesis trees. The set of branches with the N_{hypo} highest probabilities are only considered. A recursive Bayesian methodology is followed for calculating hypothesis probabilities from frame to frame.

[0092] Multi-camera fusion is helpful in tracking objects and people. Monitoring of large sites can be best accomplished only through the coordinated use of multiple cameras. A seamless tracking of humans and vehicles is preferred across the whole geographical area covered by all cameras. A panoramic view is produced by fusing the individual camera FOV's. Then object motion is registered against a global coordinate system. Multi-camera registration (fusion) is achieved by computing the Homography transformation between pairs of cameras. The homography computation procedure takes advantage of the overlapping that exists between pairs of camera FOV's. Pixel coordinates of more than 4 points are used to calculate the homography transformation matrix. These points are projections of physical ground plane points that fall in the overlapping area between the two camera FOV's. These points are selected and marked on the

ground with paint during the installation phase. Then the corresponding projected image points are sampled through the Graphical User Interface (GUI). This is a process that happens only in the beginning and once the camera cross-registration is complete it is not repeated. In order to achieve optimal coverage with the minimum number of sensors, the cameras are placed far apart from each other and at varying angles. A sophisticated warping algorithm may be used to accommodate the large distortions produced by the highly non-linear homography transformations.

[0093] An algorithm is used to compute the homography matrices. The algorithm is based on a statistical optimization theory for geometric computer vision and cures the deficiencies exhibited by the least squares method. The basic premise is that the epipolar constraint may be violated by various noise sources due to the statistical nature of the imaging problem. In FIG. **12**, the statistical nature of the imaging problem affects the epipolar constraint. O_1 and O_2 are the optical centers of the corresponding cameras. P(X,Y,Z) is a point in the scene that falls in the common area between the two camera FOV's. Ideally, the vectors $\overrightarrow{O_1p}$, $\overrightarrow{O_2q}$ and $\overrightarrow{O_1O_2}$ are co-planar. Due to the noisy imaging process, however, the actual vectors $\overrightarrow{O_1p}$, $\overrightarrow{O_2q}$ and $\overrightarrow{O_1O_2}$ may not be co-planar.

[0094] In particular, for every camera pair, a 3×3 nomography matrix H is computed such that a number of world points $P_\alpha(X_\alpha, Y_\alpha, Z_\alpha), \alpha=1, 2, \ldots, N$ and $N \geqq 4$, projected into the image points p_α and q_α the following equations holds:

$$\overline{P_\alpha} x \overline{Hq_\alpha} = 0, \alpha=1,2,\ldots, N. \qquad \text{Equation (2)}$$

Notice that the symbol (x) denotes the exterior product and also that the above equation does not hold for the actual image points p_α and q_α but for the corresponding ideal image points $\overline{p}_\alpha$ and $\overline{q}_\alpha$ for which the epipolar constraint is satisfied (see FIG. **12**). Equivalently, the above equation (2) can be written as:

$$(x_\alpha^{(k)}; H)=0, k=1,2,3, \qquad \text{Equation (3)}$$

with:

$$x_\alpha^{(k)} = e^{(k)} x \overline{p}_\alpha \overline{q}_\alpha^T, \alpha=1,2,\ldots, N, \qquad \text{Equation (4)}$$

where for any two matrices A and B $(A;B)=trA^TB)$ and $e^{(1)}=(1, 0, 0)^T$, $e^{(2)}=(0, 1, 0)$, $e^{(3)}=(0, 0, 1)^T$. In a statistical framework, homography estimation is equivalent to minimizing the sum of the following squared Mahalanobis distances:

$$J[H] = \frac{1}{N} \sum_{\alpha=1}^{N} \sum_{k,l=1}^{3} (\Delta X_\alpha^{(k)}, V(X_\alpha^{(k)}, X_\alpha^{(l)}) \Delta X_\alpha^{(1)}),$$

under the constraints described by the above equation (3). Note that the covariant tensor of the matrices $\Delta X_\alpha^{(k)}$ and $\Delta X_\alpha^{(l)}$ is denoted by:

$$V(X_\alpha^{(k)}, X_\alpha^{(l)}) = E[\Delta X_\alpha^{(k)} \oplus \Delta X_\alpha^{(l)}]$$

where $\Delta X_\alpha^{(k)} = X_\alpha^{(k)} - \overline{x}_\alpha^{(k)}$. The symbol ($\oplus$) denotes tensor product. If one uses Lagrange multipliers, estimation of the homography matrix H reduces to the optimization of the following functional J[H]

$$J[H] = \frac{1}{N} \sum_{\alpha=1}^{N} \sum_{k,l=1}^{3} (W_\alpha^{kl}(H)(X_\alpha^{(k)}; H)(X_\alpha^{(l)}; H)). \qquad \text{Equation (5)}$$

