US 20180300539A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2018/0300539 A1

Guan et al. (43) Pub. Date: Oct. 18, 2018

(54) METHOD AND APPARATUS FOR PATTERN TRACKING

(71) Applicant: **AIC Innovations Group, Inc.**, New York, NY (US)

(72) Inventors: **Lei Guan**, Jersey City, NJ (US); **Adam Hanina**, New York, NY (US)

(21) Appl. No.: **15/894,073**

(22) Filed: **Feb. 12, 2018**

Related U.S. Application Data

(63) Continuation of application No. 15/605,695, filed on May 25, 2017, now Pat. No. 9,892,316, which is a continuation of application No. 13/110,500, filed on May 18, 2011, now Pat. No. 9,665,767.

(60) Provisional application No. 61/447,243, filed on Feb. 28, 2011.

Publication Classification

(51) **Int. Cl.**

G06K 9/00 (2006.01)

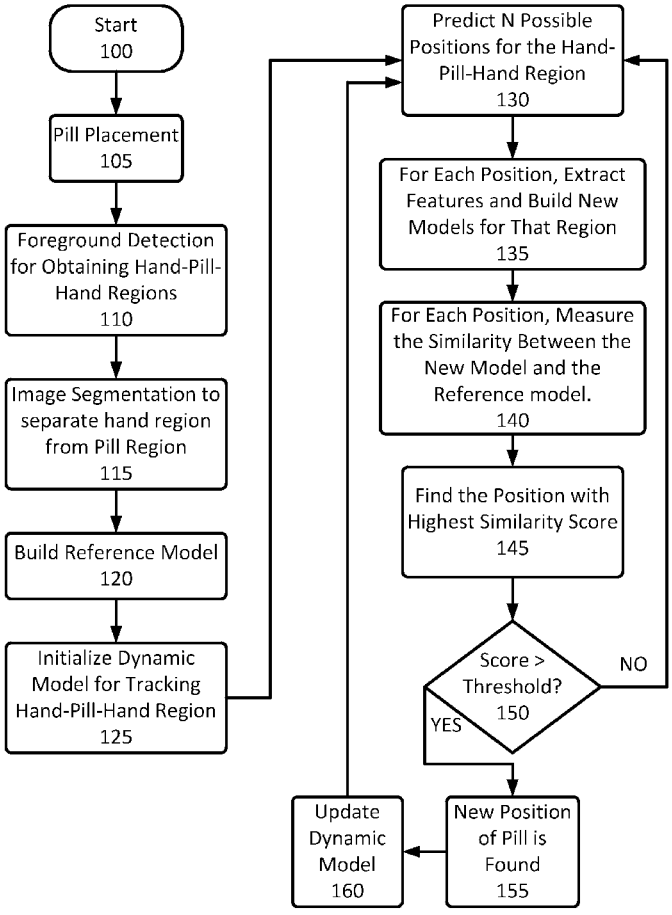
G06T 7/246 (2006.01)

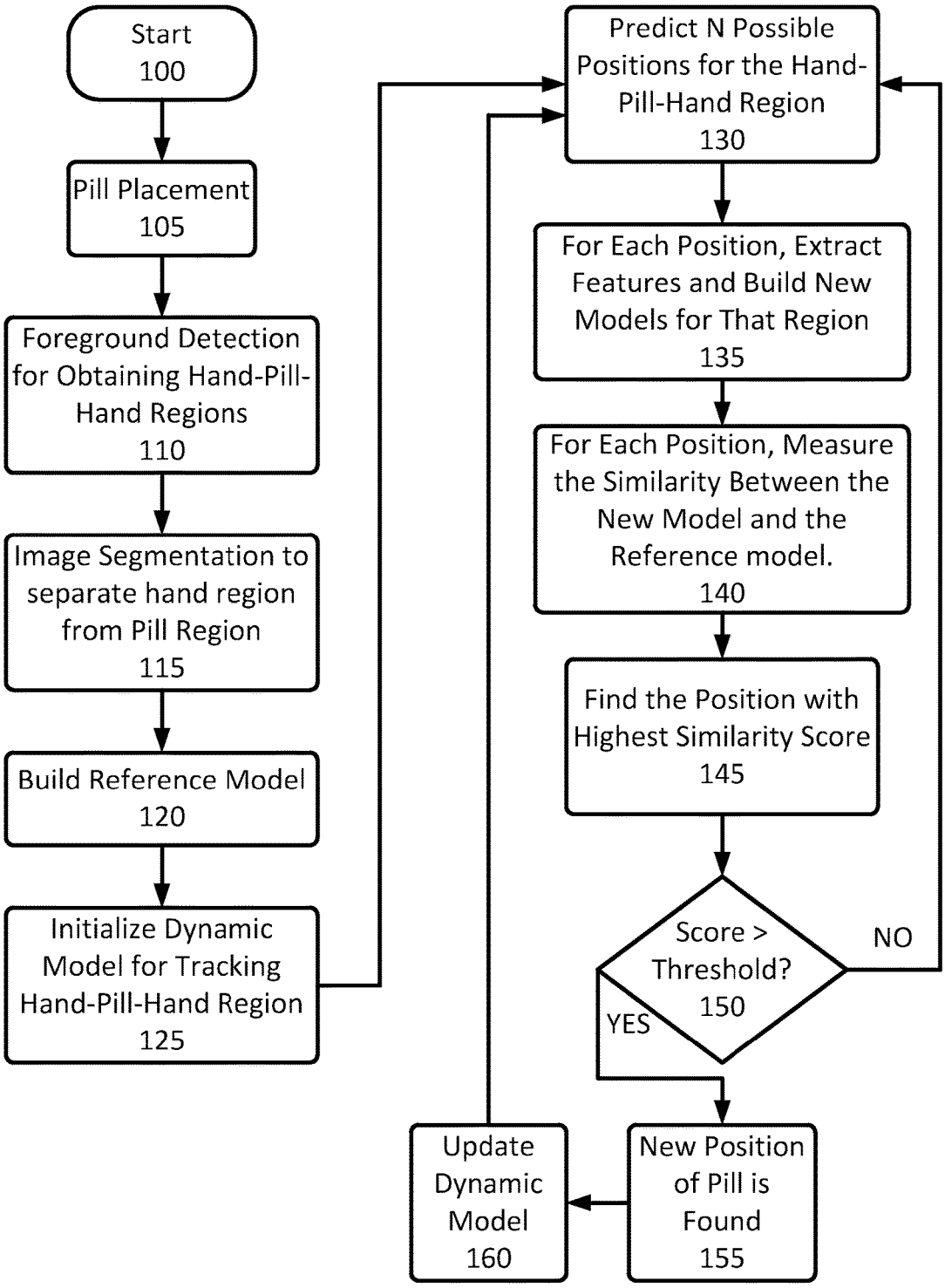
(52) **U.S. Cl.**

CPC ***G06K 9/00355*** (2013.01); ***G06T 7/251*** (2017.01); *G06T 2207/30196* (2013.01)

(57) **ABSTRACT**

A method and apparatus for pattern tracking. The method includes the steps of performing a foreground detection process to determine a hand-pill-hand region, performing image segmentation to separate the determined hand portion of the hand-pill-hand region from the pill portion thereof, building three reference models, one for each hand region and one for the pill region, initializing a dynamic model for tracking the hand-pill-hand region, determining N possible next positions for the hand-pill-hand region, for each such determined position, determining various features, building a new model for that region in accordance with the determined position, for each position, comparing the new model and a reference model, determining a position whose new model generates a highest similarity score, determining whether that similarity score is greater than a predetermined threshold, and wherein if it is determined that the similarity score is greater than the predetermined threshold, the object is tracked.