The (3×3) weight matrix $W_\alpha(H)$ is expressed as:

$$W_{\alpha(H)} = (p_\alpha x H V[q_\alpha] H^T x P_\alpha + (Hq_\alpha) \times V[p_\alpha](Hq_\alpha))_2^- \qquad \text{Equation (6)}$$

The symbol $(\cdot)_r^-$ symbolizes the generalized inverse of a matrix (N×N) computed by replacing the smallest (N−r) eigenvalues by zeros. The computation process for the optimization of the functional in Equation (5) proceeds as follows:

[0095] 1. Initialization begins by setting the parameter c=0 and the weights $W_\alpha=I$ for $\alpha=1, \ldots N$.

[0096] 2. Proceed by computing the following matrices:

$$M = \frac{1}{N} \sum_{\alpha=1}^{N} \sum_{k,l=1}^{3} W_\alpha^{(kl)} X_\alpha^{(k)} \otimes X_\alpha^{(l)}$$

$$N = \frac{1}{N} \sum_{\alpha=1}^{N} \sum_{k,l=1}^{3} W_\alpha^{(kl)} v(X_\alpha^{(k)}, X_\alpha^{(l)}).$$

[0097] 3. Next calculate the smallest eigenvalue λ_{min} of $\hat{M}=M-cN$ and the associated eigenvector H_{min}.

[0098] 4. If $\lambda_{min} \rightarrow 0$ then the estimated homography matrix $\hat{H}=H_{min}$ is returned and the program exits.

[0099] Otherwise, the weights W_α are updated according to Equation (6) and the value of C_{old} is updated according to:

$$c_{old} = c_{old} + \frac{\lambda_{min}}{(H_{min}; N H_{min})}.$$

[0100] In this latter case, the computation continues by looping back though step 2.

[0101] Due to the specific arrangement of the cameras (large in-between distances and varying pointing angles), the homographies introduce large distortions for those pixels away from the overlapping area. An interpolation scheme is used to compensate for the excessively non-linear homography transformation.

[0102] The scheme is a warping algorithm which interpolates simultaneously across both dimensions. Specifically, the warping computation proceeds as follows:

Step 1. Map the pixel grid of the original image to a warped grid as it is prescribed by the homography transformation. This in general results into mapping the regular rectangular grid of the original image to an irregular quadrilateral grid **246** in the warped space (see FIG. **13**).

Step 2. The warped pixel coordinates in general may take any positive or negative values. Scale these coordinates to a normalized positive range.

Step 3. Employ an interpolation scheme to fill out the pixel values in the warped space. One can visualize why such an interpolation is necessary if one overlays the warped pixel grid on the regular pixel grid as shown in FIG. **13**. The shaded rectangles **247** represent the intermediate regular pixel locations that mediate between the warped grid nodes **248**. The

warping algorithm should assign intensity values to these intermediate regular pixel locations in order to form properly the warped image. A blown-out view **250** of the region containing rectangles **1** (**251**) and **2** (**249**) in FIG. **13** is shown in FIG. **14**. The shaded rectangular area **29** is the 'gap' between the warped nodes (i_w, j_w) and (i_w+m, j_w+n). This area may contain full regular pixel locations in the middle and partial regular pixel locations in the border (left, right, top, and bottom). Assign intensity values to these partial or full pixel locations by weighing the intensity at the warped node (i_w, j_w) with their area A $(0<A\leqq1)$.

Step 4. Apply the above two-dimensional interpolarization scheme from left to right and top to bottom until the nodes in the warped pixel grid are exhausted.

[0103] Step 5. Finally, map the normalized warped coordinates back to the unnormalized warped coordinates for proper positioning of the image in the universal image plane of the observed area.

[0104] Image processing component **102** can provide threat assessments of an object or person tracked. Component **102** can alert security or appropriate personnel to just those objects or persons requiring their scrutiny, while ignoring innocuous things. This is achieved by processing image data in image processing component **102** through a threat assessment analysis which is done after converting the pixel coordinates of the object tracks into a world coordinate system set by image processing component **102**. Known space features, fixed objects or landmarks are used in coordinate reference and transformation. The assembly of features uses the trajectory information provided by image processing module **102** to compute relevant higher level features on a per vehicle/pedestrian basis. The features are designed to capture "common sense" beliefs about innocuous, law abiding trajectories and the known or supposed patterns of intruders. The features calculated include:

[0105] number of sample points starting position (x,y)

[0106] ending position (x,y) path length distance covered (straight line)

[0107] distance ratio (path length/distance covered)

[0108] start time (local wall clock) end time (local wall clock) duration

[0109] average speed maximum speed

[0110] speed ratio (average/maximum)

[0111] total turn angles (radians)

[0112] average turn angles

[0113] number of "M" crossings

[0114] Most of these are self explanatory, but a few are not so obvious. The wall clock is relevant since activities on some paths are automatically suspect at certain times of day—late night and early morning particularly.