Start
100
Pill Placement
105
Foreground Detection for Obtaining Hand-Pill-Hand Regions
110
Image Segmentation to separate hand region from Pill Region
115
Build Reference Model
120
Initialize Dynamic Model for Tracking Hand-Pill-Hand Region
125
Predict N Possible Positions for the Hand-Pill-Hand Region
130
For Each Position, Extract Features and Build New Models for That Region
135
For Each Position, Measure the Similarity Between the New Model and the Reference model.
140
Find the Position with Highest Similarity Score
145
Score > Threshold?
150
NO
YES
New Position of Pill is Found
155
Update Dynamic Model
160

METHOD AND APPARATUS FOR PATTERN TRACKING

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Patent Application Ser. No. 61/447,243 to Guan et al., filed Feb. 28, 2011, titled "Method and Apparatus for Pattern Tracking", the entire contents thereof being incorporated herein by reference.

FIELD OF THE INVENTION

[0002] This invention relates generally to the tracking of objects through time and space, and more particularly to the recognition and tracking of a particular object being held by a particular holder by a single web camera or the like employing still or sequential video images. The invention also relates to the tracking of a pill or other medication held between the fingers of a medication administrator.

BACKGROUND OF THE INVENTION

[0003] Automatic identification and tracking of objects in three dimensional space utilizing only a standard web cam is difficult in that there is no simple way of determining distance from the camera. Furthermore, selecting a particular object from a plurality of objects may be difficult in that lack of depth perception does not allow for the differentiation of these objects based on position in a direction comprising distance from the camera (z-axis direction). Complicated images may therefore result in an unacceptable number of false positive or false negative responses.

[0004] Application of such a tracking scheme to a pill management system may be particularly troublesome in that a pill or other medication may be small, and may be colored similarly to a background object such as the shirt of a user, wall or other object. Furthermore, the user may move the pill or other medication quickly through the field of view of the camera. If implemented on a mobile device or the like, movement of the device in addition to movement of the pill may contribute to tracking difficulties, as may various environmental difficulties, such as poor lighting, background noise and the like. These variables may contribute to a very challenging situation for pill identification and tracking over time.

[0005] Therefore, it would be desirable to provide a method and apparatus that overcome the drawbacks of the prior art.

SUMMARY OF THE INVENTION

[0006] In U.S. patent application Ser. No. 12/620,686 filed Nov. 18, 2009 titled Method and Apparatus for Verification of Medication Administration Adherence; Ser. No. 12/646,383 filed Dec. 23, 2009 titled Method and Apparatus for Verification of Clinical Trial Adherence; Ser. No. 12/646,603 filed Dec. 23, 2009 titled Method and Apparatus for Management of Clinical Trials; and Ser. No. 12/728,721 filed Mar. 22, 2010 titled Apparatus and Method for Collection of Protocol Adherence Data, the entire contents of each of these applications being incorporated herein by reference, as well as in other co-owned applications, the inventors of the present invention describe a system for automatically monitoring patient adherence to a medication protocol. As part of that application, determination of when a user places a pill in their mouth is an important step. Such a determination further requires that such a pill is first determined to be present in a field of view, and thereafter tracked through that field of view.

[0007] As noted above, such a determination may be made particularly difficult when employing a simple webcam that does not include the ability to determine distance, such as with a time of flight camera or a stereo camera pair, and in particular if the determination is to be made on a small pill with a potentially close proximity to the webcam. In such a situation, an image captured by the webcam comprises a two dimensional picture of a scene without the ability to differentiate between near and far objects. Various complications as noted above may make such determinations even more difficult.

[0008] In accordance with various embodiments of the invention, known patterns may be exploited in order to track one or more objects. Thus, in one preferred embodiment of the present invention, knowledge of the color of the skin of a user, or user of a range of possible or potential skin tones, may allow for the tracking of a pill or other medication by searching for a combination of colors including "skin-pill-skin", thus allowing for differentiation of such a pill even from a background with a color similar to that of the pill. Once identified, the color sequence may be tracked through various images over time. If the image is lost, prompting may be provided to the user to place the pill at a particular location in the display to allow for re-identification and continued tracking of the pill. Such a tracking scheme may be extended to tracking any number of types of objects, and in particular any such objects being held by a user in a manner in which skin tone from the user's hand or the like is visible in a relatively fixed relationship to such object to be tracked.

[0009] The inventive system may also be extended to use in auditing of various desired action sequences, and in particular to a sequence to be performed by, for example, a surgeon, where the skin tone and medical device color combination may be identified and tracked to provide and automated audit of actions performed. Similar audit may be performed to other actions of doctors, nurses, or other healthcare providers. The system may also be applicable to other, non-medical applications.

[0010] Still other objects and advantages of the invention will in part be obvious and will in part be apparent from the specification and drawings.

[0011] The invention accordingly comprises the several steps and the relation of one or more of such steps with respect to each of the others, and the apparatus embodying features of construction, combinations of elements and arrangement of parts that are adapted to affect such steps, all as exemplified in the following detailed disclosure, and the scope of the invention will be indicated in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] For a more complete understanding of the invention, reference is made to the following description and accompanying drawings, in which:

[0013] FIG. **1** is a flowchart diagram depicting an embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0014] The invention will now be described making reference to the following drawings in which like reference numbers denote like structure or steps.

[0015] In accordance with one or more embodiments of the present invention, when attempting to track a pill being held and administered by a user, an assumption may be made that the pill is always held between two fingers, or otherwise held in the palm of the user. Employing this assumption, a basis for tracking of the pill in accordance with embodiments of the invention may comprise tracking a specific color and/or shape pattern "fingertip-pill-fingertip", "palm-pill-palm", "cup-pill-cup" after determining a fingertip or palm color and/or shape of the user, or the color of a holding device, such as a cup or the like (use of the fingertips and color will be continue to be used through the remainder of the description, although any of the alternative embodiments may be employed), and preferably knowing the color of the pill in advance, although determination of the pill color in real or near-real time may also be provided. As noted above, this sequence may also be applicable to other than a pill, such as a medical instrument or other device to be tracked. In one embodiment of the invention, the user may be asked to place the pill so that when imaged by a webcam, fills a particular portion of a display. A determination of the pill color may be made, and the colors immediately above and below the pill may further be determined to be the fingertips, and thus comprise the fingertip color, of the user. Alternatively, a user may be asked to place their fingers or palm, or both, in a specific location to be imaged so that the color thereof may be determined. Additional characteristics of the pill or other objects to be tracked, such as shape, color combinations, markings or the like may also be employed. The inventive system thereafter may split the imaged region into three sub-regions with individual reference color models for each region (i.e. a model for the top fingertip, a model for the pill and a model for the bottom fingertip), generating a pattern signature that may be distinguishable from other objects or sequences of objects that may enter the field of view of the web cam. Alternatively, as will be described below, a single reference color model may be used for the entire fingertip-pill-fingertip area.

[0016] In order to aid in determining whether a pill is a correct pill a comparison of the imaged pill with prior reference images of one or more known pills is made. In this particular instance, features including the color of the fingers and pill, as well as any other characteristic features of the pill and/or fingers that may be relied upon. The features of these images are then each preferably represented by a feature vector. In order to determine similarity between the images, a distance between each pair of image vectors may be calculated, using one or more known techniques, such as Euclidean distance, Mahalanobis distance or one or more other known machine learning methods. Each distance value is then preferably fit into a predetermined Gaussian distribution to generate a probability value which may be used to evaluate similarity. This predetermined Gaussian distribution may be populated by acquisition of various images and processing thereof to generate the distribution. A later acquired image with its feature vector is then compared to the library of acquired images to determine a distance between its feature vector, and the distribution of feature vectors, thus providing a probability value indicative of the likelihood of a match between two images. Thus, through this process, confidence rates, indicative of confidence of a match between an acquired set of images and a reference set of images, may be calculated for each region (finger, pill, finger, for example) by measuring the similarity of an observed color and other noted characteristics, and comparing with a predetermined characteristics each of the one or more reference images. In accordance with an embodiment of the invention, features Score level fusion, decision fusion, or other appropriate processing may be applied to those confidence rates to aid in determining if the region contains the desired pill or not. As a result, if a high level of assurance can be achieved that a newly acquired image is in fact a desired image, it may be possible to reduce further processing power, by reducing resolution, frame rate, etc.; requiring only enough processing power to confirm that the object has not changed. Such features may be beneficial when employing the system on a mobile or other reduced processing powered device.