[0115] The turn angles and distance ratio features capture aspects of how circuitous was the path followed. The legitimate users of the facility tend to follow the most direct paths permitted by the lanes. 'Browsers' may take a more serpentine course.

[0116] The 'M' crossings feature attempts to monitor a well-known tendency of car thieves to systematically check multiple parking stalls along a lane, looping repeatedly back to the car doors for a good look or lock check (two loops yielding a letter 'M' profile). This can be monitored by keeping reference lines for parking stalls and counting the number of traversals into stalls.

[0117] The output of the feature assembly for trajectories is recorded from a site observed of some period of time and is

stored. That storage is used to produce threat models based on a database of features. During trial periods of time, several suspicious events can be staged (like "M" type strolls or certain inactivities) to enrich the data collection for threat assessments. Individual object or person trajectories or tracks may be manually labeled as innocuous (OK) or suspicious (not OK or a threat). A clustering algorithm assists in the parsimonious descriptions of object or person behavior. The behavior database consists of the labeled trajectories or tracks and the corresponding vectors. They are processed by a classification tree induction algorithm. Then the resultant classifier classifies incoming line data as OK or not OK.

[0118] Image processing component **102** detects and tracks moving objects or people. In the event that several people are moving alongside each other, they may be tracked as a single object, but may split into two or more objects and be detected as several tracks. Tracking can be lost because the people are obscured by equipment or natural obstructions. However, tracking will be correctly resumed once the people reappear. Additional cameras may be used if split tracks become a security loophole. Image processing **104** can recognize objects or people that disappear and appear within an FOV within short time intervals. This recognition function may be achieved by higher resolution cameras to capture detailed features of cars and especially humans. The cameras may have automated zoom mechanisms for being able to zoom in momentarily on every detected object or person and capture a detailed object signature. Tracking can be done under light or in the dark.

[0119] In summary, the tracking approach is based on a multi-Normal mixture representation of the pixel processes and on the Jeffrey divergence measure for matching to foreground or background states. This matching criterion results into dynamic segmentation performance. The tracking (MHT) algorithm and external multi-camera calibration are achieved through the computation of homographies. A warping algorithm which interpolates simultaneously across two dimensions addresses excessive deformations introduced by the nomography. A threat assessment analysis based on a tree induction algorithm reports suspicious patterns detected in the annotated trajectory data. The threat assessment analysis includes a clustering algorithm. The algorithm helps in the automation of assessment and classification of objects and people.

What is claimed is:

1. A tracking system comprising:
an imaging device;
a vision processor connected to said imaging device; and
an identifier connected to said vision processor.

2. The tracking system of claim **1** wherein said vision processor comprises:
an image processor; and
a tracker.

3. The tracking system of claim **2**, further comprising a fusion processor connected to said vision processor and said identifier.

4. The tracking system of claim **3**, further comprising a user interface component.

5. The tracking system of claim **4**, further comprising a connection to at least one database.

6. The tracking system of claim **5**, further comprising a sensor connected to said identifier.

7. The tracking system of claim **6**, wherein said user interface component comprises an operator monitor.

8. The tracking system of claim **7**, wherein said imaging device comprises a plurality of cameras.

9. The tracking system of claim **8**, wherein said sensor comprises a detector of people and objects.

10. The tracking system of claim **9**, wherein said identifier can positively identify people and objects detected by said detector.

11. A tracking system comprising:

means for imaging;

means for processing images from said means for imaging;

means for identifying;

means for fusing together said means for identifying and said

means for imaging; and

means for tracking connected to said means for imaging.

12. The tracking system of claim **11**, wherein said means for tracking can track at least one person or object and identify that person or object.

13. The tracking system of claim **12**, further comprising means for interfacing a user with said means for tracking.

14. The tracking system of claim **13**, wherein said means for interfacing a user comprises a means for monitoring said means for tracking.

15. The tracking system of claim **14**, further comprising a means for sensing connected to said means for identifying.

16. The tracking system of claim **15**, further comprising a means for interacting with a database.

17. A method for tracking comprising:

attaining images of an item to be tracked;

processing the images;

tracking the item; and

identifying the item.

18. The method for tracking of claim **17**, further comprising fusing the identifying the item with the tracking the item.

19. The method for tracking of claim **18**, further comprising interfacing a user with the method for tracking.

20. The method for tracking of claim **19**, further comprising interacting the method for tracking with a database.

* * * * *