[0017] In addition to comparing any such resulting score or confidence level to one or more thresholds to determine one or more levels of similarity, these confidence levels may be stored and compared over multiple medication administrations to determine whether the user is improving over time in taking their medication or not, and may be used to flag situations where the user may be trying to trick the system, or may benefit from additional training regarding the process of medication administration. Other factors such as lighting, background colors and other conditions may impact a particular acquired color or acquisition of other desired characteristic, and thus may also be employed to suggest a change of location, lighting, hardware, or other issues that may affect the ability of the system to properly acquire images of the system.

[0018] Advantages of the inventive method include, among others: 1) it is relatively robust to variations in shape, rotation, scale, lighting, movement of a mobile platform and the like; 2) it is relatively robust to changes background, or background that may have a color similar to the color of the pill or another object; 3) it uses the information based upon pattern identification assumptions.

[0019] As noted above, in accordance with various embodiments of the invention, a reference color model for the fingertip regions and pill region may be determined. Separating the colors in this region for proper separate acquisition, while potentially improving subsequent recognition of those elements, presents another challenge. In accordance with this particular embodiment of the invention, no previous knowledge or information regarding the pill, including its color, need be known. This information is preferably captured during run-time (i.e. a first step before the object is tracked). Therefore, the system preferably determines when a hand-pill-hand combination is within a region of interest, and then differentiates the two hand portions and the pill for the purposes of building reference color models. Embodiments of the invention may preferably employ motion information to determine whether the "hand-pill-hand" is within a predetermined region of interest of the image. If large areas of motion are measured, image segmentation techniques may be used to further analyze that region of the image. If that region is able to be segmented into three parts (hand, pill, hand), the inventive system is then able to build the color model for each region. If it is determined that the region cannot be segmented, the system

then may further determine that the image of the region being investigated may not include a proper hand-pill-hand combination. Or alternatively, if the region can be determined to include the hand-pill-hand combination, but for some reason cannot be properly segmented, the entire region may be used and included as a single entity.

[0020] Therefore, referring first to FIG. **1**, processing begins at step **100** and at step **105** a user is asked to place a pill in a particular portion of a field of view of a webcam, as may be indicated to them on a display including their image and a graphical or other locator for pill placement. Once placed in an appropriate location, a foreground detection process is performed at step **110** to determine a hand-pill-hand region in a manner as described above. Thereafter, at step **115**, image segmentation may be performed to separate the determined hand portion of the hand-pill-hand region from the pill portion thereof. Then, at step **120**, three reference models may be built, one for each hand region and one for the pill region based upon their appearances. Such appearances may include, but are not limited to, shape, color, texture, gray-scale intensity, histogram of colors, or other determinable attributes of the hand and pill. Alternatively, a single reference model for the entire hand-pill-hand region may be built. Next, at step **125**, a dynamic model for tracking the hand-pill-hand region is initialized.

[0021] Thereafter, based upon that dynamic model, N possible next positions for the hand-pill-hand region may be determined at step **130**, and at step **135** for each such determined position, various features may be determined, and a new model for that region may be built in accordance with the determined position. Next, at step **140** for each such position, a comparison is made between the new model and a reference model, and at step **145** a position whose new model generates a highest similarity score (thus, having the smallest feature vector difference) between it and the current reference model is determined. It is then inquired at step **150** whether that similarity score is greater than a predetermined threshold. If it is determined that the similarity score is not greater than a predetermined threshold, and thus the inquiry is answered in the negative, processing returns to step **130**, and an additional N possible hand positions may be presented. If on the other hand, the inquiry at step **150** is answered in the affirmative, and it is determined that the similarity score for one of the determined new model positions is grater that the predetermined threshold, processing continues with step **155**, where a new position of the pill (and thus the hand-pill-hand combination) is determined to have been found. Processing then moves to step **160** where a new dynamic model is determined, and processing then passes to step **130** where N possible new hand-pill-hand positions are determined.

[0022] As is noted in the description of FIG. **1**, both motion tracking and image segmentation techniques may be employed. Of course, while only one of these techniques may be employed in isolation, another advantage of the embodiments of the invention that employ motion and image segmentation techniques in combination is that the ratios of the three segmented sub-regions (top hand, pill, bottom hand) can be obtained. The ratios not only help to divide the tracking window into three sub-regions and but also help with subsequent tracking. Although a size of a tracking window may change over time, the approximate ratio of these segmented sub-regions is likely the same, a useful reference point for tracking.

[0023] In accordance with another preferred embodiment of the invention, a dynamic feature selection optimization scheme based on one or more decision fusion confidence levels for pill or other object recognition and tracking may be employed. As is well known, and as has been recognized by the inventors of the present invention, processing power is generally considered to be expensive, especially in mobile devices. It has further been determined by the inventors of the present invention that when confidence levels of object or pill identification or object or pill tracking are high, the need to perform high levels of processing may be significantly reduced. This may allow the system to be switched to possibly employ lower performance machine vision computations, resulting in better performance and potentially lower bandwidth requirements. A similar approach may be to downsample images based on confidence levels and mutliscaling. For example, an original image could be downsampled to 1/20, 1/10, 1/5 or the like. If the confidence is low in 1/20 scale, then the system may preferably utilize a next level until an acceptable confidence is achieved. Such adjustments may be made in a dynamic manner because the performance is continually monitored over time and within multiple frames to optimize confidence levels.

[0024] In addition, a process of feature selection may be optimized due to the fact that some features require higher computational power than others: a) in predictive color scanning, the inventive system may discard non-core pixel colors that do not match expected results to avoid wasting of computational bandwidth; b) when performing predictive shape scanning, the system may search algorithms based on identifying known corners or images; and c) the system may apply a similar methodology to unique markings on the pill or shape of interest instead of identifying points in a procedural manner. The system therefore may automatically narrow confidence levels to optimize confirmation, and in each case reduce the number of points needed for each feature to optimize accuracy.

[0025] In accordance with another embodiment of the invention, upon tracking of a pill or other object, if identification of such an object is difficult, the user may be encouraged by a proximity indicator to bring the pill close enough to the camera to ensure the correct amount of data is captured to both confirm likeness and to ensure the proper metrics are tracked. The system may then guide the user for correct proximity placement to maximize confidence levels of object verification.

[0026] Other objects may be tracked with a simple webcam by the user of similar image segmentation and other aspects of the invention. For example, a brand of a pair of shoes may be identified, and the wearer may be offered a discount when walking by. Based upon known patterns, cans of soda or other food items may be similarly tracked, perhaps requesting the user brings such a can closer to the webcam to interact with the user on a new sales promotion. The system may further be applied to injectable medications and the like, inhalers, or other medication delivery systems, and in particular may be employed to confirm activation and/or positioning thereof.

[0027] Therefore, in accordance with various embodiments of the invention, determination of a pill being in a patient's mouth may be tracked and confirmed. Because webcams cannot see depth, as long as the pill or other object is substantially surrounded by an unbroken ring of red color (the inner parts of the mouth), then one can safely assume

that the pill is no longer held by the fingers or the hand and has been placed in the mouth. The unbroken circle is another pattern to determine placement. The unbroken circle may be determined and tracked in place of the finger tip or palm as noted above. Identification sequences may be similar to those noted above with respect to the finger-pill-finger combination, but rather employing a color determination of mouth-pill-mouth.

[0028] It will thus be seen that the objects set forth above, among those made apparent from the preceding description, are efficiently attained and, because certain changes may be made in carrying out the above method and in the construction(s) set forth without departing from the spirit and scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

[0029] It is also to be understood that this description is intended to cover all of the generic and specific features of the invention herein described and all statements of the scope of the invention which, as a matter of language, might be said to fall there between.

What is claimed:

1. A method for tracking a pill in a patient's hand, comprising:
- performing a foreground detection process to determine a hand-pill-hand region;
- performing image segmentation to separate the determined hand portion of the hand-pill-hand region from the pill portion thereof;
- building three reference models, one for each hand region and one for the pill region;
- initializing a dynamic model for tracking the hand-pill-hand region;
- determining N possible next positions for the hand-pill-hand region;
- for each such determined position, determining various features to be employed in a comparison to a reference model;
- building a new model for that region in accordance with the determined position;
- for each position, comparing the new model and the reference model;
- determining a position whose new model generates a highest similarity score;
- determining whether that similarity score is greater than a predetermined threshold;
- wherein if it is determined that the similarity score is greater than the predetermined threshold, the object is tracked.

2. The method of claim **1**, wherein the reference models are combined into a single reference model.

3. The method of claim **1**, wherein the feature may be selected from the group of:
- shape, color, texture, gray-scale intensity and histogram of colors.

4. A method for tracking a pill in a patient's hand, comprising the steps of:
- performing a foreground detection process of an image or series of images to determine a hand-pill-hand region;
- performing image segmentation of the image or sequence of images to separate the determined hand portion of the hand-pill-hand region from the pill portion thereof;
- building one or more reference models including at least the hand region and the pill region;
- determining a feature vector of the hand hand-pill-hand region;
- comparing the determined feature vector to one or more reference feature vectors;
- determining a distance between the determined feature vector and each of the one or more reference feature vectors; and
- determining an image or set of images corresponding to the determined feature vector to be similar to an image or set of images whose corresponding feature vector when the determined distance is less than a predetermined threshold.

5. The method of claim **4**, further comprising the step of fitting each determined distance into a Gaussian distribution to determine a confidence probability of a match between the two corresponding images or sets of images.

6. A method for tracking a pill in a user's hand comprising the steps of:
- acquiring one or more images;
- determining a hand-pill-hand portion of one or more of the acquired images;
- storing an indication of one or more characteristics of and
- determining a location of the hand-pill-hand portion in a next one or more of the acquired images in accordance with the one or more stored indications.

7. The method of claim **6**, further comprising the step of storing a color signature of the hand-pill-hand portion of the one or more of the acquired images as the indication of the one or more characteristics.

8. The method of claim **7**, further comprising the step of determining the location of the hand-pill-hand portion in a next one or more of the acquired images in accordance with the stored color signature.

9. The method of claim **7**, further comprising the step of distinguishing a pill colored similarly to a background in accordance with the stored color signature.

10. The method of claim **9**, wherein the background is a user shirt.

11. The method of **9**, wherein the background is an environmental surface.

12. The method of claim **6**, wherein the step of determining a hand-pill-hand portion of one or more of the acquired images further comprises the steps of:
- performing a foreground detection process of an image or series of images to determine a hand-pill-hand region;
- performing image segmentation of the image or sequence of images to separate the determined hand portion of the hand-pill-hand region from the pill portion thereof; and
- building one or more reference models including at least one of the hand region and the pill region.

13. The method of claim **12**, wherein the method of claim **12** further comprises the step of determining a feature vector difference between at least one of the one or more reference models and a hand-pill-hand portion of one or more of the one or more acquired images.

14. The method of claim **13**, wherein if the vector difference is less than a predetermined threshold, there is determined to be a match.

15. The method of claim **12**, wherein the foreground detection process takes into account one or more environmental factors.

16. The method of claim **15**, wherein the one or more environmental factors comprises ambient light.

17. The method of claim **6**, wherein the one or more characteristics comprises shape.

18. The method of claim **6**, wherein the one or more characteristics comprises color.

19. The method of claim **6**, further comprising repeating the processing for a plurality of consecutive acquired images.

20. The method of claim **19**, wherein processing is performed on a subset of the plurality of consecutive acquired images.

* * * * *